IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received February 14, 2019, accepted February 21, 2019, date of publication February 26, 2019, date of current version March 18, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2901864

Detecting Malicious Social Bots Based on Clickstream Sequences

PEINING SHI[1,2], ZHIYONG ZHANG[1,2], (Senior Member, IEEE), AND KIM-KWANG RAYMOND CHOO[3], (Senior Member, IEEE)

[1]Information Engineering College, Henan University of Science and Technology, Luoyang 471023, China
[2]Henan Joint International Research Laboratory of Cyberspace Security Applications, Luoyang 471023, China
[3]Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, TX 78249-0631, USA

Corresponding author: Zhiyong Zhang (xidianzzy@126.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61772174 and Grant 61370220, in part by the Plan For Scientific Innovation Talent of Henan Province under Grant 174200510011, in part by the Program for Innovative Research Team (in Science and Technology) in the University of Henan Province under Grant 15IRTSTHN010, in part by the Program for Henan Province Science and Technology under Grant 142102210425, in part by the Natural Science Foundation of Henan Province under Grant 162300410094, and in part by the Project of the Cultivation Fund of Science and Technology Achievements of Henan University of Science and Technology under Grant 2015BZCG01.

ABSTRACT With the significant increase in the volume, velocity, and variety of user data (e.g., user-generated data) in online social networks, there have been attempted to design new ways of collecting and analyzing such big data. For example, social bots have been used to perform automated analytical services and provide users with improved quality of service. However, malicious social bots have also been used to disseminate false information (e.g., fake news), and this can result in real-world consequences. Therefore, detecting and removing malicious social bots in online social networks is crucial. The most existing detection methods of malicious social bots analyze the quantitative features of their behavior. These features are easily imitated by social bots; thereby resulting in low accuracy of the analysis. A novel method of detecting malicious social bots, including both features selection based on the transition probability of clickstream sequences and semi-supervised clustering, is presented in this paper. This method not only analyzes transition probability of user behavior clickstreams 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 clickstreams increases by an average of 12.8%, in comparison to the detection method based on quantitative analysis of user behavior.

INDEX TERMS Online social network, social bots, user behavior, semi-supervised clustering.

I. 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 [1]. 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 QuakeBot [2] 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 [3]. 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 [4], disseminate malicious or fake information [5], [6], malicious comment, promote or advance certain political or ideology agenda and propaganda [7],

The associate editor coordinating the review of this manuscript and approving it for publication was Wei Yu.

and influence the stock market and other societal and economical markets [8]. 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 [9]–[11]. 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 [12]. 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.

The rest of this paper is organized as follows. The second section briefly reviews related research. The third section presents the method for detection algorithms for malicious social bots, followed by the experiment and result analysis in the fourth section. The final section concludes this paper.

II. RELATED WORKS

A. BEHAVIOR ANALYSIS

Malicious users in social network platforms are likely to exhibit behavior patterns that different from normal users, because their goals in maximizing their own needs and purposes (e.g., promote a certain product or certain political beliefs or ideology). User behavior analysis is not only helpful in gaining an in-depth understanding of user intent, but it is also important to the detection of malicious social bots' accounts in online social networks. User behavior likely change under different situations. Chang [12] proposed that situation analytics can be included in software service requirement analysis, which can facilitate the analysis of any change in user's requirements. Such an analysis is useful to understand the dynamic needs of a software service environment. Zhang *et al.* [13] presented a framework to the discovery of user behavior pattern in multimedia video recommendation services on online social networks. Their framework is based on social context and analyzes the changes in user need for different social situations. Such user behavior data can be obtained if we have access to the user's logs [14] or user's clickstreams (e.g., recorded by social network platforms). The difference in user behavior can be obtained, for example, by analyzing the image search logs of users to study the search intention of different users [15], and this approach can facilitate optimization of search engines. Wang *et al.* [16] used user clickstream data to construct a clickstream graph model to represent user behavior and identify different user groups, in order to detect malicious accounts. There have also been other researches that indicate user intent and abnormal accounts can be determined through behavior analysis, and social situation in facilitating the understanding of users' dynamic behavior. Liu *et al.* [17] constructed a new convolutional neural network architecture based on user behavior, search engine content and context information to construct a click model and find out the user's click preferences to improve search quality. Al-Qurishi *et al.* [18] collected a large amount of user information on the Twitter and YouTube, about 13 million channel activities, analyzing and detecting abnormal behaviors that deviate significantly from large-scale specifications through user behavior in two social networks.

