International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue: 3 Volume No.8May - June – 2024
DOI:10.46647/ijetms.2024.v08i03.025 ISSN: 2581-4621

Automated Bot Detection on Twitter Using URL Patterns and Learning Automata

Mr. S.V Hemanth[1], S Sneha Reddy[2], R Nithin[3], G Keerthi[4], Shinde Vinayak Rao Patil[5]

[1]Associate Professor, Computer Science Engineering, Hyderabad Institute of Technology and Management. [2]Student, Computer Science Engineering, Hyderabad Institute of Technology and Management.
[3]Student, Computer Science Engineering, Hyderabad Institute of Technology and Management.
[4]Student, Computer Science Engineering, Hyderabad Institute of Technology and Management.
[5] Student, Computer Science Engineering, Hyderabad Institute of Technology and Management

Abstract— The fight against fake news and propaganda on social media becomes increasingly difficult as malicious bots impersonate real users. These imposters spread misinformation through compromised or inauthentic accounts, often tricking users with shortened URLs that contain viruses and lead to malicious websites. Therefore, distinguishing between these bots and genuine Twitter users is crucial. Analyzing user interactions within the social network can be time-consuming. This research proposes a more efficient approach: LA-MSBD, an algorithm that relies on learning automata. LA-MSBD focuses on URL-based features like signs of spam and redirection frequency to identify bots. It also incorporates a trust assessment model that combines two methods: direct trust based on Bayes' theorem and indirect trust derived from Dempster-Shafer Theory (DST). This approach helps identify trustworthy individuals on Twitter.

Keywords—online social networks(OSNs), learning automata (LA), deceitful social bots, and dependability.

I. INTRODUCTION

Social media platforms struggle with the rise of fake accounts, also known as malicious bots. These bots act like real people but engage in harmful activities like spreading spam, creating fake identities, manipulating reviews, and launching phishing attacks. Twitter's character limit encourages users to shorten URLs with services like bit.ly, but this can be exploited by bots to spread deceptive shortened URLs that lead to dangerous websites. These tactics highlight vulnerabilities in Twitter's system, particularly its methods for detecting spam. Researchers have proposed various approaches to identify spam on Twitter, focusing on the content of tweets, user connections, and profile information. However, malicious bots can still bypass location-based detection by manipulating their tweets and profiles. While analyzing user connections is more adaptable, the vast amount of social network data makes it time-consuming to analyze effectively. As a result, differentiating between bots and real users remains a challenge

Current methods for identifying malicious URLs rely on technical details of web addresses (DNS components and URL properties). However, bots can use redirection techniques to evade detection, making it difficult to catch them all. To ensure online safety, it's crucial to identify these malicious URLs posted by bots on Twitter.

Existing location-based methods often rely on pre- programmed algorithms that prioritize pre-defined characteristics over analyzing social behavior. These methods struggle to distinguish genuine information patterns from the ever-changing tactics of bots. This has led researchers to explore alternative learning techniques like learning automata (LA), which show promise in improving bot detection accuracy.

The proposed LA-MSBD algorithm, based on learning automata, uses URL features to differentiate between legitimate and harmful tweets. This method can effectively identify malicious bots on Twitter by incorporating a trust evaluation model. This model combines two approaches: direct trust based on user behavior analysis (Bayesian learning) and indirect trust based on connections with trusted users (Dempster-Shafer Theory). This strategy represents a significant step forward in the fight against malicious social bots.

II. PROBLEM STATEMENT

Phony accounts, also known as malicious social bots, pose a major threat to Twitter. These bots pollute the platform with fabricated tweets, automated interactions, and the spread of risky URLs. Thwarting these bots is crucial to safeguarding the trustworthiness of Twitter. Conventional methods that rely on analyzing social connections (social graph features) are often sluggish and easily fooled by bots that can manipulate these interactions.

