ZKG INTERNATIONAL
ISSN: 2366-1313

TWITTER NETWORK LEARNING AUTOMATA URL CHARACTERISTICS FOR HARMFUL SOCIAL BOT IDENTIFICATION

[1]EEDUNURI UMA MAHESHWARI, [2]M.UDAY KUMAR

[1] M.Tech scholar, [2]Associate Professor, Dept of CSE,

JNTUH UNIVERSITY COLLEGE OF ENGINEERING, JAGTIAL, T.S, 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.

ISSN: 2366-1313

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

1. INTRODUCTION:

In our daily lives, social media has become increasingly crucial. People naturally flock to this medium to read and s hare news, given that billions of users produce and consume information every day. Social media bots are little programmes that can be deployed on social media platforms to perform a variety of useful and destructive functions while encouraging human behaviour. Some social media bots provide helpful services like weather and sports scores. These excellent social media bots are clearly labelled as such, and those who connect with them are aware that they are bots.

A huge majority of social media bots, on the other hand, are harmful bots masquerading as human users. Users lose faith in social media platforms' ability to offer accurate news as a result of these bots, since they suspect that the stories at the top of their feeds were "pushed" there by manipulative bots. Because so many individuals are using social media, malevolent users such as bots have begun to manipulate conversations in the direction that their makers desire. These malicious bots have been used for nefarious purposes such as spreading false information about political candidates, inflating celebrities' perceived popularity, deliberately suppressing protestors' and activists' messages, illegally advertising by spamming social media with links to commercial websites, and influencing financial markets in an attempt to manipulate stock prices. Furthermore, these bots have the ability to alter the outcomes of standard social media analysis. Social media bots use a variety of attack strategies, including: Sleeper bots are bots that sleep for lengthy periods of time before waking up to unleash an attack of thousands of postings in a short period of time (perhaps as a spam attack), and then sleep again. jacking the trend - the use of

ISSN: 2366-1313

top trending topics to focus on a certain audience for the purpose of targeting, An attacker employs a watering hole assault to estimate or watch which websites a company frequently visits and infects one or more of them with malware. Click farming or like farming-inflate fame or popularity on a website by like or reposting content via click farms, and hashtag hijacking- use of hashtags to focus an assault (e.g. spam, harmful links) on a specific audience using the same hashtag.

Twitter is regarded as a credible source of information. In addition to this, bot operated accounts can be 2.5 times more influential than human operated accounts. Malicious bots have been able to influence measures on Twitter, including the trending topics. These bots can also influence statistics performed on Twitter data, such as the top hashtags and the most important users in the data. Traditionally, the detection of malicious social bots are done through the usage of blacklisting methods. These are essentially lists of URLs collected by anti-virus groups which are known to be malicious. While these methods are fast (requiring a simple database lookup), and are expected to have low False Positive rates, a major shortcoming is that they fail against newly generated URLs. This is a severe limitation as new URLs are generated everyday. To address these limitations, there have been several attempts to solve this problem through the use of machine learning.

Here, 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 attacks, our proposed LA-based malicious social bot detection (LA-MSBD) algorithm integrates a trust computational model with a set of URL-based features for the detection of malicious social bots. The proposed trust computational model contains two parameters, namely, direct trust and indirect trust.

ISSN: 2366-1313

2. BOTS

Internet bots, also known as web robots, WWW robots or simply bots, are software applications that run automated tasks over the Internet. Typically, bots perform tasks that are both simple and structurally repetitive, at a much higher rate than would be possible for a human alone. The largest use of bots is in web spidering, in which an automated script fetches, analyzes and files information from web servers at many times the speed of a human. Each server can have a file called robots.txt, containing rules for the spidering of that server that the bot is supposed to obey. In addition to their uses outlined above, bots may also be implemented where a response speed faster than that of humans is required (e.g., gaming bots and auction-site robots) or less commonly in situations where the emulation of human activity is required. .

