Hindawi
Security and Communication Networks
Volume 2021, Article ID 6653430, 14 pages
https://doi.org/10.1155/2021/6653430

*Research Article*

# Detecting Abnormal Social Network Accounts with Hurst of Interest Distribution

**Xiujuan Wang, Yi Sui, Yuanrui Tao, Qianqian Zhang, and Jianhua Wei**

*Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China*

Correspondence should be addressed to Yi Sui; 17864307856@163.com

Received 8 December 2020; Revised 19 May 2021; Accepted 27 May 2021; Published 9 June 2021

Academic Editor: Manjit Kaur



With the rapid development of the Internet since the beginning of the 21st century, social networks have provided a significant amount of convenience for work, study, and entertainment. Specifically, because of the irreplaceable superiority of social platforms in disseminating information, criminals have thus updated the main methods of social engineering attacks. Detecting abnormal accounts on social networks in a timely manner can effectively prevent the occurrence of malicious Internet events. Different from previous research work, in this work, a method of anomaly detection called Hurst of Interest Distribution is proposed based on the stability of user interest quantifiable from the content of users' tweets, so as to detect abnormal accounts. In detail, the Latent Dirichlet Allocation model is adopted to classify blog content on Twitter into topics to calculate and obtain the topic distribution of tweets sent by a single user within a period of time. Then, the stability degree of the user's tweet topic preference is calculated according to the Hurst index to determine whether the account is compromised. Through experiments, the Hurst indexes of normal and abnormal accounts are found to be significantly different, and the detection rate of abnormal accounts using the proposed method can reach up to 97.93%.

## 1. Introduction

With the rapid development and popularization of the Internet, people have become increasingly more interested in using social media. Online social platforms enable people to share their daily lives, express emotions, and access global news hotspots without leaving home. Owing to the convenience offered by such platforms, the number of social network users has increased dramatically. Statistics show that more than 1.3 million new users on average joined social media platforms every day during 2020, with nearly half a billion new users taking the global user total to almost 4.2 billion by the start of 2021 [1]. However, with the complexity and diversification of social networks, several problems have emerged. The terabyte (TB) level user data and high user traffic generated by social platforms present criminals with opportunities.

After manipulating social robots to hijack real accounts, criminals steal personal privacy data or implement Telecom fraud by spreading malicious links and sending spam and phishing e-mails. According to statistics, the share of spam in e-mail traffic amounted to 50.37% in 2020 [2]. Even social robots make negative comments or post false news on specific topics to control the direction of public opinion and affect social stability or political elections [3, 4]. In Ref. [5], it is pointed out that 9%–15% of active Twitter accounts are robots, so it is a challenging task to detect such robots. In addition, it is of great significance to create a green and safe network environment, protect users' privacy and security, and maintain political and social stability.

In previous cases of compromised accounts on social platforms, account thieves usually made a large number of repeated statements through the software to achieve dissemination. It was easy to distinguish whether the account had been stolen based on the content and language features. Today, the behaviors of social robots are more similar to those of real human beings [6, 7]. Therefore, it is no longer effective to determine whether an account has been compromised by robots simply based on the virtue of the grammatical and semantic features of the content.

As an increasing amount of behavior patterns of users on social platforms is quantified as features for abnormal detection, results have been achieved in the early stage of applications. At the same time, the information acquisition of criminals has gradually become symmetrical. Through the learning and imitation of normal user characteristics, a number of accounts that are difficult to distinguish between true and false are gradually constructed to evade abnormal detection.

To solve the aforementioned problems, it is argued herein that the psychological characteristics of human beings are not easy to change rapidly and are difficult to imitate, which can effectively help detect compromised accounts. The characteristics of human beings mainly refer to the uniqueness of the user, such as their personality, hobbies, and emotional tendencies. It takes a long time for one person to get to know another person intimately in daily communication [8]. It is even harder for an attacker to mimic a person's human characteristics through a single online message. If one can grasp the changes in user characteristics of human beings, it will contribute to a new approach of achieving abnormal detection. It is considered that most accounts with stable features are normal accounts, those with disordered features are robot accounts, and those with features that change at some point in time are compromised.

The main difference between our work and existing models lies in the detection based on the judgement of stability of user interest distribution. The contributions of this paper are described in detail as follows.

First, a new feature for anomaly detection is introduced. According to the individual differences of users and the stability of psychological characteristics, the user interest distribution is extracted from the content of users' tweets. Compared with the features used in previous studies, the process of extracting user interest distribution is easier to carry out, and the features that can reflect individual differences are not easy to imitate and will not mutate.

Then, an interest presentation algorithm with the Latent Dirichlet Allocation (LDA) model is proposed. There are too many points of the characteristics of human beings to investigate, and it is even more difficult to describe and quantify them. In the work described in this paper, hobbies, one of the characteristics of human beings, were chosen to detect compromised accounts since they can be easily quantified.

Finally, a compromised-account-detection algorithm is proposed in this paper. The Hurst of Interest Distribution (HoID) is introduced to measure the stability of user hobbies. Stability refers to whether the changing trend of user preferences is within an acceptable range, while the status update of abnormal accounts hijacked by robots is random, which is inconsistent with the previous psychological characteristics. Therefore, the stability of the distribution of hobbies of social accounts is used to identify the existence of abnormal accounts.

The rest of this paper is organized as follows. In Section 2, the related work is divided into two parts: the first part is abnormal-account-detection methods based on user characteristics, and the second part is characteristics of mining of human beings based on LDA. In Section 3, the detection method based on the HoID algorithm is elaborated. In Section 4, the experiments and corresponding analysis are described. Conclusions are drawn in Section 5.

## 2. Related Work

With the rapid development of the Internet era, social networks provide a significant number of conveniences for work, study, and entertainment but also bring various information-security problems. Cyberspace threats emerge endlessly and cause huge losses to Internet users. In the field of information security, to prevent illegal attacks and protect the security of private data in the process of transmission and storage, information-encryption technology is constantly improving and great progress has been made.

In recent years, many image-encryption approaches have been proposed on the basis of chaotic maps, in which it is very crucial to assign value to chaotic map parameters. The existing solutions are based on metaheuristics, which have the problems of slow computing speed and falling into local optima. Aiming to resolve this issue, Kaur et al. proposed a strength Pareto evolutionary algorithm-II-based metaheuristic approach to tune the hyperparameters of a four-dimensional chaotic map [9]. Comparative analyses showed that the proposed approach outperformed the competitive approaches in terms of entropy. Furthermore, dual local-search-based multiobjective optimization (DLS-MO) was used to obtain the optimal parameters of a hyperchaotic map and encryption factors in another study, which also achieved good performance [10]. In addition, the parameter estimation of hyperchaotic maps involves extensive computational search. Kuar et al. proposed a minimax differential evolution-based seven-dimensional hyperchaotic map to generate secret keys for image encryption [11]. The fitness of the parameters was evaluated using correlation coefficient and entropy. The proposed approach achieved significantly good encryption results compared to the competitive approaches. In addition, the proposed approach resisted various security attacks.