A novel approach called LA-MSBD tackles this problem by focusing on characteristics extracted from URLs, which are more difficult for bots to forge. LA-MSBD examines how often URLs are redirected and shared, and incorporates a trust assessment model that considers both direct trust (using Bayes' theorem) and indirect trust (leveraging Dempster-Shafer theory). This combination of URL analysis and trust scores allows for a more precise way to differentiate between real users and bots. Tested with real data from Twitter, LA-MSBD surpasses existing methods by achieving superior precision, recall, F-measure, and accuracy. This makes LA-MSBD a dependable solution for identifying and eliminating
Bots.

III. OBJECTIVES

The primary objectives of the LA-MSBD algorithm for identifying malicious social bots on Twitter are:

- To introduce a novel approach using learning automata (LA) to detect harmful bots.
- To combine a trust assessment model with features extracted from URLs within tweets for effective bot detection.
- To apply the trust assessment model based on the principle that users with trustworthy connections are likely to be trustworthy themselves.
- To aggregate the trustworthiness scores of a user's immediate neighbors (one-hop neighbors) on Twitter using Dempster's combination rule, assuming independence between the trust scores from each neighbor.
- To demonstrate the effectiveness of the proposed method in real-world scenarios.
- To enhance the overall security and reliability of social interactions on Twitter.
-

IV. LITERATURE REVIEW

The increasing influence of social media platforms like Twitter on public opinion and information dissemination has raised concerns about automated accounts, or "bots," which can spread misinformation or provide automated services. Accurate detection of these bots is essential to maintain the integrity of online interactions. Research in this area has explored various approaches, including a study by Chen et al. (2017) that proposed a machine learning method for real-time Twitter spam detection using statistical features of tweets and user accounts. Rndic and Laskov (2014) examined how attackers could evade machine learning classifiers, demonstrating the need for robust classifiers. Yazidi et al. (2013) used learning automata to track spatiotemporal event patterns, highlighting the importance of adaptive systems in monitoring evolving data. Khojasteh and Myoid (2006) investigated learning automata for cooperation in multi-agent environments, emphasizing its potential for fostering cooperative behaviors in complex systems like Robo soccer. These studies collectively underscore the need for advanced, adaptive algorithms to enhance the security and reliability of digital interactions.

V. EXISITING SYSTEM

Currently, rule-based systems and machine learning algorithms that examine static information from user accounts and tweet content are the main methods used by Twitter to identify malicious social bots. Although these approaches have shown considerable success, they have a number of drawbacks. They mostly rely on past data, which may not be up to date to reflect new threats or changing bot strategies, which eventually reduces accuracy. Additionally, models trained on incomplete or biassed datasets frequently perform poorly in real-world scenarios and have difficulty adjusting to emerging trends, raising concerns about bias and generalization. Additionally, because bots can swiftly alter their strategies to avoid detection, traditional systems struggle to handle the dynamic nature of bot behavior. Additionally, these techniques can be resource- intensive, needing a large amount of computer power and knowledge, which restricts their use and accessibility, especially for researchers or smaller organizations.

VI. PROPOSED SYSTEM

To detect harmful social bots on Twitter, the proposed Learning Automata-based harmful Social Bot Detection (LA-MSBD) model combines URL-based features and a trust computation model. The model uses a framework that assesses tweets and user behaviors in order to analyze user behavior within the Twitter network. One important component of this methodology is the trust computation, which evaluates users' credibility based on URL features included in their tweets. Redirection chains, content classification, and domain reputation are some of these aspects. Through the use of learning automata, the model continuously refines its detection tactics, increasing its robustness and accuracy against new threats and changing bot behavior. The model's capacity to identify patterns suggestive of bot activity is improved by this adaptive learning in conjunction with the examination of URL characteristics.

VII. IMPLEMENTATION

I. *Methodology*

The LA-MSBD model is used in a multi-step process to identify harmful social bots on Twitter. To begin with, Data Collection entails compiling a dataset of Twitter URLs classified as dangerous or benign, which forms the basis of the model. Feature Extraction is the next step, in which important features to help with classification are extracted from the URLs, including redirection chains, URL shorteners, and domain repute. Supervised techniques such as KNN are employed in the Classification stage to classify the URLs.

