Review on Bot Detection on Social Media

Geetanjali Kuswaha[1], Dr. Nitesh Dubey[2]
[1] Dept of CSE
[2]Professor, Dept of CSE
[1, 2] Global Nature Care Sangathan Group of Institution, Jabalpur, Madhya Pradesh, India.

Abstract- *Social bots are computer algorithms in online social networks. They can share messages, upload pictures, and connect with many users on social media. Social bots are more common than people often think. Twitter has approximately 23 million of them, accounting for 8.5% of total users; and Facebook has an estimated 140 million social bots, which are between 5.5%–1.2% total users almost. In recent years Twitter bots have become increasingly sophisticated, making their detection more difficult. In contrast, many examples exist of cases where accounts created by bots fake or computers have been detected successfully using machine learning models. This paper presents a review of social bot and various approaches for detection of social bots, which find out open research questions and challenges in present methods.*

Keywords- Bot, Social bots, Twitter bots, Bot detection, Machine learning, Artificial intelligence.

I. INTRODUCTION

A lot of people use social media platforms not only to keep in touch with friends and family, but also to gather information and news from around the world. Thus, social media play a fundamental role in the news fruition. The case study for Britain reported in [1] shows an increase in the usage of social media, and more importantly their relevance to news consumption. Social media are powerful tools connecting millions of people across the globe. These connections form the substrate that supports information dissemination, which ultimately affects the ideas, news, and opinions to which we are exposed. *Social bots* are accounts controlled by software, algorithmically generating content and establishing interactions. A social bot is software to automate user activities. These activities can be (i) generating pseudo posts which look like human generated to interact with humans on a social network, (ii) reposting post, photographs or status of the others, and (iii) adding comments or likes to posts, (iv) building connections with other accounts. Therefore, the level of the sophistication of the bots is diverging. A social bot [2, 3] could be dummy like bots aggregating information from news, weather news, and blog posts and then reposts them in the social network. On the other hand, they also can be extremely sophisticated such as infiltrating human conversations. These capabilities have pros and cons for users of OSN and they can be used for good or bad intentions. The main research question is focused on "How we detect malicious activities on OSN". Many techniques are proposed to detect social bots on OSN in the literature. We review these techniques within a methodological categorization and unveil possible research avenues for each category for the social bot detection.

1.2 Bot Detection

An Internet bot is an automated software application. It can run any range of tasks and does so repetitively. The implementation of bots on the Internet is so widespread that bots made up 50% of all online traffic in 2016 [4]. Some of the tasks that bots perform are feed fetchers, commercial crawlers, monitoring, and search engine bots. For example, feed fetchers change the display of websites when they are accessed for mobile users and search engine bots collect metadata that allows the search engine to perform. These tasks shape the Internet as we see it daily.

A Twitterbot is an Internet bot that operates from a Twitter account. Some of the tasks that can be automated from a Twitter bot are writing Tweets, retweeting, and liking. Twitter does not mind the use of Twitter bot accounts as long as they do not break the Terms of Service through actions such as Tweeting automated messages that are spam or Tweeting misleading links. Twitter bots, like bots in general, serve a variety of purposes ranging from simple tasks such as following a user to more complex tasks like engaging in discussion with other users. Social bots are a type of bot that interacts with users and whose purpose is to generate content that promotes a particular viewpoint. The veracity of the content is irrelevant to the detection of the social bot. It is estimated that between 9 and 15 percent of Twitter accounts are bots [7]. The goal of our bot detection research is to develop refined techniques that are able to detect social bots that are actively avoiding being caught by traditional bot detection techniques. There are many types of bots on Twitter. One type of bot exists only to artificially increase the number of followers that an account has [5]. The number of Twitter followers determines its influence because the extent of the followers determines how widely spread is the account's

message. And the weight its message receives. People are more likely to trust an account with 1 million Twitter followers than 100 [6]. Using bots to artificially inflate the number of followers and account is a way to increase one's popularity and attract more human followers.

Figure 1.1 [11] provides a quick view of various kinds of fake profiles and several other kinds of profiles found in different online social networks. Real profiles have to be categorized into compromised and non-compromised ones which are also shown in the figure 1.1.