Although encryption technology ensures the privacy and security of information to a large extent, in view of the openness and sharing concept of cyberspace, in addition to the application of encryption technology, one should also combine an anomaly-detection algorithm to further purify the network environment and enhance network security. Detecting abnormal accounts on social networks in a timely manner can effectively prevent the occurrence of malicious Internet events. There are many studies on abnormal account detection on social platforms, and the LDA model has also been applied to the field of abnormal account detection.

The detection of social robots must process a pre-collected dataset and then select some representative and distinguishing features from the content information, behavior information, and social relationship graph. Finally, a supervised-machine-learning algorithm is used to classify the features to obtain a more accurate detection effect [12]. Earlier researchers include Wang [13], who extracted graph-

and content-based features and designed an algorithm to detect spam robots in Twitter. Efthimion et al. [14] proposed a new machine-learning algorithm that utilized a series of features, including the length of user name, time pattern, emotional expression, and the ratio of followers to friends. Logistic regression (LR) as a classifier could effectively detect robots with an error rate of 2.25%.

In recent years, with the development of big data and the improvement of computer performance, deep learning has gradually become popular. Cai et al. proposed a behavior-enhanced deep model (BeDM) for bot detection [15]. BeDM fused content information and behavior information and regarded user content as temporal text data to extract latent temporal patterns. They combined convolutional neural networks (CNNs) with a long short-term memory (LSTM) model, and the garbage robot of tweets was detected efficiently and accurately. Sneha et al. proposed a deep neural network based on the LSTM model that extracted context features from user metadata that were fed as auxiliary input into a LSTM deep network to process tweet text [16]. In addition, a technique based on synthetic minority oversampling (called SMOTE) was proposed to generate a large-scale labeled dataset suitable for deep network training from the minimum number of labeled data. Experiments showed that this structure could achieve high classification accuracy (AUC > 96%, where AUC denotes area under the curve) in the process of detecting bots through only one tweet.

*2.1. Abnormal-Account-Detection Methods Based on User Characteristics.* Feature representation is commonly adopted to detect abnormal accounts. Because there are significant differences between abnormal and normal accounts in some characteristics, the accuracy of account classification can be effectively improved by selecting features with a large degree of differences. The existing feature representation is mainly divided into features including attribute features, content features, network features, and activity features.

Attribute features include user name, avatar, number of followers, and other basic information, which are easily obtained. The user's age, educational background, e-mail address, emotional status, and other characteristics are also involved, which are not easily obtained due to the influence of user-privacy settings. Teams in the 2015 DARPA (Defense Advanced Research Projects Agency) Twitter Bot challenge used the users name, user avatar, geographical location, and other attributes to detect abnormal accounts in Twitter [17]. Results of these experiments indicated that a bot-detection system must be semisupervised, but all teams used human judgement to augment automated bot-identification processes. The credibility features of some social platforms (such as Twitter) can also be used for abnormal detection.

Content features refer to the features extracted from content information posted by users, which can be mainly divided into grammatical and semantic features [18, 19]. The semantic features refer to the subjects or emotions of the published content items, while the grammatical features refer to the features including sentence structure, word frequency statistics, and punctuation. Several special features are also used, such as the use of "#," "@," and "http://." Kumar et al. built an automatic classification system that used features such as text length and text composition ratio to detect abnormal users in Wikipedia [20]. Results of the experiments showed that the algorithms could utilize these additional signals rather than article appearance features to accurately identify hoaxes.

Network features mainly refer to the correlations between social users, which are quantified by scholars into indicators such as degree, clustering coefficient, and centrality. Most of the normal accounts have social circles, there are many friends who follow each other, and the number of followers and followees is relatively balanced, while abnormal accounts have great differences in the above aspects. Graph data are generally used to represent the feature and structure information of nodes. With the rise of deep learning, a large number of researchers have considered using deep-learning models to automatically model graph data, including graph embedding [21] and graph neural networks [22]. Kirill et al. used a graph-embedding model to extract node representations from social network user profiles and used different classifiers to classify features, such as Multilayer Perceptron (MLP), K-Nearest Neighbor (KNN), and Gradient Boosting (GB) [23]. In addition, a stacking-based ensemble was created, which not only extracted graph features but also utilized text features. Empirical evaluation proved the effectiveness of the proposed method for bot detection and showed that stacking of first-layer classifiers with graph-embedding features allowed boosting the best single-classifier scores by 1%–4% in AUC accuracy.

To better detect malicious accounts and social bots, Seyed Ali and others think that account classification should employ a feature set and social graph at the same time. Therefore, a detection model based on a graph CNN was proposed, which effectively gathered the features of a node neighborhood [24]. Experimental results showed the superiority of the method, which increased the AUC accuracy by 8%. Given the growing scale of social networks, it will consume a significant amount of computing resources to construct the Twitter graph structure based on the follower and friend relationships in social accounts.

Activity features refer to a user's behavior patterns, such as active time, frequency of information published, and common clients [3, 25, 26]. Xin et al. divided users' social behaviors into two categories: extroversive behavior features such as activity sequence and introversive behavior features such as request latency [27]. European distance is used to quantify the differences between the incoming clickstream and the behavior pattern represented by the behavioral profile, so as to identify whether the clickstream is from real users or abnormal users. This method is applicable to users that directly access Online Social Network (OSN) pages, but it is difficult to trace the behavior patterns of users who access OSNs solely through application programming interfaces. Wu et al. used the published information quantity matrix feature to detect abnormal users in Sina Weibo [28]. Yamak et al. used the time intervals between user registration and first posting to detect fake accounts in Wikipedia [29]. The results from several machine-learning algorithms were compared to show that new features and training data

enabled the detection of 99% of fake accounts, improving previous results from the literature.

In addition, some studies have combined the above features. Chavoshi et al. illustrated that the presence of highly synchronous cross-user activities revealed abnormalities and thus developed the DeBot system to identify bots in Twitter's network [30]. DeBot is an unsupervised method that calculates cross-user activity correlations to detect bots in a parameter-free fashion. Its evaluation showed that DeBot detected bots at a rate higher than the rate Twitter was suspending them.