B. SOCIAL BOTS DETECTION

Botnets become widespread in wired and wireless networks. In particular, bots in a botnet are able to cooperate towards a common malicious purpose [19]. In recent years, social bots have become very popular in social networks, and they can imitate human activities in social networks. They are also programmed to work together to fulfill the prescribed tasks. There are a wide range of methods (e.g., sophisticated techniques and tools that may be associated with nation states and state-sponsored actors) used by some users with malicious or nefarious intent as well as social bots. For example, in order to imitate the features of human users successfully, social bots may 'crawl' for words and pictures from online social networks to complete fabricated user profiles and so on. Semi-social bots between humans and social bots have also reportedly emerged in social networks [20], which are highly complex social bots that bear the characteristics of human behavior and social bot behavior. The automated procedure for semi-social bot is generally activated by humans, and the subsequent actions are automatically performed by social bots. This process further increases the uncertainty of the operation time of social bots [21]. Social bots are generally more intelligent and they can more easily imitate human behavior, and they cannot be easily detected.

In existing literature, social bots are generally detected using machine learning-based approaches, such as BotOrNot [22] released by the Twitter in 2014. In BotOrNot, the random forest model is used in both training and analysis by using historical social information of normal users and social bots accounts. Based on six features (i.e. network, user, making friends, time, content and emotion), this model distinguished normal users from social bots. Morstatter *et al.* [1] proposed a heuristic-type supervised BoostOR model with increasing recall rate to detect malicious bots, which using the proportion of tweets forwarded to the published tweets on the Twitter, the mean length of tweets, URL, and forwarding interval. Wang *et al.* [16] constructed a semi-supervised clickstream similarity graph model for user behavior to detect abnormal accounts in Renren. 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 [23]. A time act model, namely, Act-M, was constructed focusing on the timing of user behavior activities [24], which can be used to accurately determine the interval between different behaviors of social media users to accurately detect malicious users. There have been focused on detecting semi-social bots too. For example, a management framework relying on entropy component, spam detection component, account attribute component, and decision maker was proposed by Chu *et al.* [20]. In the approach, Naive Bayes is adopted to categorize automated Twitter accounts into human, social bots, or semi-social bots. Previous studies have also shown that quantitative features such as friends, fans, forwarders, and tweets can be used in feature selection. 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. In other words, such an approach is generally ill-suited for real-time detection of malicious social bots on social networking platforms. Unsupervised learning, on the other hand, it does not require manual labeling of data. However, unsupervised learning approaches are sensitive to initial values and can only classify different results. It is not possible to determine which cluster is normal and which cluster is abnormal.

We also observe that social bots usually have similar features and same purpose. Unsupervised clustering algorithms can classify users into different clusters based on the similarity of users. To identify potential malicious social bots in online social networks in real-time, we analyze the social situation behavior of users in online social networks. We also evaluate user behavior features and select the transition probability of user behavior on the basis of general behavior characteristics. We then analyze and classify situation aware user behaviors in social networks using our proposed semi-supervised clustering detection method. This allows us to promptly detect malicious social bots using only a small number of tagged users.

III. PROPOSED METHOD FOR DETECTING OF MALICIOUS SOCIAL BOTS

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.

A. DEFINITIONS

Definition 1 [13]: SocialSitu(t) denotes the situational information at moment t. SocialSitu(t) is a four-tuple SocialSitu(t) $= \{ID, d, A, E\}$, where ID refers to the user's identity information (including the group to which the user belongs to and the role of user in the group), d refers to user's wishes at the t time, A refers to user operation corresponding to d at the particular moment (namely, behavior), and E refers to environmental information (e.g., terminal devices, equipment information and location information).

Definition 2: Clickstream is the order of clicks when users visit some websites or use the mobile terminals. The user's click event is a single point of operation. Clickstream is a series of point operations, and it refers to the SocialSitu(t) sequence of user from start point to target achievement. The sequence of the clickstream I $= \{SocialSitu(1), SocialSitu(2), \ldots, SocialSitu(n)\}, n \in N$, $SocialSitu(1)$ refers to the user's first click behavior on the platform. $SocialSitu(n)$ refers to the last click event that the user performs prior to exiting the platform.

Definition 3: The collection of user clickstream consists of many clickstream sequences, the collection of clickstreams, namely Click(s) $= \{I(1), I(2), \ldots, I(m)\}$, s refers to the user id.