3.RELATED WORK:

In [1] Sneha Kudugunta, Emilio Ferrara (2018) proposed a deep neural network based on contextual LSTM (Long ShortTerm Memory) architecture allowing the use of both tweet content and metadata to detect bots at the tweet level. The contextual features are extracted from user metadata and fed as auxiliary input to LSTM deep nets processing the tweet text. From a single tweet, the model can achieve an extremely high accuracy exceeding 96% AUC.They also proposed methods based on SMOTE (Synthetic Minority Oversampling Technique) that yield a near perfect user-level detection accuracy (> 99% AUC) to enhance existing datasets by generating additional labeled examples. Both these methods use a very minimal number of features that can be obtained in a straightforward way from the tweet itself and its metadata. The system outperforms previous state of the art while leveraging a small and interpretable set of features yet requiring minimal training data.

In [2] Mohammed AL - Janabi, Ed de Quincey, Peter Andras (2017) proposed a supervised machine learning classification model to detect the distribution of malicious content in online social networks (ONSs). The multi-source features have

ISSN: 2366-1313

been used to detect social network posts that contain malicious Uniform Resource Locators (URLs). These URLs could direct users to websites that contain malicious content, drive-by download attacks, phishing, spam, and scams. For the data collection stage, the Twitter streaming application programming interface (API) was used and Virus Total was used for labelling the dataset. A random forest classification model was used with a combination of features derived from a range of sources. The random forest model without any tuning and feature selection produced a recall value of 0.89. After further investigation and applying parameter tuning and feature selection methods, however, we were able to improve the classifier performance to 0.92 in recall.

In [3] Chongzhen Zhang, Yanli Chen, YangMeng (2020) proposed A Novel Framework Design of Network Intrusion Detection Based on Machine Learning Techniques. We propose a novel intrusion detection framework to improve classification capabilities. Simultaneously, the retraining of the classifier in the classification module is realized through the database module and the feedback module so as to ensure the high accuracy rate of the classification module continuously. We developed a novel classification method by combining SAE and RF. Our approach realizes the potential of effective representation and dimensionality reduction to improve the classification results for traditional ML algorithms in binary and multiclass classification. We take full use of the characteristics of the SAE model, combined with the feature library in the database module, to restore the flow before compression, which can be used for post event analysis and forensics. We evaluate our proposed framework using the CICIDS2017 dataset and

give the training and testing times. Compared with different methods in the related work using the same dataset, we have achieved the best value in the binary and multiclass classification.

In [4] Sylvio barbon JR,Gabriel F.C.Campos,Gabriel M.Tavares(2018) proposed a Detection of Human, Legitimate Bot, and Malicious Bot in Online Social Networks

ISSN: 2366-1313

Based on Wavelets. The proposed approach was modeled in five steps: Acquisition, Profiling Setup, Features Extraction, Feature Selection, and Classification. Our model is suitable to any OSN, i.e., the Acquisition step can be adapted according to each OSN API. Classification step is also flexible. In this work, we adopted Random Forests (RFs) as classifiers. This choice was based on works such as Singh et al. (2014) and del Río et al. (2014), which reported good results by applying RFs in big data environments and Igawa et al. (2016) that treats specifically bots on OSNs. We have proposed an algorithm for classifying authors as being a human, a legitimate robot, or a malicious robot, in OSNs. The algorithm was based on Discrete Wavelet Transform to obtain a pattern of writing style embedded in post contents. Experiments have been conducted by classifiers with two different datasets: single and miscellaneous theme. It was observed that the proposed method yields the high average classification accuracies of 94.47% for both datasets. Considering the results, the text-based model we have developed gives promising accuracies in classifying the user type based on its writing style. We believe that the proposed algorithm can be very helpful to combat frauds in OSN. Further exploration of different machine-learning approaches can yield more interesting results.

4 .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,

ISSN: 2366-1313

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.