*2.2. Characteristics of Mining of Human Beings Based on LDA.* LDA is a document theme model, through which the thematic tendency of an article can be obtained, and thus the expression of users' interests can be obtained. Some scholars have used a LDA model to carry out some studies related to the characteristics of human beings. Liu et al. proposed the probabilistic topic model (PT-LDA model) to predict personality characteristics under the framework of the five-factor model and considered that each topic not only has the multinomial distribution of words but also has the Gaussian distribution of personality, which provides a new method to reveal user behaviors in social networks [31]. Zhang et al. proposed the concept of GROUP-LDA, which integrates book-related information into the LDA model to describe the subject relevance among documents to accurately detect the book audience [32]. According to the evaluation results, it outperformed Latent Semantic Analysis (LSA), LDA, the author-topic model (ATM), and several other collaborative filtering methods in terms of precision, recall, $F1$-score, and mean average precision (MAP) for book-audience detection. Shinjee et al. combined the LDA theme model of television viewers with the LDA theme model of program descriptors to effectively improve the user-prediction accuracy of new TV programs [33]. Gao et al. proposed a mechanism called SECO-LDA to construct service co-occurrence (SECO) documents by studying the potential topic model in the history of service collaboration to extract potential SECO topics. The derived knowledge of these topics will help reveal the tendency of service composition, aid the understanding of the cooperation behaviors between services, and provide a better service recommendation [34]. Yan et al. considered that traditional search engines only collect documents containing keywords in the query without considering the real intention hidden by users. To solve this problem, a personalized retrieval algorithm based on query-intention recognition and a subject model is proposed. A LDA topic model is used to model the historical search data of users. When a new query appears, the underlying topic of the query is identified by the topic model of its user-history search, and the appropriate document is recommended [35].

## 3. Materials and Methods

In this paper, a method of detecting abnormal accounts on social platforms based on the judgement of stability of user interest distribution is proposed. The LDA model is adopted to calculate the distribution of users' interests based on body content items published by users, and the Hurst parameter is used to measure the stability of interest distribution. In detail, the aforementioned HoID algorithm detects compromised accounts on social platforms through four steps: sorting out user tweets, training the LDA model, obtaining user interest distribution, and determining whether the interest distribution is stable, as shown in Figures 1 and 2. Among them, $\Omega$ refers to user tweets, $L$ refers to the LDA model, $D$ refers to the interest distribution, $H$ refers to the stability of interest, and $L = f(\Omega)$, $D = g(\Omega, L)$, and $H = h(\mathrm{D})$.

*3.1. Sorting Out User Tweets.* The content items published by users on social platforms in a certain period of time were summarized and sorted out. Taking users as a unit, statistics on the word usage frequency of each tweet were generated, and then the tweets were transformed into the form of word vectors, with stop words removed. The Baidu English stop word list, which contains 891 stop words, was used in this work.

*3.2. Interest Distribution.* It is considered that the contents posted by users on social platforms are closely related to their interests, and the topic distribution of tweets can reflect the distribution of users' interests and hobbies. Consequently, the processed tweets are input into the LDA model for training, which is used to predict the topic of the blocking tweets to obtain the interest distribution of users.

An article can cover more than one topic, and the words in the article reflect the specific set of topics it covers. In the proposed method, each topic is taken as a probability distribution on the word, and the document is taken as a probability mix of those topics. If one has $N$ topics, the probability of the word $i$ in a given document can be written as

$$P(w_i) = \sum_{j=1}^{N} P(w_i|z_i = j)P(z_i = j), \tag{1}$$

where $z_i$ is the potential variable representing the topic of the $i$th word. $P(w_i|z_i = j)$ is the probability of word $w_i$ being under the $j$th topic. $P(z_i = j)$ represents the probability of selecting a word from the $j$th topic in the current document, which varies from document to document. The $j$th topic is represented as a polynomial distribution $\varphi^j_{w_i} = P(w_i|z_i = j)$ of $V$ words in the word list. The text is represented as a random mix of $\varphi^d_j = P(z_i = j)$ on $K$ implied topics. Thus, the probability of word $w$ "occurring" in text $d$ is

$$P(w|d) = \sum_{j=1}^{N} \varphi^j_{w_i} \cdot \theta^d_j. \tag{2}$$

The maximum-likelihood estimators $\alpha$ and $\beta$ of the maximum-likelihood function (equation (3)) are obtained by the expectation-maximization algorithm, and the parameter values of $\alpha$ and $\beta$ are estimated, so as to determine the LDA model:

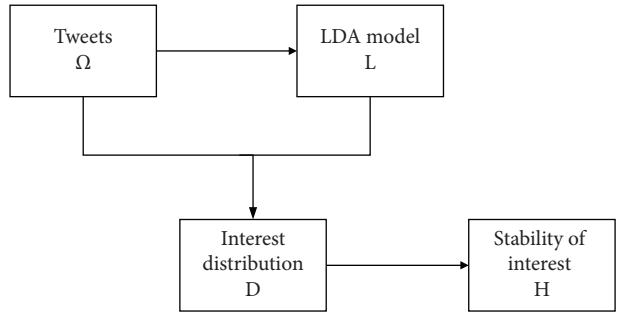


Figure 1: Module diagram of account interest mining.

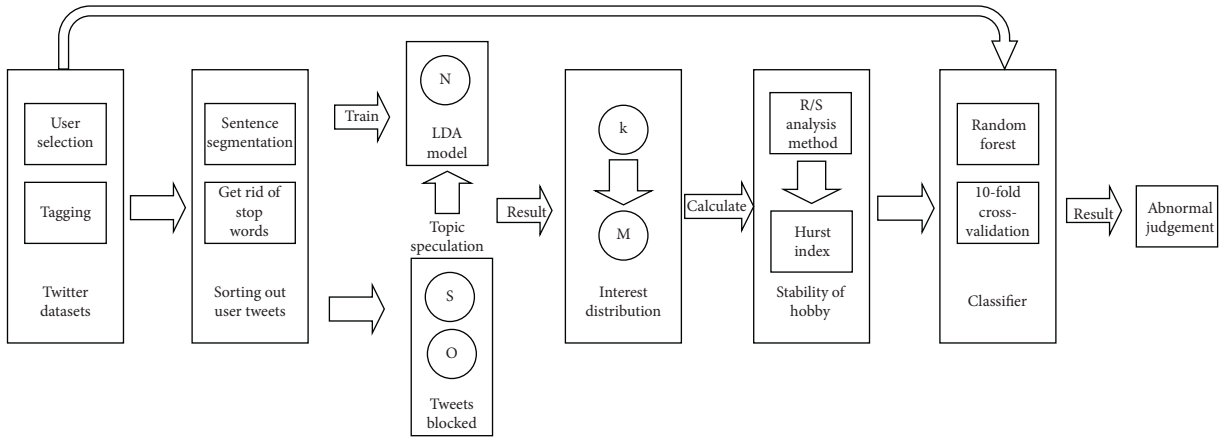


Figure 2: Account interest mining.

$$l(\alpha, \beta) = \sum_{i=1}^{M} \log\ p(d_i|\alpha, \beta), \tag{3}$$

where the conditional probability distribution of the "occurrence" of text $d$ can be obtained from

$$p(d|\alpha, \beta) = \frac{\Gamma(\sum_i a_i)}{\prod_i \Gamma(a_i)} \int \left( \prod_{i=1}^{k} \theta_i^{a_i - 1} \right) \left( \sum_{n=1}^{N} \sum_{i=1}^{k} \prod_{j=1}^{V} (\theta_i \beta_{ij})^{w_n^j} \right) d\theta. \tag{4}$$

Since the paired variables $\theta$ and $\beta$ exist in equation (4), the analytical equation cannot be calculated and an approximate solution is required. Griffiths et al. proposed that Gibbs sampling is better in terms of confusion and running speed. Gibbs sampling as proposed by Griffiths et al. is a better way in terms of perplexity and running speed.