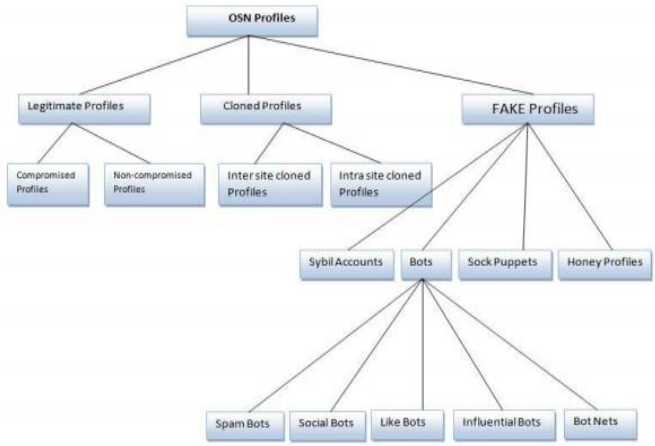

Figure 1.1: Evolution of Fake Profiles in OSNs [11].

Following are categories of bots [11]:

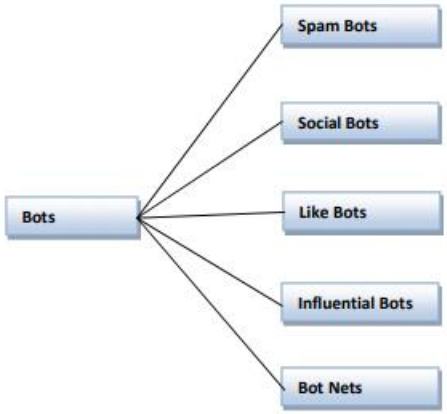

Figure 1.2: Type of Bots in OSNs [11].

According to the Global digital report 2019 [8] out of the world's total population of 7.676 billion, there are 4.388 billion internet users and 3.484 billion social media users. Almost half of the world's total population depends upon the internet for their knowledge. However, how much or up to what extent the circulated facts are verified is still a big question. How much we can rely on the information content that we are browsing every day. False information is created and initiated by a small number of people. People, relations, content and time are four critical dimensions of networked data analysed multi dimensionally by proposing an iOLAP framework based on polyadic factorization approach [9]. This framework handles all types of networked data such as microblogs, social bookmarking, user comments, and discussion platforms with an arbitrary number of dimensions. Origination, propagation, detection and Intervention are the four main facets of information pollution.

II. RELATED WORK

Generally, social bot detection on social networks is performed by one or more of the three common methods mentioned earlier: Graph-based, crowdsourcing, and machine learning. Figure below represents taxonomy of social bot detection approaches [10]:

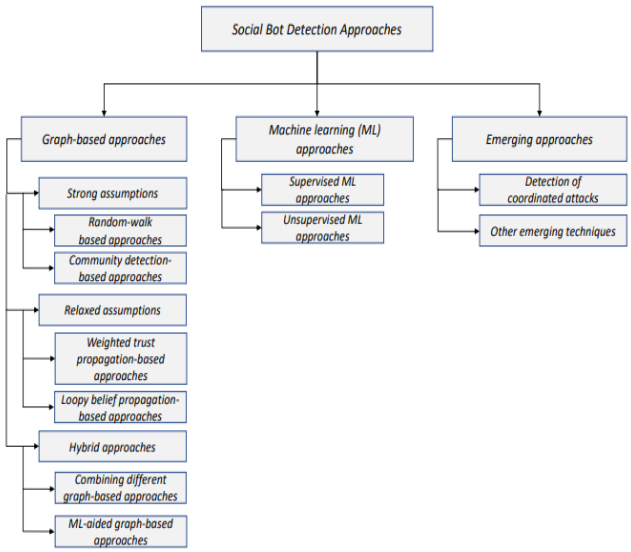

Figure 1.3: Taxonomy of Social Bot Detection Approaches[10]

The graph-based method involves using the social graph of a social network to understand the network information and the relationships between edges or links across accounts to detect bot activity. The crowdsourcing method involves using expert annotators to identify, evaluate, and determine social bot behaviors. Finally, the machine learning method involves developing algorithms and statistical methods that can develop an understanding of the revealing features or behavior of social network accounts in order to distinguish between human- and computer-led activities.