Definition 4: Transition probability between click steams: $P_{(i,j)}$ represents the probability that the click event is j at $t+1$ moment when the click event is i at t moment. Transition probability $P_{(i,j)}$ refers to (1). Here, $\sum\limits_j X(t) = i$ means the total number of transitions that may occur in click status i among all user click events. $\sum\limits_j \{X(t+1) = j, X(t) = i\}$ means the total number of click events when the click event i at t and click event j at $t+1$.

$$P_{(i,j)} = \frac{\sum\limits_k \{X(t+1) = j, X(t) = i\}}{\sum\limits_k X(t) = i}, \quad k \in \mathrm{N} \qquad (1)$$

B. FEATURE SELECTION BASED ON TRANSITION PROBABILITY OF CLICKSTREAM SEQUENCES

The malicious behavior of social bots refers to a variety of behaviors performed by social bots for a specific purpose. However, the behaviors involved in this paper are not necessarily malicious behaviors, which are related operations that malicious users are most likely to perform for different social network platforms to achieve their goals. For example, social bots may achieve different purposes by performing the main function-related operations in Twitter, such as posting tweets, comments, forwarding tweets and so on. In the social networking platform, we usually determine whether the corresponding behavior is normal or malicious based on the final result of the user behavior. For instance, we determine whether a comment is malicious by analyzing whether the user's comment content contains ads. However, with the constant evolution of social bots, simple text analysis is difficult to detect comments because they can spread the message by posting images or more subtle text. As we all know, social bots achieve different purposes according to the main functions of the platform, and they perform different behaviors in different social networks. Therefore, in this paper, we focus on the operations related to the main functions of the experimental platform. These operations are not necessarily malicious, but are most likely to be performed by malicious social bots to meet different purposes.

Malicious social bots search the Internet for information and picture to fill personal information and simulate the human time features in content production and consumption. The user's profile picture and other personal data features, likes, comments, and some quantitative features are easily imitated by malicious social bots. Thus, the detection efficiency is also gradually reduced. To explore robust features, user behavior features should be deeply analyzed and expanded. The clickstream sequences can reflect the dynamic changes of the user behavior, while also hiding the important behavior features of the user. We get more information on the click behavior in three ways, namely: (1) In terms of user behavior data acquisition, we employ user clickstream sequences under situation aware environments, rather than simply click events. Social situation analytics can be used to acquire the external observable environment of applied scenarios and the hidden environment of user information in time. (2) In terms of user behavior features selection, we extend user behavior features from the single click behavior to the linear features of clickstream sequences, which can better reflect user intent in special situations. (3) In the dimension of user behavior features, we add temporal dimension features to the spatial dimension of user behavior features, and analyze user behavior features in multiple dimensions, which make user behavior features more robust.

The differences between different users can be described by sequence analysis on user clickstream behavior. The transition probability between clickstreams is an important hidden feature in user clickstream sequences, which can reflect the user behavior habits and preferences in different situations. Compared with the quantitative feature, the transition probability features are more robust and not easily imitated. The malicious social bots can imitate the quantitative features of normal users by setting the number of related behaviors (e.g., the number of likes, the number of comments, the number of friends, etc.), and the user cannot observe the transition probability features on the online social network platform. It is also difficult for malicious social bots to mimic the features of normal users. Meanwhile, the transition probability between user click streams also cannot be obtained directly by querying the data in the database or the user's click stream log.

Based on the function provided by the experimental platform, we find that the main function of the experimental platform is playing the video and the malicious social bot often use playback-related operations such as comments and likes to achieve their goals. The paper mainly focused on the following behaviors, such as playing, liking, sharing, commenting, sharing and reporting, as well as their combinations based on user's click events and clickstream sequences. By analyzing clickstream sequences, we use the window sliding method to obtain the number of transitions between specific click events. At the same time, we choose the inter-arrival times (IATs) between user-specific click events based on the time feature in the user's click behavior. Based on the playing behavior, we can get the difference between the normal user and social bot in the time and space dimension. For instance, the normal user prefers clicking the like or comment button after some time of watching the video, rather than clicking the comment button as soon as you click the paly button.

C. CLASSIFICATION ALGORITHM OF MALICIOUS SOCIAL BOTS

Real-time detection of malicious social bots in online social platforms can detect and block social bots in a timely manner. We propose the detection method of malicious social bots based on semi-supervised clustering method, which can reduce the time of artificial marking, and the detection program can run periodically in the background of the website. Simultaneously, we choose the hybrid feature of transition probability features and time feature can be used to increase the robustness of the features, thus improving the accuracy of detection. In the meantime, the user's transition probability features and inter-arrival times can be obtained. We can analysis user behavior and social bots behavior based on features of temporal and spatial dimensions. Based on the constrained seed K-means algorithm [24], we set the sample mean square error threshold to determine the number of iterations, then obtain the social bots detection algorithm. The detection algorithm for malicious social bots is described in Algorithm 1.