Because of the 140-word limitation of a single tweet, the users' interests cannot be precisely reflected by so few words. To solve this problem, user tweets must be divided into blocks. Each tweet block includes user tweets in a certain period, so as to reflect the distribution of user interests and hobbies in that period. The concepts of tweet-block size $S$ and tweet-block overlap degree $O$ are introduced here. Tweet-block size refers to the number of tweets in a single tweet block, while tweet-block overlap degree refers to the overlap degree of two adjacent tweet blocks. Figure 3 shows an example of $S = 20$ and $O = 10$.

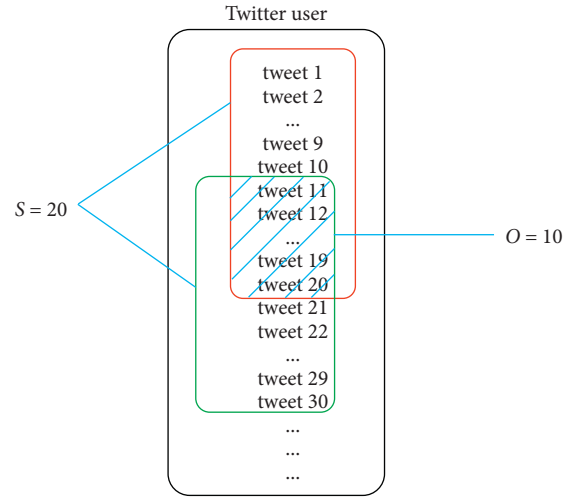


Figure 3: Example of user tweet separation when $S = 20$ and $O = 10$.

$T$ is defined as the total number of tweets of the user. Taking $K$ tweet blocks as an example, where $K = T/(S - O) + 1$, tweet block $k$ contains $((k-1) \cdot (S-O) + 1, (k-1) \cdot (S-O) + S)$ tweets. Those tweets in one block are jointed as one text.

The LDA model is used to calculate the topic distribution of $K$ tweet blocks. The topic distribution of the $k$th tweet block is represented as $\vec{k} = (P_{k1}, P_{k2}, P_{k3}, \ldots, P_{kN})$. So far, the interest distribution matrix $M$ of order $K * N$ has been obtained. Each row of the matrix represents the distribution

of the user tweets on $N$ topics in a certain period, and each column represents the variation of users' interest on a certain topic over time:

$$M = \begin{pmatrix} (P_{11}, P_{12}, P_{13}, \ldots, P_{1N}) \\ (P_{21}, P_{22}, P_{23}, \ldots, P_{2N}) \\ (P_{31}, P_{32}, P_{33}, \ldots, P_{3N}) \\ (P_{41}, P_{42}, P_{43}, \cdots, P_{4N}) \\ \cdots \\ \cdots \\ (P_{K1}, P_{K2}, P_{K3}, \ldots, P_{KN}) \end{pmatrix}. \quad (5)$$

*3.3. Stability of Interest.* Stability refers to whether the changing tendency of user preferences is within an acceptable range. Since the characteristics of human beings are not easy to imitate and not easy to change in a short time, it is considered here that the distribution of hobbies of a healthy account user should be stable. If there is a mutation, the account may be compromised. The method of analyzing the stability of a group of data is different in terms of application backgrounds. In this paper, the Rescaled Range Analysis (R/S analysis) method is used to calculate data stability.

R/S analysis is usually used to analyze the fractal characteristics of time series and the long-term memory process. It was originally proposed by British hydrologist Harold Edwin Hurst when he was studying the Nile Dam project. It was later used in the analysis of various time series.

In this study, the Hurst index is used to indicate the degree of stability of user interest. From the user-interest distribution matrix obtained above, the interest distribution sequence $\overrightarrow{n}$ of a user under the topic $n$ is shown in equation (6). It is divided into $[k/10]$ subintervals. For each subinterval, the cumulative dispersion $X_{t,l}$ is calculated according to equation (7), where $M_L$ is the average value of $P$ in the interval $l$:

$$\overrightarrow{n} = (P_{1n}, P_{2n}, P_{3n}, \ldots, P_{kn}), \quad (6)$$

$$X_{t,l} = \sum_{u=1}^{10} (x_u - M_L), \quad (7)$$

$$R = \max\left(X_{t,l}\right) - \min\left(X_{t,l}\right). \quad (8)$$

The fluctuation range $R$ is defined by equation (8) and is equal to the difference between the maximum and minimum values of the accumulated deviation. The standard deviation of the subinterval is denoted as $S$, and the rescaled range (R/S) is defined, which increases with increasing sequence length. Through a long period of practice, Hurst established the relationship as shown in in the following equation:

$$\frac{R}{S} = Kl^H. \quad (9)$$

Taking the logarithm of both sides of equation (9), one obtains

$$\log\left(\frac{R}{S}\right)_l = H \log(l) + \log(K). \quad (10)$$

The least-squares regression analysis of $\log(l)$ and $\log(R/S)_l$ in equation (10) can be used to calculate $H$, which is called the Hurst exponent. The corresponding Hurst exponent of the $n$th topic is denoted as $H_n$. So far, $N$ Hurst exponentials have been obtained, i.e., $H = (H_1, H_2, H_3, \ldots, H_n)$. $H$ represents the stability of the user's interests, and it is a group of characteristics of human beings that are affected by parameters $N$, $S$, and $O$ in this study. Classifiers can be used to complete classification work through traditional machine-learning methods.

## 4. Results and Discussion

In this section, the $H$ feature defined at the end of the preceding section is used to classify each user.

*4.1. Dataset.* There are no public datasets that were used in the related research on the detection of compromised accounts, so it is difficult to accurately determine whether an account has been compromised or not. The varol-2017 dataset used by Varol et al. [36] was selected in the present work. Varol et al. monitored approximately 10% of the public tweets in Twitter for a period of three months starting from October 2015 and selected users who sent at least 90 tweets during the three-month observation period and sent more than 200 tweets overall. This dataset has since been adopted by many researchers and offers tweets in terms of users. It is easy to observe a user's interests change by analyzing their tweets over time.

