International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

DETECTING MALICIOUS SOCIAL BOTS BASED ON CLICKSTREAM SEQUENCE

S. Pradyotha Rao[1], L. Sudha[2]

[1]Student, [2]Assistant Professor, Department of Computer Science and Engineering, S.A. Engineering College, Chennai – 77.

ABSTRACT: *The presence of bots has been felt in many aspects of social media. Twitter has especially felt the impact, with bots accounting for a large portion of its users. These bots have been used for malicious tasks such as spreading false information about political candidates and inflating the perceived popularity of celebrities. These bots can change the results of common analyses performed on social media. Malicious social bots have also been used to disseminate false information (e.g., fake news), and this can result in real-world consequences. Therefore, detting and removing malicious social bots in online social networks is crucial. A novel method of detecting malicious social bots, including both features selection based on the transition probability of click stream sequences and semi-supervised clustering, is presented in this paper. This method not only analyzes transition probability of user behavior click streams but also considers the time feature of behavior. Findings from our experiments on real online social network platforms demonstrate that the detection accuracy for different types of malicious social bots by the detection method of malicious social bots based on transition probability of user behavior click streams increases by an average of 12.8%, in comparison to the detection method based on quantitative analysis of user behavior.*

INTRODUCTION

In online social networks, social bots are social accounts controlled by automated programs that can perform corresponding operations based on a set of procedures. The increasing use of mobile devices (e.g., Android and iOS devices) also contributed to an increase in the frequency and nature of user interaction via social networks. It is evidenced by the significant volume, velocity and variety of data generated from the large online social network user base. Social bots have been widely deployed to enhance the quality and efficiency of collecting and analyzing data from social network services. For example, the social bot SF Quake Bot is designed to generate earthquake reports in the San Francisco Bay, and it can analyze earthquake related information in social networks in real-time. However, public opinion about social networks and massive user data can also be mined or disseminated for malicious or nefarious purpose. In online social networks, automatic social bots cannot represent the real desires and intentions of normal human beings, so they are usually looked upon malicious ones. For example, some fake social bots accounts created to imitate the profile of a normal user, steal user data and compromise their privacy, disseminate malicious or fake information, malicious comment, promote or advance certain political or ideology agenda and propaganda, and influence the stock market and other societal and economical markets. Such activities can adversely impact the security and stability of social networking platforms. In previous research, various methods were used to protect the security of online social network. User behavior is the most direct manifestation of user intent, as different users have different habits, preferences, and online behavior (e.g., the way one clicks or types, as well as the speed of typing). In other words, we may be able to mine and analyze information hidden in user's online behavior to profile and identify different users. However, we also need to be conscious of situational factors that may play a role in changing user's online behavior. In other words, user behavior is dynamic and its environment is constantly changing i.e., external observable environment (e.g., environment and behavior) of application context and the hidden environment in user information. In order to distinguish social bots from normal users accurately, detect malicious social bots, and reduce the harm of malicious social bots, we need to acquire and analyze social situation of user behavior and compare and understand the differences of malicious social bots and normal users in dynamic behavior. Specifically, in this paper, we aim to detect malicious social bots on social network platforms in real-time, by (1) proposing the transition probability features between user clickstreams based on the social situation analytics; and (2) designing an algorithm for detecting malicious social bots based on spatiotemporal features.

EXISTING SYSTEM:

Recent statistics show that more than 50% of Twitter accounts are not human users. Social network administrators are well aware of these harmful activities and try to delete these users using their suspension/removal systems. By one estimate 28% of accounts created in 2008 and half of the accounts created in 2014 have been suspended by Twitter. What is not well taken care of is the role of bots in facilitating these malicious activities.

According to the social interactions between users of the Twitter user to identify the active, passive and inactive users, a supervised machine learning method was proposed to identify social bots on the basis of age,

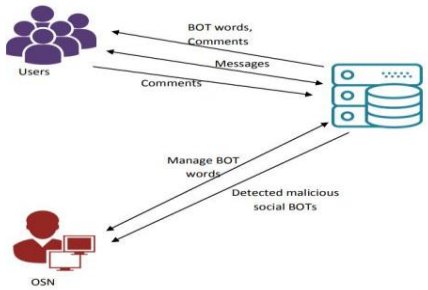
location and other static features of active, passive, and inactive users in the Twitter, as well as interacting person, interaction content, interaction theme, and some dynamic characteristics

The supervised learning method can be effective in detecting social bots. However annotation and training for large amounts of data are required in supervised learning. Tagging data requires time, manpower, and is generally unsuitable for the big data social networking environment.