To overcome the restrictions of techniques we have introduced the deep learning technique dependent of Faster RCNN and RNN Using Twitter API we can post tweets including hash tag for particular topic. We generate Twitter data set to differentiate between human and bots. The tweets can be posted by registered users or by computer program. Computer programs are designed in such a way that they continuously tweet from some account on particular topic on behalf or against it for some neutral tweets. Usually when normal users tweet or reply to a particular topic the frequency of tweets will be normal. We will train the data sets using RCNN based algorithm. The input will be processed using RCNN to find out the sentiments. To detect if it is a human generated or not we will check the time interval of the Accounts suppose its within 10 minutes. If it is more than 8-10 minutes then it was suspected. We will also analyze the type of treats that has been posted from a user account. In the course of identification of hu an account deception large corpus of data is obtained. This data is recycled since it will contain information that do not contribute to the end results. The data is mined, impure data is removed and then stored in a database. This data is applied to machine learning models and results are obtained.

4.1. ACQUISITION AND PROFILING SETUP

The acquisition step consists of getting data from an OSN to create a textual data set. Our model requires the collected textual set to be grounded on one main subject .This is requirement is necessary, because wavelet-based text mining takes analysis of signalized key terms to obtain any further knowledge. In case the dataset is not grounded on one main term, we suggest the extraction of relevant terms. The use of a supervised deep-learning approach, we need a l dataset to induce a method. The text of each user is concatenated cumulatively following the profile-based paradigm and a textual repository with several users' posts. This step is called Profiling Setup and

ISSN: 2366-1313

provides the data that will be processed to obtain the feature vector that describes each class.

4.2. DATA BOOTSTRAP

Data Bootstrap Consists of Social Bots, unlabeled Data normal user data which contains all the details each of them such as network, user, making friends, content,tweeting and emotion, the attribute where can we detect malicious social bots.

4.3. FEATURE EXTRACTION

This phase which aims to collect relevant information about the malicious bots. This includes information such as presence of the URLs in a blacklist, features obtained from the URL String, information about the host, the content of the website such as HTML and JavaScript, popularity information, etc., gives an example to demonstrate various types various types of information that can be collected from an online social network to obtain the feature representation.

4.4. FEATURE PREPROCESSING

In this phase, the unstructured information about the data is appropriately formatted, and converted to a numerical vector so that it can be fed into Deep learning algorithms. For example, the numerical information can be used as is, and Bag-of-words is often used for representing textual or lexical content.

4.5. TRAINING CLASSIFIERS

We propose a general framework used for online classification and offline training. A classification problem consists of taking an input vector with data and deciding. It

ISSN: 2366-1313

follows a supervised learning process based on training from instances of each class the most important learning feature is the generalization: the algorithm should produce sensible output for inputs that were not encountered during learning.