Our study uses Twitter official API interface Tweepy [37] to crawl user data, including user's published tweets and user metadata. This dataset comprised 940 original accounts. After data screening and cleaning, 616 users were selected as normal accounts. These original data are used to depict normal users and construct compromised accounts.

*4.2. Compromised-Account Construction.* A method proposed by Trang et al. in [38], which first obtained the data of normal users, then randomly paired 616 normal accounts several times, and exchanged some tweets in the paired accounts, was adopted to construct abnormal accounts. As shown in Figure 4, the top $m$ tweets of account $U_1$ itself were selected as the normal data before being hijacked. The $(m+1)$th tweet to the $N$th tweet is exchanged with the same part of the account $U_2$ matched with the user, as the abnormal data after being hijacked. Accordingly, the "compromised" accounts, $U_1^*$ and $U_2^*$, are constructed.

Based on the above method, the raw data were used to construct the compromised accounts using $N = 190$ and $m = 171$. Thus, 190 tweets from each of the 616 normal accounts were selected as experimental data. The first to 171st tweets served as original tweets, and the 172nd to 190th tweets were exchanged from paired accounts and served as the abnormal part. What is more, any two accounts in the normal accounts can be paired, and the pairing process is

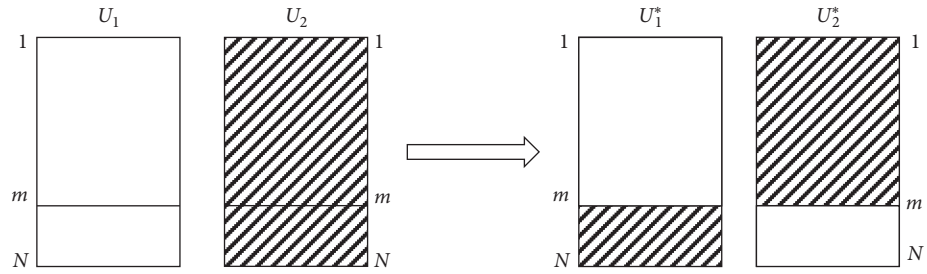


Figure 4: Compromised-account construction.

random and nonrepetitive. As a result, a total of 8,340 abnormal accounts were constructed.

*4.3. Data Preprocessing.* The above-described 940 normal accounts and 8,340 abnormal accounts exhibit a serious class-imbalance problem that will affect the correctness of the experimental results. Therefore, the SMOTE algorithm was used in this work to solve the class-imbalance problem [39].

SMOTE is an upsampling algorithm in which $M$ samples are randomly selected from the $k$-nearest neighbors of each few samples, where $M$ is the sampling multiplier. For each randomly selected neighbor $b$, a new sample $c$ is constructed with its original sample $a$ as follows:

$$c = a + \text{rand}(0, 1) * (b - a). \tag{11}$$

Since it is impossible to carry out SMOTE sampling on tweets, in this experiment, after calculating the $N$-dimensional Hurst value of the normal accounts, the above sampling was carried out on the tweets, and finally 8,340 normal and 8,340 abnormal accounts were constructed.

*4.4. Evaluation Metrics.* In this experiment, abnormal accounts were taken as positive samples. A random-forest classifier was used to conduct ten-fold cross-validation classification. Accuracy, precision, recall, and $F1$-measure were used as the evaluation indexes of the classification effect, calculated, respectively, as follows:

$$\begin{aligned} \text{precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{accuracy} &= \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}, \\ \text{recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}}, \\ F1 &= \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}. \end{aligned} \tag{12}$$

Here, TP refers to true positive, which equals the number that is correctly divided into a positive example. TN refers to true negative and is the number that is correctly divided into a negative example. FP refers to false positive and shows the number that is judged to be a positive sample of a negative sample. FN refers to false negative and indicates the number that is judged to be a negative sample of a positive sample.

*4.5. Experiment 1.* This experiment explored the influence of different topic numbers $N$ and overlap degrees $O$ on the classification effect when the tweet-block size was fixed to $S = 20$. The number of topics was taken as $N = 2, 3, \ldots, 10$, and the overlap degree of tweet blocks was taken as $O = 0, 1, \ldots, 9$. Thus, a total of 90 sets of classifications with different $N$ and $O$ values were involved.

Figures 5–8 depict the classification metrics with varying parameters. It can be seen from these figures that the changing tendency of different evaluation indexes is similar.

Longitudinal observation shows that when $O$ is constant, the accuracy/precision/recall/$F1$-measure is the lowest when $N = 2$, which indicates that interest distribution among two topics cannot accurately depict the interests of users. The performance becomes better when $N = 7, 8, 9, 10$. Almost each metric reaches the maximum when $N = 8$. This means that the interests of users distribute among approximately eight topics.

$O$ refers to the degree of overlap between two adjacent tweets. When the degree of overlap is high, the proportion of repeated content between two tweets is high, and the change of a user's topic tendency is small. In contrast, when the degree of overlap is low, the change of a user's topic tendency is large. The change of topic tendency will be reflected in the change of Hurst value and, in turn, will affect the experimental results.

$N$ refers to the number of topics. Different numbers of topics will lead to different distributions of user topic preference. When $N$ is small, the user's tendency to assign each topic is relatively large. When $N$ is large, the tendency to assign each topic is relatively small, which will also lead to the change of Hurst value and affect the experimental results.

Transverse observation shows that when $N$ is constant, with increasing $O$, each index bounces back at several nodes, but the overall change tendency is downward. This means that if the overlap between blocks is too large, the change of user interests is less obvious to be quantified, thus affecting the "judgement" of the HoID algorithm.

In abnormal detection, missing an abnormal account can be fatal. Therefore, recall is considered the most important index in the present work. Figures 9 and 10 show the average recall rates of $O = 1$–$9$ when $N$ is constant and the average recall rate of $N = 2$–$10$ when $O$ is constant.

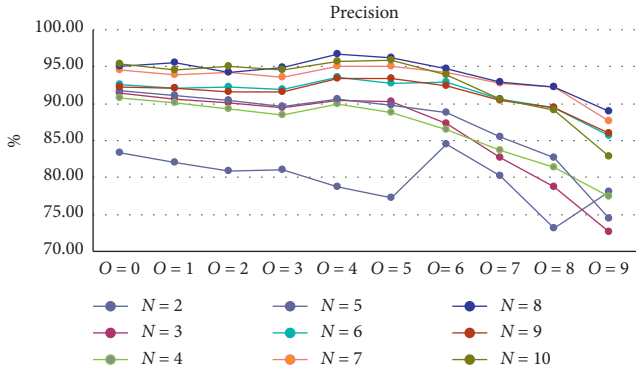


Figure 5: Precision for different $N$ and $O$ values.