IV. EXPERIMENT AND RESULT ANALYSIS

A. DATA COLLECTION

The experimental platform in this study is the online media social network platform CyVOD [25]. CyVOD is an Internet plus technology information service application platform that integrates science and technology policy, scientific and technological achievements, and technology and social interaction. 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). The acquisition of user's action in our own website is shown in Figure 1. You can pass the corresponding buried data to the server and record it in the server through the code when

Algorithm 1 The Detection Algorithm for Malicious Social Bots

```
Input: The log set of users' click event: $DS$, cluster number $k = 2$, a small number of labeled samples: $S = \bigcup_{j=1}^{k} S_j$, global threshold $\tau$
Output: Normal user set, Social bots set
SocialBotsDection ($DS$, $S$)
Begin
for $s \leftarrow 1$ to n // $s$ refers to the users id
    $C_s = \{I(1), I(2), \ldots, I(m)\}$ // generate the user's intent sequence sets $Cs$
    According to formula (1), Calculate the transition probability $P_{(play,like)}$, $P_{(play,feedback)}$, $P_{(play,comment)}$, $P_{(play,share)}$, $P_{(play,more)}$, $P_{(play,paly)}$
    $\mathrm{IAT}(s) = \frac{\sum T((t-1)=play)-T(t)}{N(play)}$ // calculate the inter-arrival times
    $x_s = \{P_{(play,like)}, P_{(play,feedback)}, P_{(play,comment)}, P_{(play,share)}, P_{(play,more)}, P_{(play,paly)}, \mathrm{IAT}(s)\}$ // generate the sets of transition probabilities and time feature
endfor
for $j \leftarrow 1$ to k
    $\mu_j = \frac{1}{S_j} \sum_{x \in S_j} x_s$ // initialize the cluster center
endfor
repeat
    $C_j = \varnothing \ (1 \le j \le k)$
for $j = 1, 2$
    for all $x \in S_j$
        $C_j = C_j \cup \{x_s\}$
    endfor
endfor
for all $x_s \in D \backslash S$
    Calculate the distance from the sample $x_s$ to mean vectors $\mu_s (1 \le j \le k)$:   $d_{sj} = \|x_s - \mu_s\|_2$
    Find out the nearest cluster to sample $x_s$:
        $r = \arg\min_{j \in \{1,2\}} d_{sj}$
    $C_r = C_r \cup \{x_s\}$ // divide sample $x_s$ into the
        corresponding cluster
    $M = \frac{1}{n} \sum_{s=1}^{n} (x_s - \mu_s)^2$ // calculate mean square error
    $M_t = M$
endfor
until $M - M_t < \tau$
End
```

you manipulate some controls of the UI layer. In the real social network platform, many platforms use the burying technology to obtain the user's behavior data. In the research, many scholars choose to cooperate with the social platform or call the corresponding API of the social platform to obtain data.

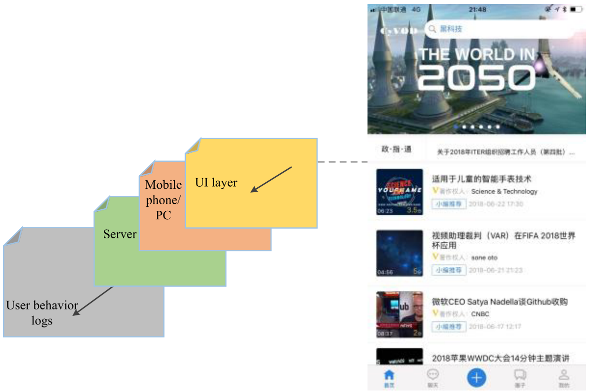

User behavior logs
Server
Mobile phone/ PC
UI layer
中国联通 4G
21:48
CyVOD
黑科技
THE WORLD IN 2050
政·指·通
关于2018年ITER组织招聘工作人员（第四批）...
适用于儿童的智能手表技术
Science & Technology
小编推荐 2018-06-22 17:30
视频助理裁判（VAR）在FIFA 2018世界杯应用
sone oto
小编推荐 2018-06-21 21:23
微软CEO Satya Nadella谈Github收购
CNBC
小编推荐 2018-06-17 12:17
2018苹果WWDC大会14分钟主题演讲

FIGURE 1. Acquisition process of user clickstream behavior.