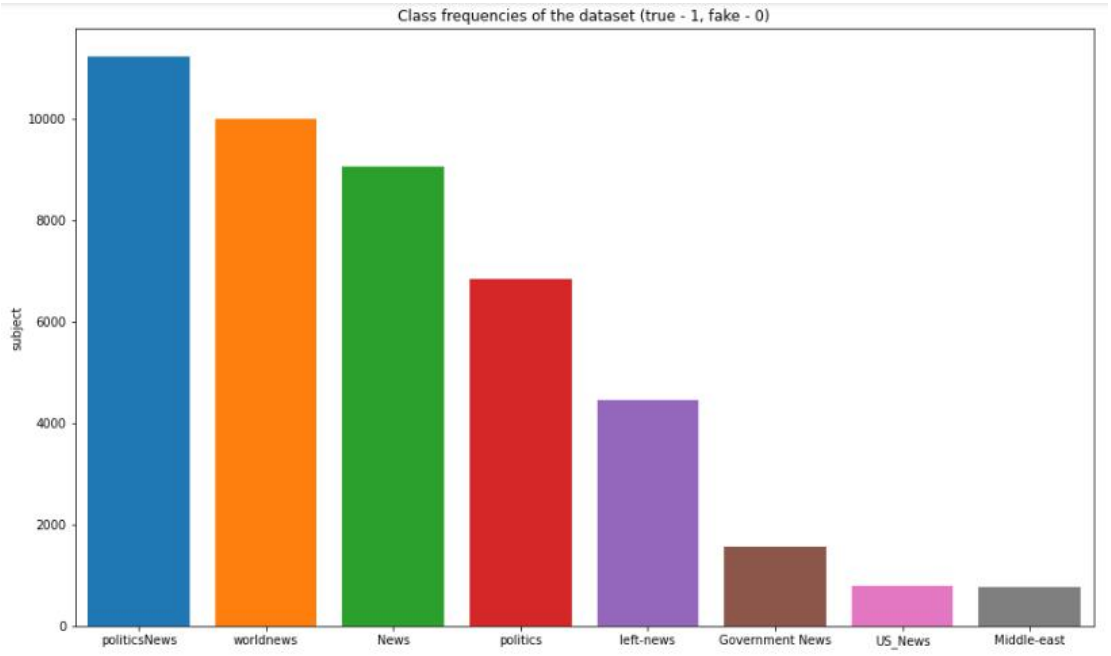

Figure 1: Categories of labels in the tweet data

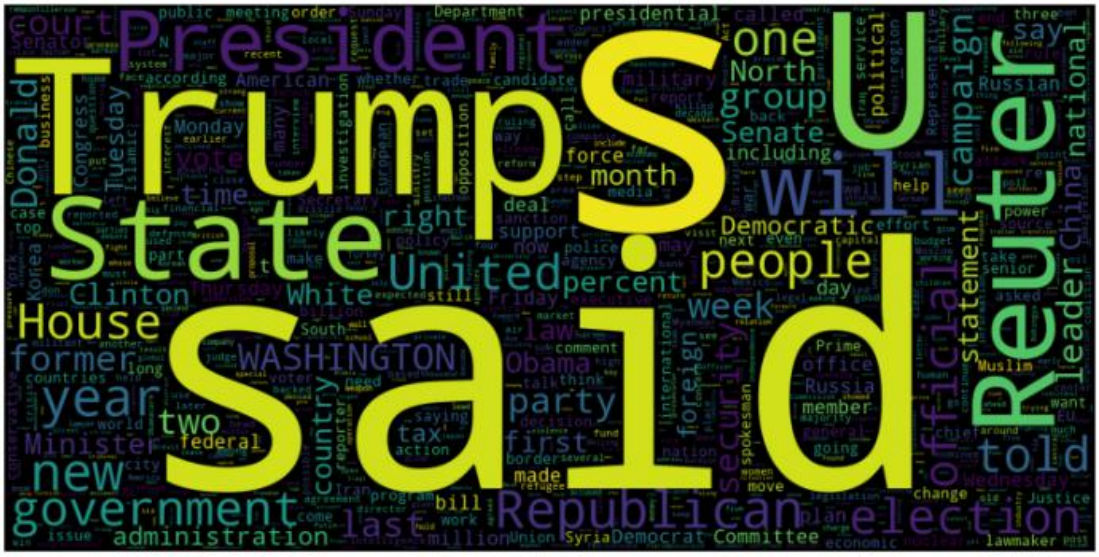

Figure 2 : Word vector in corpus representing the words

5 .CONCLUSION:

ISSN: 2366-1313

The need for new, low-cost Bot detection systems is evident given the frequency of detecting malicious bots on social media sites such as Twitter. We suggested a Naive Bayes and Random Forest (RF) algorithm for detecting tweets or URLs that are potentially fraudulent or damaging to users. So far, we have downloaded and installed all of the software that is required for the planned system. The dataset was obtained from the Kaggle website, and the preparation stage was completed. The features of preprocessed data will be extracted in the next phase, and the method will be implemented, with a model saved that can be used to categories the data. .