Machine learning (ML) has played significant roles in identifying malicious accounts in social networks. In fact, the majority of the articles on malicious accounts detection focused on machine learning. ML incorporates a variety of

methods, such as supervised, unsupervised, and semi-supervised learning. Supervised ML algorithm acquires a labeled dataset and learns a model as output, which can predict the class label for new data [12]. In supervised learning, the classifier learns from a large quantity of label data to build a model during training. Unsupervised learning (i.e clustering) differs in the sense that, no labeled data is present during the training stage, and the system learns from the data itself by identifying relationships or similarities among the instances in the dataset. Because the process of obtaining labeled data is tedious, a semi-supervised algorithm takes little labeled data in addition to a large amount of unlabeled data to produce a model.

2.1 Machine Learning Model

Ahmed et al. [13] use six supervised machine learning classifiers SVM, LSVM, KNN, DT, SGD, LR to detect fake reviews of hotels and fake news articles on the web using text classification. Their experiments achieve a significant accuracy of 90% and 92%respectively. Different content-based, features based, behavior-based and graph-based approaches can be used to detect opinion spams present in different formats of fake reviews, fake comments, social network posting and fake messages. In addition to the mainstream news media; there is also a concept of alternative media that aims to just present the facts and let readers use their critical thinking to explore reality by means of discussions.

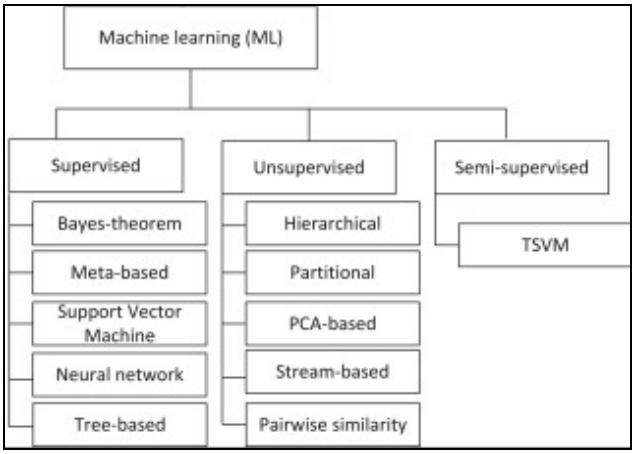

Figure 2.1: Machine Learning Methods[12]

The most recent work on the text in the field of fake news detection is given as follows: [14] assess the problem related to information credibility on Twitter. They have proposed an automated classification system, including four major components:

1) The reputation based technique,
2) A credibility classifier engine,
3) A user experience component, and
4) A feature rank algorithm.

Novelty and pseudo feedback (PF) based features have been introduced by [15] to detect rumours on early basis, along with features based on the presence of several URLs, hash-tags and user-names, POS tags, punctuation characters as well as eight different categories of sentiment and punctuation emotions. Many authors have worked on veracity classification task. [16] Introduced three sets of features related to linguistic, user-oriented, and temporal propagation. The Twitter dataset has been used for evaluation. Study reveals that the best performing features were those in the temporal category. Sarcasm is also one of the crucial issues over social media. M. Bouazizi*et al.* [17] assessed the problem related to sarcasm on twitter using pattern-based approach and introduced four sets of features that cover the different type of sarcasm and classified tweets as sarcastic and non-sarcastic. Social media is an open community where anyone can create their content, without any check on its veracity. Also, data present on social media is highly heterogeneous. Though, many credible sources are there whose integrity cannot be questioned and the content produced by them is verified and double checked. Inspired by these ideas, we exploit this property in our work

Support Vector Machines (SVMs) are one of the most widely used machine learning method for classification in a number of research areas. SVMs are discriminative classifiers formally defined by a separating hyperplane. According to the experiments in [18], SVMs have outperformed a number of supervised machine learning approaches for deception detection in text, obtaining an F-measure F 1 of 0.84. However, as pointed out by the authors themselves, there exists a significant variation in performance depending on the dataset selected for training [17].

Content-based features (e.g. linguistic and visual features) were exploited in most SVM-based approaches to fake news and deception detection. In particular, Afroz et al. [19] has obtained highly competitive scores for the task of deception detection on a number of datasets by exploiting only lexical, syntactic, and content-specific features. Rubin et al. [20] has trained an SVM for satirical fake news detection with a number of content-based features, obtaining an F 1 of 0.87.