According to the features that have been extracted. In the Testing stage, the model checks the URLs to see if they are malicious, frequently concentrating on trends such as bots frequently utilizing specific domains. Lastly, URL Feature Prediction classifies URLs based on these features, which aids in differentiating bots that spread dangerous links from legitimate users, increasing the precision of twitter bot detection.

Through the utilization of these characteristics, the LA- MSBD model efficiently recognizes and categorizes URLs, hence augmenting the strong identification of malevolent social.

Getting information links and tweets from users of the social media site is the first stage .After that, they purify the data to facilitate analysis. This involves determining the security of the links. Subsequently, a trust models specialized computer program is used to determine if the tweets are authentic or not. Lastly, they search for any phony accounts such as bots that are disseminating false information.

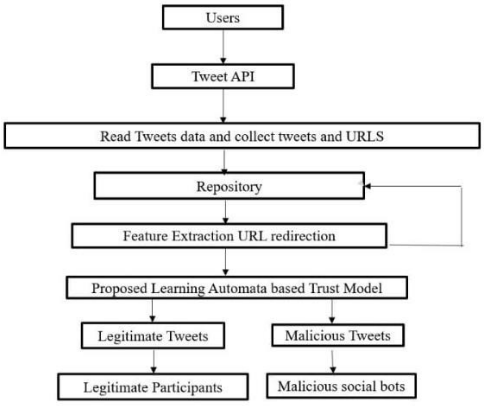

Users
Tweet API
Read Tweets data and collect tweets and URLS
Repository
Feature Extraction URL redirection
Proposed Learning Automata based Trust Model
Legitimate Tweets
Malicious Tweets
Legitimate Participants
Malicious social bots

Fig 1 : Methodology

II. *Algorithms*

a. *Direct Trust Computation*

The Direct Trust Computation algorithm starts by initializing the direct trust set as an empty set. For each tweet that a participant posts, the algorithm checks if the tweet was sent to a neighboring participant (friend). If it was, the feature set of the tweet is extracted and the features are ranked. Each feature is then evaluated to compute the probability that it is associated with malicious content. Based on this evaluation, a trust value for the tweet is computed. This trust value is then concatenated with the existing direct trust set, updating the set with the new value. After processing all tweets, the overall direct trust for the participant is computed.

b. *Indirect Trust Computation*

The Indirect Trust Computation algorithm begins by initializing the indirect trust set as an empty set. If a participant has one or more one-hop neighboring participants, the algorithm computes the direct trust for each of these neighbors using the direct trust computation method. The indirect trust set is then updated by concatenating it with the direct trust values of the neighbors. After processing all neighbors, the overall indirect trust for the participant is computed. If there are no one- hop neighboring participants, the indirect trust value is set to zero.

c. *LA-MSBD Algorithm*

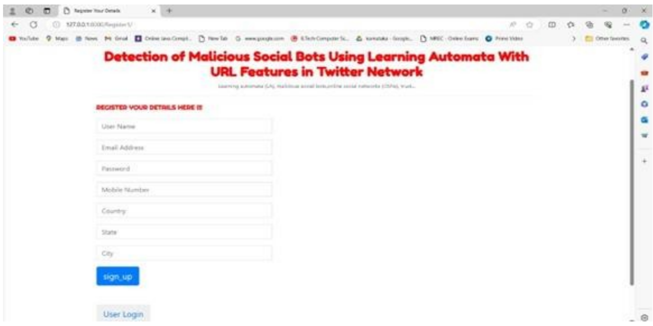
LA-MSBD algorithm begins by initializing three sets: one for storing malicious social bots (initially empty), another for temporarily storing decision-making data (also initially empty), and a third for storing trust values for all participants (initially empty). Learning automata are then activated for each participant in the network, with each participant having its own automaton to adjust their trust values based on behavior over time.