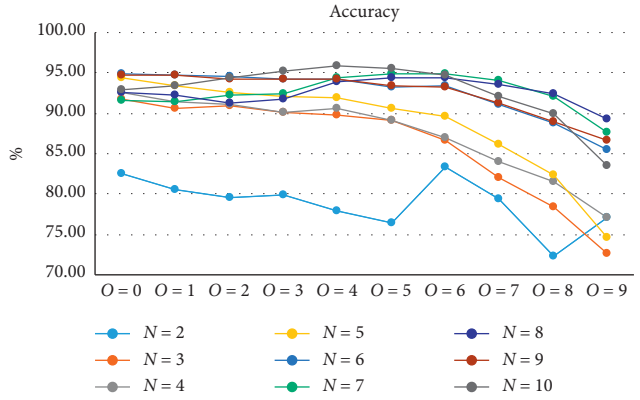


Figure 6: Accuracy for different $N$ and $O$ values.

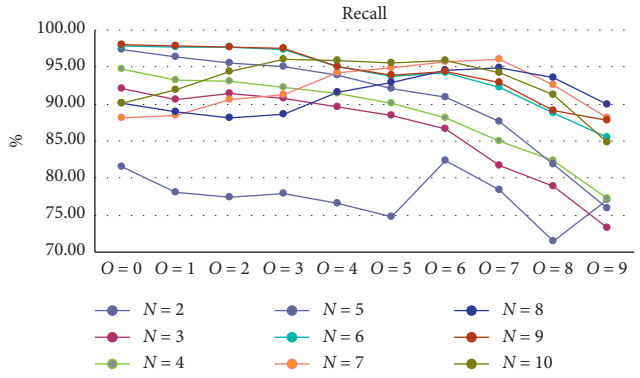


Figure 7: Recall for different $N$ and $O$ values.

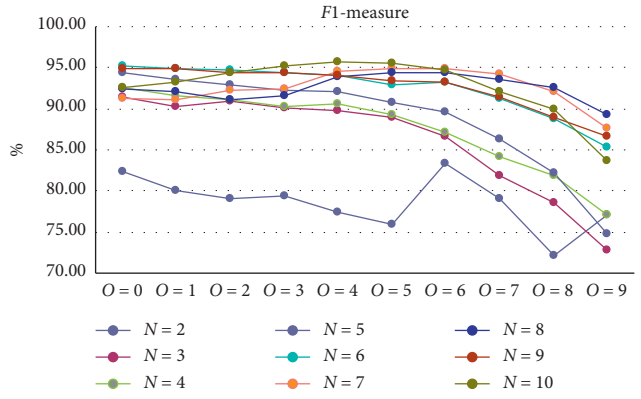


Figure 8: $F1$-measure for different $N$ and $O$ values.

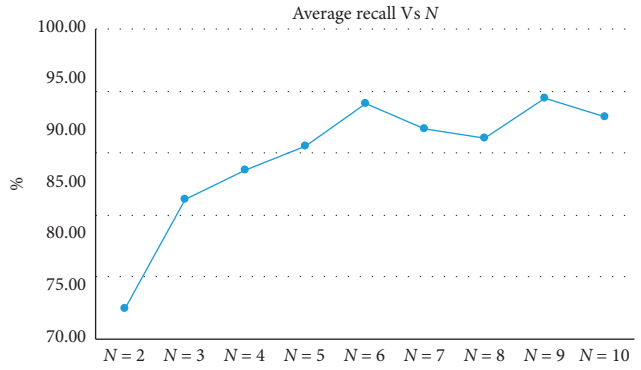


Figure 9: Average recall rate of $N = 2$–10 when $O$ is certain.

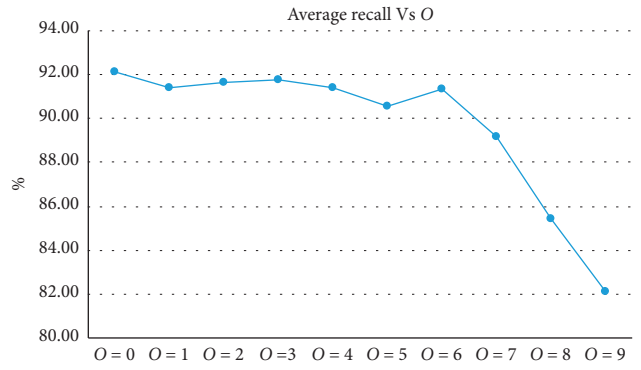


Figure 10: Average recall rate of $O = 2$–9 when $N$ is certain.

As can be seen from Figures 9 and 10, when $O$ is constant, the abnormal-account recall rate increases with increasing $N$, and when $N$ is constant, the abnormal-account recall rate shows a downward tendency with increasing $O$. This confirms the previous conclusion; that is, low topics cannot classify users' interests, and large overlap cannot reflect the tendency change of user interests. By comparing the data among groups, it can be concluded that, in this dataset, $N = 9$ and $O = 0$ are the best parameters to achieve a maximum recall rate when $S = 20$, with a precision of 92.12%, accuracy of 94.77%, recall of 97.93%, and $F1$-measure score of 94.93%. This group of parameters was used for the experiment.

*4.6. Experiment 2.* This experiment was designed to visually examine the Hurst-index distribution of normal and abnormal accounts.

Table 1 shows the differences between normal and abnormal accounts in the mean and variance of $H$.

The results in the table indicate that in the majority of cases, the $H$ mean of normal accounts is larger than that of abnormal accounts. This is consistent with the feature of the

Table 1: Mean and variance of $H$.

| Hurst index | $H$ mean | | $H$ variance | |
|---|---|---|---|---|
| | Normal | Abnormal | Normal | Abnormal |
| $H1$ | 1.037 | 0.919 | 0.469 | 0.285 |
| $H2$ | 1.105 | 0.831 | 0.319 | 0.469 |
| $H3$ | 0.445 | 0.684 | 0.989 | 0.670 |
| $H4$ | 1.075 | 0.905 | 0.346 | 0.328 |
| $H5$ | 0.979 | 0.898 | 0.544 | 0.327 |
| $H6$ | 0.841 | 0.825 | 0.744 | 0.515 |
| $H7$ | 1.094 | 0.967 | 0.360 | 0.207 |
| $H8$ | 0.990 | 0.926 | 0.556 | 0.293 |

Hurst exponent, namely, the larger the exponent, the greater the stability. The $H$ variance of normal accounts is also larger than that of abnormal accounts. The smaller difference in $H$ of abnormal accounts is due to the similar ways the abnormal accounts are constructed.