TABLE 1. Classification of malicious social bots.

Categories	Tasks
Social bots for batch registration	Batch registration of accounts
Social bots for malicious likes	Perform video like tasks
Social bots for malicious comments	Perform video comment tasks
Social bots for malicious reports	Perform video report tasks
Social bots for malicious acquisition of user information	Acquire much user information
Mixed social bots 1	Perform one or more of four special tasks as required
Mixed social bots 2	Add friends, imitate user friendship

B. EXPERIMENTAL DESIGN

A total of 1500 malicious social bots accounts on the CyVOD platform are assigned different tasks, including malicious 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. The mixed social bot, the values of six transition probability features maybe average, which looks like normal user. In this experiment, four malicious social bots that perform different specific purposes and two malicious social bots with mixed behavior are set up. Malicious social bots are classified as shown in Table 1. We designed an Android application called "SocialBot" to simulate the behavior of social bots. According to the functional characteristics of the experimental platform, we designed such seven categories of social bot that perform different tasks. The social bot program can be activated automatically or manually by clicking these buttons.

In this paper, a total of 450 thousand items of data were collected from July 1 to September 30, 2018. These data were clickstream data of normal users and social bots on CyVOD. Based on the corresponding functions provided by CyVOD platform, 46 click events with 4 categories of user behavior features were recorded, including user information viewing, video broadcasting, comment related behaviors, friend related behaviors, comment releasing in circles, and other related behaviors.

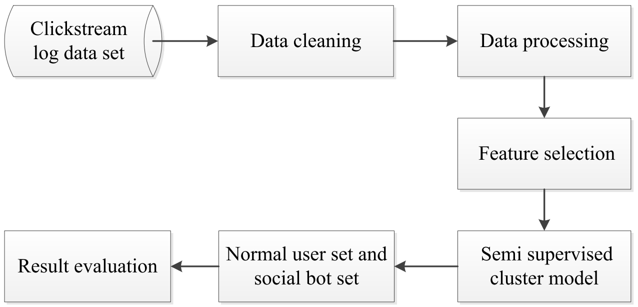


FIGURE 2. Experiment procedure.

C. MALICIOUS SOCIAL BOTS DETECTION

Data set cleaning and screening, data feature processing, data classification, and a series of operations were conducted after acquiring clickstream data set of the user. The detailed steps are shown in Figure 2.

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 playback 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 F_1 Score (F_1 is the harmonic average of Precision and Recall, $F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$). In the meantime, we use Accuracy as a 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.

D. FINDINGS

In this paper, the corresponding user data from the CyVOD web platform and Android client are collected. The iOS client is discarded because of its lower amount of data compared with the other two. To protect user privacy, users are assigned a unique and anonymous id. The main function of the platform is audio and video playback. The experimental features are selected around the main features of the platform. The accuracy of the detection method for malicious social bots proposed is compared and analyzed by acquiring the corresponding click times of different categories of malicious social bots and transition probability features between clickstreams. In this paper, three categories of features are selected to train the proposed semi-supervised detection method and detect its classification. The comparison of three categories of features is shown in Table 2.

TABLE 2. Comparison of three categories of characteristics.

Categories	Corresponding Characteristics
Quantitative feature	Times of play, number of comments, point of praise, number of report, and times of sharing
Transition probability feature	$P_{(play,\ play)}$, $P_{(play,\ \mathrm{like})}$, $P_{(play,\ feedback)}$, $P_{(play,\ comment)}$, $P_{(play,\ share)}$, $P_{(play,\ more)}$
Mixed feature	$P_{(play,\ play)}$, $P_{(play,\ \mathrm{like})}$, $P_{(play,\ feedback)}$, $P_{(play,\ comment)}$, $P_{(play,\ share)}$, $P_{(play,\ more)}$, IATs (inter-arrival times)

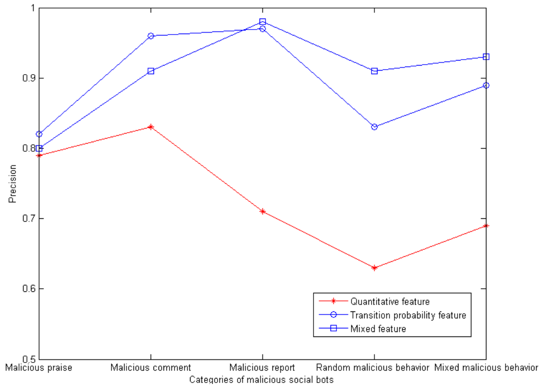


FIGURE 3. Precision of detection methods based on different features for different types of malicious social bots.