6.REFERENCES:

[1] G. Lingam, R. R. Rout, and D. V. L. N. Somayajulu, "Adaptive deep Q-learning model for detecting social bots and influential users in online social networks," Appl. Intell., vol. 49, no. 11, pp. 3947–3964, Nov. 2019.

[2] D. Choi, J. Han, S. Chun, E. Rappos, S. Robert, and T. T. Kwon, "Bit.ly/practice: Uncovering content publishing and sharing through URL shortening services," Telematics Inform., vol. 35, no. 5, pp. 1310– 1323, 2018.

[3] S. Madisetty and M. S. Desarkar, "A neural networkbased ensemble approach for spam detection in Twitter," IEEE Trans. Comput. Social Syst., vol. 5, no. 4, pp. 973–984, Dec. 2018.

[4] 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.

[5] 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

[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.

ISSN: 2366-1313

[7] 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.

[8] 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.

[9] 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. [27] G. Lingam, R. R. Rout, and D. V. L. N. Somayajulu, “Detection of social botnet using a trust model based on spam content in Twitter network,” in Proc. IEEE 13th Int. Conf. Ind. Inf. Syst. (ICIIS), Dec. 2018,

[10] A. Moayedikia, K.-L. Ong, Y. L. Boo, and W. G. S. Yeoh, “Task assignment in microtask crowdsourcing platforms using learning automata,” Eng. Appl. Artif. Intell., vol. 74, pp. 212–225, Sep. 2018.

[11] G. Lingam, R. R. Rout, and D. Somayajulu, “Learning automatabased trust model for user recommendations in online social networks,” Comput. Electr. Eng., vol. 66, pp. 174–188, Feb. 2018.

[12] Manju, S. Chand, and B. Kumar, “Target coverage heuristic based on learning automata in wireless sensor networks,” IET Wireless Sensor Syst., vol. 8, no. 3, pp. 109– 115, Jun. 2018.

[13] G. Wang, X. Zhang, S. Tang, C. Wilson, H. Zheng, and B. Y. Zhao, ‘‘Clickstream user behavior models,’’ ACM Trans. Web, vol. 11, no. 4, Jul. 2017, Art. no. 21.

[14] Y. Liu, C. Wang, M. Zhang, and S. Ma, ‘‘User behavior modeling for better Web search ranking,’’ Front. Comput.Sci., vol. 11, no. 6, pp. 923–936, Dec. 2017.

ISSN: 2366-1313

[15] M. Al-Qurishi, M. S. Hossain, M. Alrubaian, S. M. M. Rahman, and A. Alamri, ‘‘Leveraging analysis of user behavior to identify malicious activities in large-scale social networks,’’ IEEE Trans. Ind. Informat., vol. 14, no. 2, pp. 799– 813, Feb. 2018.

[16] A. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” IEEE Communications Surveys & Tutorials, vol. 18, no. 2, pp. 1153–1176, 2017. [2] N. Farnaaz and M. A. Jabbar, “Random forest modeling for network intrusion detection system,” Procedia Computer Science, vol. 89, pp. 213–217, 2016.

[17] H. Wang, J. Gu, and S. Wang, “An effective intrusion detection framework based on SVM with feature augmentation,” Knowledge-Based Systems, vol. 136, pp. 130–139, 2017.

[18] I. M. Akashdeep, I. Manzoor, and N. Kumar, “A feature reduced intrusion detection system using ann classifier,” Expert Systems with Applications, vol. 88, pp. 249–257, 2017.

[19] Y. Chuan-Long, Z. Yue-Fei, F. Jin-Long et al., “A deeplearning approach for intrusion detection using recurrent neural networks,” IEEE Access, vol. 5, pp. 21954–21961, 2017.

[20] M. Lopez-Martin, B. Carro, and A. SanchezEsguevillas, “Application of deep reinforcement learning to intrusion detection for supervised problems,” Expert Systems with Applications, vol. 141, Article ID 112963, 2019.