Limited by the difficulties in visualizing high-dimensional data, Figures 11 and 12 only illustrate the two-dimensional joint distribution in normal and abnormal accounts, taking $H1$ and $H2$ as examples. It can be seen that the $H1$ and $H2$ values of most normal accounts are between 1.0 and 1.5 at the same time, while those of the abnormal accounts are between 0.8 and 1.2. The distribution of normal accounts is quite different from that of the abnormal ones, which proves the assumption of HoID, i.e., most accounts with stable features are normal accounts.

In order to verify the difference between abnormal and normal accounts, an independent samples $t$-test was adopted on the Hurst index of abnormal and normal accounts. The $P$ values are shown in Table 2.

All $P$ values are less than 0.05, which proves that there are significant differences between the Hurst indexes of normal and abnormal accounts.

*4.7. Experiment 3.* This experiment explored the influence of different tweet block sizes, $S = 5$, 10, 15, 20, and 25, on the classification effect when the number of topics was $N = 9$ and the overlap degree of tweet blocks was $O = 0$. The results are shown in Figure 13.

It can be seen from Figure 13 that the recall rate is the best at $S = 20$, and when $S = 15$ and 25, the comprehensive level of the four indexes is better, while when $S = 5$ and 10, all evaluation indicators are at a low level. It can be considered that the classification effect becomes better with increasing tweet block size $S$. It seems that 15 tweets in one block is the best way to judge user interest distribution in a certain timeframe.

Different classifiers were used to conduct ten-fold cross-validation classification. Results for accuracy, precision, recall, and $F1$-measure in classification by random forest (RF), support vector machine (SVM), and KNN are shown in Figure 14. The results show that KNN performs the best followed by RF, while SVM performs the worst.

*4.8. Contrast Experiment.* Egele et al. proposed COMPA, a method to detect compromised accounts on social networks

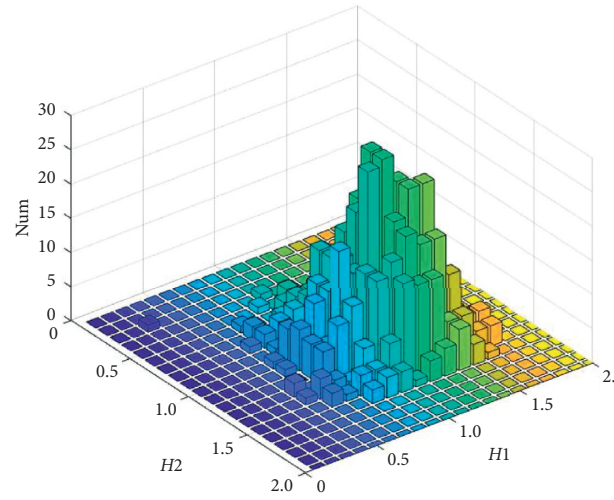


Figure 11: Normal-account joint distribution of $H1$ and $H2$.

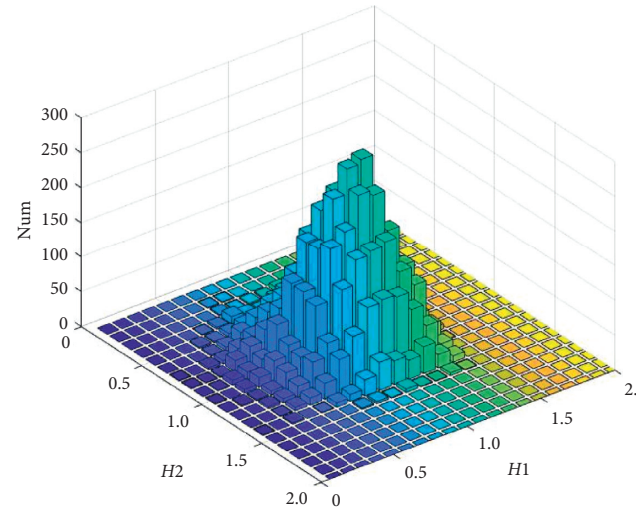


Figure 12: Abnormal-account joint distribution of $H1$ and $H2$.

[40]. The features selected by COMPA include terminal situation, user-mention situation, link-addition situation, time-point situation, language situation, and topic-participation situation. COMPA extracted features from the message flows published by users in chronological order and established a behavioral model to observe whether the new

Table 2: $T$-test results on normal and abnormal accounts.

| Hurst index | $P$ value |
|---|---|
| $H1$ | $4.075 \times 10^{-111}$ |
| $H2$ | 0.0 |
| $H3$ | $9.417 \times 10^{-78}$ |
| $H4$ | $1.090 \times 10^{-298}$ |
| $H5$ | $2.432 \times 10^{-34}$ |
| $H6$ | 0.025 |
| $H7$ | $3.136 \times 10^{-202}$ |
| $H8$ | $1.311 \times 10^{-30}$ |
| $H9$ | $2.603 \times 10^{-22}$ |

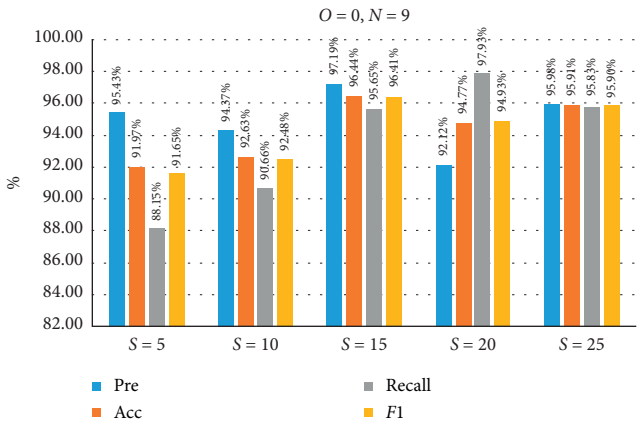


Figure 13: Results for $N = 9$, $O = 0$, and $S = 5$, 10, 15, 20, and 25.

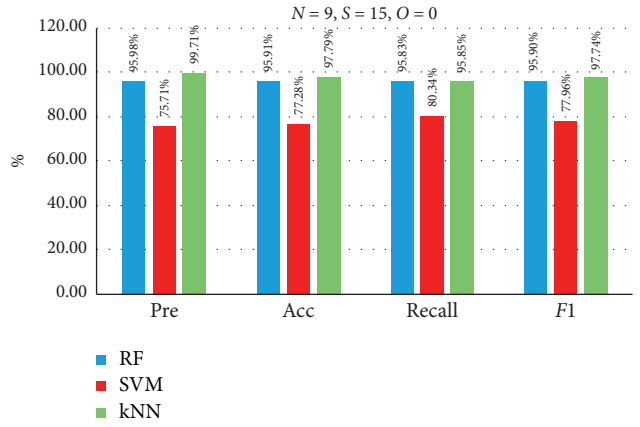


Figure 14: RF, SVM, and KNN classification results.

message flows conformed to the expected behavior and evaluate the abnormal score of each individual feature. Finally, the outlier scores of each feature were combined to obtain the global score of each message. The accuracy of the test results depended largely on the established behavioral profile and threshold value of selection.