To verify the effectiveness of the proposed features, different types of malicious social bots are modeled using three categories of features to obtain precise detection. The precision of detection for different types of malicious social bots by using three categories of features is shown in Figure 3. The recall of detection for different types of malicious social bots by using three categories of features is shown in Figure 4. We find that (1) the precision of the semi-supervised clustering method for the detection of the same type of malicious social bots based on transition the probability features and mixed features is higher than that of the semi-supervised clustering method based on the quantitative feature; (2) for simple malicious social bots, the application of transition probability feature and mixed feature can effectively detect malicious social bots accounts. However, for malicious social

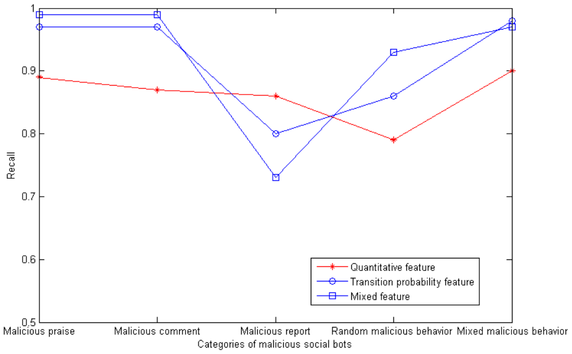


FIGURE 4. **Recall of detection methods based on different features for different types of malicious social bots.**

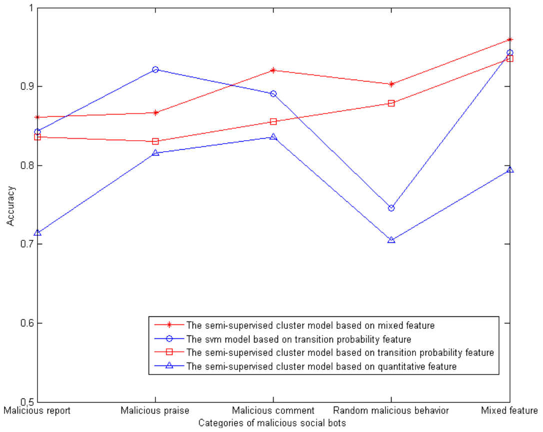
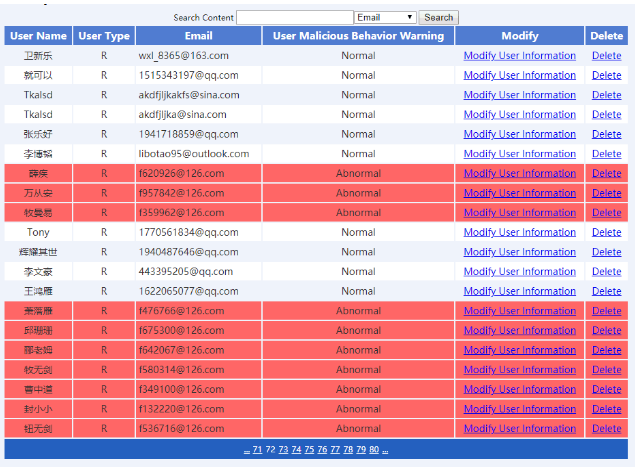


FIGURE 5. **Detection results of malicious social bots using different methods.**

bots with mixed feature and random behavior, the application of mixed feature can obtain better results. The experimental result shows that the precision of a semi-supervised clustering method based on mixed features for detecting malicious social bots with mixed malicious feature can be as high as 93.1%, the recall rate is 97.5%, and the F_1 Score is 95.2%. Compared with the semi-supervised clustering method with quantitative features, our method can detect malicious social bots accounts in online social platforms more accurately.

To verify the accuracy of the method, the support vector machine model based on transition probability, the semi-supervised clustering method based on mixed feature, the semi-supervised clustering method based on transition probability, and the semi-supervised clustering method based on quantitative feature are established in the same data set. The detection accuracy of different methods for malicious social bots are shown in Figure 5. The experiment proves that the proposed semi-supervised clustering method based on transition probability between user clickstreams can effectively detect malicious social bots in online social platforms. The comparison between different methods shows that the precision and accuracy of the detection method of malicious social bots based on transition probability can reach 95% or higher. Compared with the traditional detection method based on the quantitative feature, accuracy is improved by 12.8% on average. The method can effectively detect malicious accounts on social platforms. Finally, the malicious social bots detection program was deployed and run on the CyVOD platform. In the background user information list, malicious accounts of social bots are marked in red for convenience in addressing malicious social bots. Malicious social bots are automatically labeled on the website as shown in Figure 6.