For each participant and each time slot within the total number of time slots, the algorithm computes direct trust using a direct trust computation method and indirect trust using an indirect trust computation method. These values are combined to calculate an overall trust value for the participant. Based on this trust value, an action probability value is determined. The trust value is then compared with a predefined threshold: if the trust value is below the threshold, the algorithm appends "1" to the temporary decision-making set; otherwise, it appends "0". The action probability value is updated accordingly for the next time slot.

After evaluating all time slots for a participant, the algorithms classify the participant based on the contents of the temporary decision-making set. If the number of "1"s exceeds the number of "0"s, the participant is classified as a malicious social bot, added to the set of malicious social bots, and penalized by a small reduction in their trust value. If the number of "0"s is greater than or equal to the number of "1"s, the participant is classified as legitimate, added to the list of legitimate participants, and their trust value is appended to the set of trust values. The temporary decision-making set is then reset for the next participant.

Finally, the algorithm returns the set of trust values along with the list of legitimate participants and the set of malicious social bots. This process systematically detects malicious social bots by analyzing both direct and indirect trust values over multiple time slots and adapting classifications based on the participants' trustworthiness.

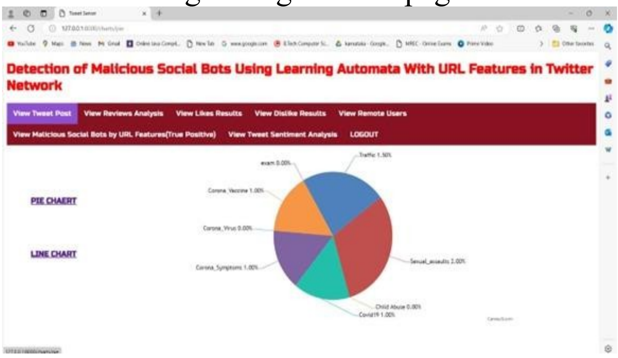
VIII. RESULTS



Detection of Malicious Social Bots Using Learning Automata With URL Features in Twitter Network

REGISTER YOUR DETAILS HERE !!!

Fig 2: Registration page



Detection of Malicious Social Bots Using Learning Automata With URL Features in Twitter Network

Fig 3: Pie Chart Analysis

IX. CONCLUSION

The goal of the research is to combine a trust computation model with various URL-based attributes to produce an LA- MSBD approach for MSBD. Furthermore, we use Bayesian learning and DST on each participant's tweets to evaluate their trustworthiness. Moreover, only a limited set of learning actions is trained by the suggested LA-MSBD method to update action probability values (i.e., the probability that a participant will tweet harmful URLs). The suggested LA-MSBD method produces benefits for incremental learning. We evaluate our suggested LA-MSBD algorithm's performance using two Twitter datasets.

X. FUTURESCOPE

Our goal is to investigate how attributes are interdependent and how that affects malicious social bot detection (MSBD). Subsequent versions may include real-time data integration, allowing Twitter to apply this feature to their app. Furthermore, integration with numerous commercially accessible social networking networks is a possibility. Currently, the detection dataset is provided manually, but our objective is to move the project forward to the point when the model collects the required dataset autonomously for bot identification.

References

[1] 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.

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

[3] M. R. Khojasteh and M. R. Meybodi, "Evaluating learning automata as a model for cooperation in complex multi-agent domains,"in Robot Soccer World Cup. Springer,2006, pp. 410–417.

[4] 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.

[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, p. 3.

[6] N. Rndic and P. Laskov, "Practical evasion of a learning-based classifier: A case study," in Proc. IEEE Symp. Secur. Privacy, May 2014, pp. 197–211.

[7] C. Chen, J. Zhang, X. Chen, Y. Xiang, and W. Zhou, "6 million spam tweets: A large ground truth for timely Twitter spam detection," in Proc. IEEE Int. Conf. Commun. (ICC), Jun. 2015, pp. 7065–7070.

[8] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of Twitter accounts: Are you a human, bot, or cyborg?" IEEE Trans. Dependable Secure Comput.,vol.9,no.6,pp.811–824,Nov.2012.