PROPOSED SYSTEM:

In this paper, we aim to detect malicious social bots on social network platforms in real-time, by (1) proposing the transition probability features between user clickstreams based on the social situation analytics; and (2) designing an algorithm for detecting malicious social bots based on spatiotemporal features. In order to better detect malicious social bots in online social networks, we analyze user behavior features and identify transition probability features between user clickstreams Based on the transition probability features and time interval features, a semi-supervised social bots detection method based on space-time features is proposed.

Architecture diagram:



Modules:

1. DATA COLLECTION

The CyVOD platform comprises the website platform and Android and iOS applications. On CyVOD, the user clickstream behavior is obtained by a data burying point, and user clickstream data is collected server-side. In the realistic environment, for your own website, you can use the buried technology to get the corresponding data; for other websites, you need to get the data by working with the website or by calling the corresponding API (if provided).

2. EXPERIMENTAL DESIGN

Social bots that perform a single task, malicious social bots that coordinate to perform tasks, and malicious social bots that perform mixed tasks. For example, a user can perform two or more actions in the actions of liking, comment, sharing and so on. The social bot for malicious likes, the value of the P(play, like) (the transition probability of ‘‘the current click event is and the next click event is liking’’) would be high and the value of other transition probability features would be small or zero.

3. MALICIOUS SOCIAL BOTS DETECTION

1) Data cleaning:
Data that are clicked less must be cleaned to remove wrong data, obtain accurate transition probability between clickstreams, and avoid the error of transition probability caused by fewer data.

2) Data processing:
Some data are selected randomly from the normal user set and social bots set to the label. Normal user account is labeled as 1, and the social bots account is labeled as −1. Seed users are classified as the category of clusters.

3) Feature selection:
In the spatial dimension: according to the main functions of the CyVOD platform, we select the transition probability features related to the play-back function: P(play, play), P(play, like) , P(play, feedback), P(play, comment), P(play, share) and P(play, more) ; in the time dimension: we can get the inter-arrival times (IATs). Because if all transition probability matrixes of user behavior are constructed, extremely huge data size and sparse matrix can increase the difficulty of data detection.

4) Semi-supervised clustering method:
First, the initial centers of two clusters are determined by labeled seed users. Then, unlabeled data are used to iterate and optimize the clustering results constantly.

5) Obtain the normal user set and social bots set:
The normal user set and social bots set can be finally obtained by detecting.

6) Result evaluation:
We evaluate results based on three different metrics: Precision, Recall, and F1 Score (F1 is the harmonic average of Precision and Recall, F1 = 2· Precision·Recall Precision+Recall). In the meantime, we use Accuracy as metric and compare it with the SVM algorithm to verify the efficiency of the method. Accuracy is the ratio of the number of samples correctly classified by the classifier to the total number of samples.

FUTURE ENHANCEMENT

The existing system identifies the malicious users and blocks it from the particular website (for example: twitter,

facebook etc.). The proposed system works in a way that the malicious user will not be able to enter any website and will be blocked from all websites in the future.

CONCLUSION

We proposed a novel method to accurately detect malicious social bots in online social networks. Experiments showed that transition probability between user clickstreams based on the social situation analytics can be used to detect malicious social bots in online social platforms accurately. In future research, additional behaviors of malicious social bots will be further considered and the proposed detection approach will be extended and optimized to identify specific intentions and purposes of a broader range of malicious social bots.

REFERENCES

[1] F. Morstatter, L. Wu, T. H. Nazer, K. M. Carley, and H. Liu, ``A new approach to bot detection: Striking the balance between precision and recall,'' in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining, San Francisco, CA, USA, Aug. 2016, pp. 533_540.
[2] C. A. De Lima Salge and N. Berente, ``Is that social bot behaving unethically?'' Commun. ACM, vol. 60, no. 9, pp. 29_31, Sep. 2017.
[3] M. Sahlabadi, R. C. Muniyandi, and Z. Shukur, ``Detecting abnormal behavior in social networkWebsites by using a process mining technique,'' J. Comput. Sci., vol. 10, no. 3, pp. 393_402, 2014.
[4] F. Brito, I. Petiz, P. Salvador, A. Nogueira, and E. Rocha, ``Detecting social network bots based on multiscale behavioral analysis,'' in Proc. 7th Int. Conf. Emerg. Secur. Inf., Syst. Technol. (SECURWARE), Barcelona, Spain, 2013, pp. 81_85.
[5] T.-K. Huang, M. S. Rahman, H. V. Madhyastha, M. Faloutsos, and B. Ribeiro, ``An analysis of socware cascades in online social networks,'' in Proc. 22nd Int. Conf. World Wide Web, Rio de Janeiro, Brazil, 2013, pp. 619_630.
[6] H. Gao et al., ``Spam ain't as diverse as it seems: Throttling OSN spam with templates underneath,'' in Proc. 30th ACSAC, New Orleans, LA, USA, 2014, pp. 76_85.