Search Content [ ] Email ▾ Search

User Name	User Type	Email	User Malicious Behavior Warning	Modify	Delete
卫新乐	R	wxl_8365@163.com	Normal	Modify User Information	Delete
就可以	R	1515343197@qq.com	Normal	Modify User Information	Delete
Tkalsd	R	akdfjljkakfs@sina.com	Normal	Modify User Information	Delete
Tkalsd	R	akdfjljka@sina.com	Normal	Modify User Information	Delete
张乐好	R	1941718859@qq.com	Normal	Modify User Information	Delete
李博韬	R	libotao95@outlook.com	Normal	Modify User Information	Delete
薛疾	R	f620926@126.com	Abnormal	Modify User Information	Delete
万从安	R	f957842@126.com	Abnormal	Modify User Information	Delete
牧曼易	R	f359962@126.com	Abnormal	Modify User Information	Delete
Tony	R	1770561834@qq.com	Normal	Modify User Information	Delete
辉耀其世	R	1940487646@qq.com	Normal	Modify User Information	Delete
李文豪	R	443395205@qq.com	Normal	Modify User Information	Delete
王鸿雁	R	1622065077@qq.com	Normal	Modify User Information	Delete
萧落雁	R	f476766@126.com	Abnormal	Modify User Information	Delete
邱珊珊	R	f675300@126.com	Abnormal	Modify User Information	Delete
郦老姆	R	f642067@126.com	Abnormal	Modify User Information	Delete
牧无剑	R	f580314@126.com	Abnormal	Modify User Information	Delete
曹中道	R	f349100@126.com	Abnormal	Modify User Information	Delete
封小小	R	f132220@126.com	Abnormal	Modify User Information	Delete
钮无剑	R	f536716@126.com	Abnormal	Modify User Information	Delete

... 71 72 73 74 75 76 77 78 79 80 ...

FIGURE 6. **Partial detection results of malicious social bots.**

V. 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.

ACKNOWLEDGMENT

The authors would like to thank the reviewers and editor for their valuable comments, questions and suggestions.

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 network Websites 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.

[7] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," *Commun. ACM*, vol. 59, no. 7, pp. 96–104, Jul. 2016.

[8] T. Hwang, I. Pearce, and M. Nanis, "Socialbots: Voices from the fronts," *Interactions*, vol. 19, no. 2, pp. 38–45, Mar. 2012.

[9] Y. Zhou *et al.*, "*ProGuard*: Detecting malicious accounts in social-network-based online promotions," *IEEE Access*, vol. 5, pp. 1990–1999, 2017.

[10] Z. Zhang, C. Li, B. B. Gupta, and D. Niu, "Efficient compressed ciphertext length scheme using multi-authority CP-ABE for hierarchical attributes," *IEEE Access*, vol. 6, pp. 38273–38284, 2018. doi: 10.1109/ACCESS.2018.2854600.

[11] C. Cai, L. Li, and D. Zengi, "Behavior enhanced deep bot detection in social media," in *Proc. IEEE Int. Conf. Intell. Secur. Inform. (ISI)*, Beijing, China, Jul. 2017, pp. 128–130.

[12] C. K. Chang, "Situation analytics: A foundation for a new software engineering paradigm," *Computer*, vol. 49, no. 1, pp. 24–33, Jan. 2016.

[13] Z. Zhang, R. Sun, X. Wang, and C. Zhao, "A situational analytic method for user behavior pattern in multimedia social networks," *IEEE Trans. Big Data*, to be published. doi: 10.1109/TBDATA.2017.2657623.

[14] S. Barbon, Jr., G. F. C. Campos, G. M. Tavares, R. A. Igawa, M. L. Proença, Jr., and R. C. Guido, "Detection of human, legitimate bot, and malicious bot in online social networks based on wavelets," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 14, no. 1s, Feb. 2018, Art. no. 26.

[15] J. Y. Park, N. O'Hare, R. Schifanella, A. Jaimes, and C.-W. Chung, "A large-scale study of user image search behavior on the Web," in *Proc. 33rd Annu. ACM Conf. Hum. Factors Comput. Syst.*, Seoul, South Korea, 2015, pp. 985–994.

[16] 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.

[17] 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.

[18] 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.