III. CONCLUSION

Information pollution, fake news, rumours, misinformation, disinformation, Bot accounts has become a by-product of the digital communication ecosystem, which

proves to be very dangerous. This thesis work presents the impact analysis, characterization, compare and comprehensively evaluate the current scenario of methods, technologies, tools to quarantine the malice of information pollution through bot accounts in social media. This research tries to provide a holistic view of information pollution ecosystem in terms of taxonomy of fraudulent contents, lifecycle of a complete ecosystem, different social digital communication platforms, primary driving forces behind disinformation spread and different credibility analysis platforms.. This work may be helpful to the new researchers to understand the different components of digital online communication from a social and technical perspective. Improving the reliability and future of online information ecosystem is a joint responsibility of the social community, digital policymakers, administration, technical and research scholars.

REFERENCES

[1] N. Newman, W.H. Dutton, G. Blank, Social media in the changing ecology of news: the fourth and fifth estates in Britain, Int. J. Internet Sci. 7 (1) (2012) 6–22.

[2] Z. Chu, S. Gianvecchio, H. Wang, and S. Jajodia, "Detecting automation of twitter accounts: Are you a human, bot, or cyborg?" IEEE Transactions on Dependable and Secure Computing, vol. 9, pp. 811-824, 2012.

[3] V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, et al., "The darpa twitter bot challenge," arXiv preprint arXiv:1601.05140, 2016.

[4] Zeifman, Igal. "Bot Traffic Report 2016", Incapsula.com, Imperva,www.incapsula.com/blog/bot-traffic-report-2016.html

[5] Cresi, S., R., Petrocchi, M., Spognardi, A., & Tesconi, M. "Fame for sale: Efficient detection of fake Twitter followers". Decision Support Systems, 80, 56-71, 2015.

[6] A. Java, X. Song, T. Finin, and B. Tseng, "Why We Twitter: Understanding Microblogging Usage and Communities, "Proc. Ninth WebKDD and First SNA-KDD Workshop Web Mining and Social Network Analysis, 2007.

[7] Varol, Onur, et al. "Online Human-Bot Interactions: Detection, Estimation, and Characterization." Online Human-Bot Interactions: Detection, Estimation, and Characterization, 9 Mar. 2017

[8] Chi, Y., Zhu, S., Hino, K., Gong, Y., & Zhang, Y. IOLAP: A framework for analyzing the internet, social networks, and other networked data. IEEE Transactions on Multimedia, 11(3), 372–382., (2009)

[9] Vosoughi, S., Deb, S., The Spread of True and False News Online, Science, 359(6380), 1146–1151, 2018.

[10] Majd Latah, "The Art of Social Bots: A Review and a Refined Taxonomy", arXiv: 1905.03240v1, 8 May 2019

[11] Mudasir Ahmad Wani, Suraiya Jabina, "A sneak into the Devil's Colony- Fake Profiles in Online Social Networks", ArXiv 2017

[12] Kayode Sakariyah Adewole, Nor Badrul Anuar, Amirrudin Kamsin, Kasturi Dewi Varathan and Syed Abdul Razak, "Malicious accounts: Dark of the social networks", Journal of Network and Computer Applications, 2016

[13] Ahmed, H., Traore, I., &Saad, S. (2017), detecting opinion spams and fake news using text classification. Security and Privacy, 1(1), e9. https://doi.org/10.1002/spy2.9.

[14] Zhao, J., Cao, N., Wen, Z., Song, Y., Lin, Y. R., & Collins, C. (2014). Flux Flow: Visual analysis of anomalous information spreading on social media. IEEE Transactions on Visualization and Computer Graphics, 20(12), 1773–1782.

[15] Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., &Hovy, E. (2016).Hierarchical Attention Networks for Document Classification. 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 1480–1489.

[16] Shelke, S., & Attar, V. (2019), Source detection of rumor in social network – A review. Online Social Networks and Media, 9, 30–42.

[17] M. Bouazizi and T. O.Ohtsuki, "A Pattern-Based Approach for Sarcasm Detection on Twitter," IEEE Access, vol. 4, pp. 5477-5488, 2016.

[18] K. Shu , A. Sliva , S. Wang , J. Tang , H. Liu , Fake news detection on social media: a data mining perspective, ACM SIGKDD Explore Newslett. 19 (1) (2017)22–36.

[19] S. Afroz, M. Brennan, R. Greenstadt, Detecting hoaxes, frauds, and deception in writing style online, in: Proceedings of the Symposium on Security and Privacy, IEEE Computer Society, Washington, DC, USA, 2012, pp. 461–475.

[20] V.L. Rubin, Y. Chen, N.J. Conroy, Deception detection for news: three types of fakes 52 (1) (2015) 1–4.