[19] J. Liu, Y. Xiao, K. Ghaboosi, H. Deng, and J. Zhang, "Botnet: Classification, attacks, detection, tracing, and preventive measures," *EURASIP J. Wireless Commun. Netw.*, vol. 2009, Dec. 2009, Art. no. 692654. doi: 10.1155/2009/692654.

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

[21] E. Van Der Walt and J. Eloff, "Using machine learning to detect fake identities: Bots vs humans," *IEEE Access*, vol. 6, pp. 6540–6549, Jan. 2018.

[22] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, "BotOrNot: A system to evaluate social bots," in *Proc. 25th Int. Conf. Companion World Wide Web*, Montreal, Canada, 2016, pp. 273–274.

[23] M. Fazil and M. Abulaish, "Identifying active, reactive, and inactive targets of socialbots in Twitter," in *Proc. Int. Conf. Web Intell.*, Leipzig, Germany, 2017, pp. 573–580.

[24] A. F. Costa, Y. Yamaguchi, A. J. M. Traina, C. Traina, Jr., and C. Faloutsos, "Modeling temporal activity to detect anomalous behavior in social media," *ACM Trans. Knowl. Discovery Data*, vol. 11, no. 4, Aug. 2017, Art. no. 49.

[25] S. Basu, A. Banerjee, and R. Mooney, "Semi-supervised clustering by seeding," in *Proc. 19th Int. Conf. Mach. Learn.*, Sydney, NSW, Australia, 2002, pp. 19–26.

[26] Z. Zhang, R. Sun, C. Zhao, J. Wang, C. K. Chang, and B. B. Gupta, "CyVOD: A novel trinity multimedia social network scheme," *Multimedia Tools Appl.*, vol. 76, no. 18, pp. 18513–18529, Sep. 2017.

PEINING SHI is currently pursuing the Ph.D. degree, majoring in computer science, with the College of Information Engineering, Henan University of Science and Technology. Her research interests focus on social situation analytics and social user behaviors security.

ZHIYONG ZHANG (M'06–SM'11) received the master's degree from the Dalian University of Technology, and the Ph.D. degree in computer science from Xidian University, China. He is a post-doctoral fellowship at School of Management, Xi'an Jiaotong University, China. He is currently a full-time Henan Province Distinguished Professor, the Director of the Henan Joint International Research Laboratory of Cyberspace Security Applications, and the Dean of the Department of Computer Science, Information Engineering College, Henan University of Science and Technology. He is also a Visiting Professor with the Computer Science Department, Iowa State University. His research interests include multimedia social networks, digital rights management, trusted computing, and usage control. In recent years, he has published over 80 scientific papers and has edited four books in the above research fields. He holds eight authorized patents. He is a Senior Member of the ACM, the Chair of the IEEE Multimedia Communications Technical Committee DRMIG, a member of the IEEE Systems, Man, Cybermetics Society Technical Committee on Soft Computing, a member of the World Federation on Soft Computing Young Researchers Committee, and a member of the Digital Rights Management Technical Specialist Workgroup Attached to China National Audio, Video, Multimedia System, and Device Standardization Technologies Committee. He is the Chair/Co-Chair and TPC member for numerous international conferences/workshops on digital rights management and cloud computing security. Moreover, he is an Editorial Board Member and an Associate Editor of the IEEE Access, the *Multimedia Tools and Applications*, the *Neural Network World*, the *EURASIP Journal on Information Security* (Springer), and a leading Guest Editor or Co-Guest Editor of the *Computer Journal* and the *Applied Soft Computing and Future Generation Computer Systems*.

KIM-KWANG RAYMOND CHOO received the Ph.D. degree in information security from the Queensland University of Technology, Australia, in 2006. He currently holds the Cloud Technology Endowed Professorship at The University of Texas at San Antonio. He is also a Fellow of the Australian Computer Society. In 2016, he was named the Cybersecurity Educator of the Year – APAC (Cybersecurity Excellence Awards are produced in cooperation with the Information Security Community on LinkedIn). In 2015, he and his team won the Digital Forensics Research Challenge organized by the University of Erlangen-Nuremberg, Germany. He is a recipient of the 2018 UTSA College of Business Col. Jean Piccione and Lt. Col. Philip Piccione Endowed Research Award for Tenured Faculty, the IEEE TrustCom 2018 Best Paper Award, the ESORICS 2015 Best Paper Award, the 2014 Highly Commended Award by the Australia New Zealand Policing Advisory Agency, the Fulbright Scholarship, in 2009, the 2008 Australia Day Achievement Medallion, and the British Computer Society's Wilkes Award, in 2008.