Tang et al. proposed the supervised analytic hierarchy process (SAHP) for abnormal user detection [41]. In the process of abnormal user detection, different characteristics often reflect different degrees of user abnormality. Compared with COMPA, to establish more comprehensive profile features, SAHP took user expression habits into account and combined information gain rate with an analytic hierarchy process to ensure the accuracy of feature weight. SAHP then made detection decisions according to different thresholds. At high thresholds, the accuracy of the method is high, but it is slightly worse when a lower threshold is selected.

Kaur et al. utilized text-based continuous authentication (TB-CoAuth) for detecting compromised accounts in social networks [42]. Four categories of features, namely, content free, content specific, stylometric, and folksonomy, are extracted and evaluated by TB-CoAuth. In addition, various statistical and similarity-based feature-selection techniques are used to rank and select optimal features for each user, which are further combined using a popular rank-aggregation technique called Borda Count. Moreover, performance of various supervised-machine-learning classifiers is analyzed on the basis of different evaluation metrics.

HoID (parameters $S = 25$, $N = 9$, and $O = 0$), COMPA, and SAHP were tested separately on the same dataset. Furthermore, in the face of one of the most significant challenges in the domain of compromised accounts, i.e., the non-availability of ground-truth data consisting of the point of compromise and the compromised tweets, Kaur also used the artificial practice of creating ground-truth data to verify the model, manually injecting spam and randomness into the accounts. Therefore, the proposed method was also applied as a baseline model for performance comparison, and the results of HoID, TB-CoAuth, COMPA, and SAHP are compared and shown in Figure 15.

In previous research work, several researchers started from the user's external information, network features, content features, and activity features using active duration, commonly used devices, account update status, and text characteristics based on the content of tweets as the main features for anomaly detection. Results have been achieved in the early stage of applications, but there is no reasonable use of tweet content, which truly expresses the characteristics of individual user differences. However, these traditional characteristics are considered to be easier to imitate and have a high degree of deception, which affects the prediction effect of the model. More information of the features that these models used is provided in Table 3.

However, as per Pariser's filter theory, every user unknowingly builds their own bubble space based on their interests and search patterns. Hence, social users will be active in different social spaces, and their tweet patterns, interest topics, and social circles have established their own unique patterns. Even if user interests fluctuate, this will evolve over time, without a sudden change. In view of the uniqueness and stability of personal style, even if criminals use personal data to obtain user interests to maintain the active status of the account after hijacking it, it cannot fit with the exclusive mode of real users, so this behavior pattern earns an automatic strict violation.

Based on the above reasons, the distribution of interests and hobbies implicit in the content of user tweets was fully investigated and the main characteristics of users were quantified. The stability of the distribution of user interests

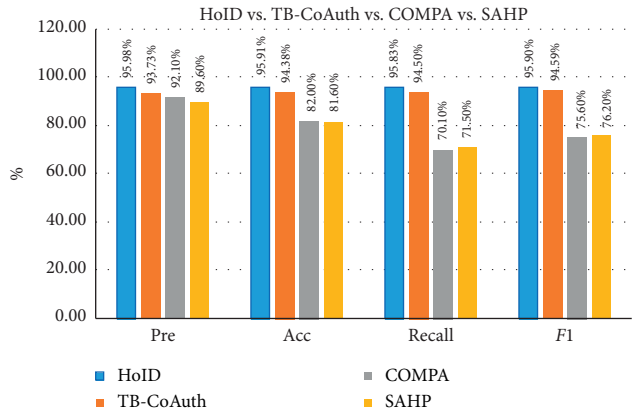


Figure 15: Results comparison of HoID, TB-CoAuth, COMPA, and SAHP.

Table 3: Comparison of feature selection with competitive methods in field of abnormal-account detection on social networks.

| Model | Feature types | Feature used | Technique | Replicability of features | Remarks |
|---|---|---|---|---|---|
| Spot 1.0 | Attribute features | Including the number of followers and followees, reputation, frequency of tweets, average number of URLs, hashtag, and trends | Machine-learning classification and statistical analysis | Easy | (1) Presented a tool developed for scoring suspicious profiles on Twitter through a three-dimensional indicator<br>(2) Limited features for each category were examined<br>(3) Text and semantics in tweets were completely ignored |
| OddBall | Network features | Number of nodes, number of edges, weights, eigenvalues, and number of friends | Unsupervised method to detect abnormal nodes in weighted graphs | Easy | (1) Discovery of new patterns that egonets follow<br>(2) Huge size of social network made it difficult to expand and gather network features |
| DARPA | Attribute features, network features, and content features | User name/avatar, geographical location, and number of followers/followings; tweet syntax and tweet semantics, such as frequent topics; sentiment inconsistency; average number of tweets per day, average clustering coefficient of retweet, and number/percentage of bots in cluster | Step 1: initial bot detection by manually inspecting<br>Step 2: clustering-based outlier detection (non-negative matrix factorization and KNN search) and network analysis<br>Step 3: classification/outlier analysis (SVMs) | Easy | (1) Algorithm detected all bots in set scene<br>(2) System needed to be semisupervised, with help of human judgement to augment automated bot-identification processes<br>(3) Powerful visualization tools were needed to help analysts capture suspicious robots |
| COMPA | Activity features and content features | Time (hour of day), message source, message text (language), message topic, links in messages, direct user interaction, and proximity | Based on user behavioral profile, anomaly detection used content and URL similarity measures | Easy | (1) Created behavioral profiles of users to detect deviation from normal model<br>(2) Compared to previous version, COMPA looked at isolated compromises that affect high-profile accounts<br>(3) It took a significant amount of time and computational resources to collect profile information from users<br>(4) Accuracy of detection results depended on established behavioral profile and selected threshold |

Table 3: Continued.

| Model | Feature types | Feature used | Technique | Replicability of features | Remarks |
|---|---|---|---|---|---|
| SAHP | Activity features and content features | Active time, message source (terminals), message topic, link, stop word, keyword, and mention (@) | Combines information gain ratio with analytical hierarchy process algorithm | Easy | (1) Presented profile features of users more comprehensively<br>(2) Improved on previously established COMPA methods for detecting compromised accounts<br>(3) Detection behavior of proposed algorithm was highly dependent on threshold value, selection of which may introduce bias |
| TB-CoAuth | Content features | Content free, content specific, stylometric, and folksonomy | Continuous authentication of textual content, incremental learning, and supervised-machine-learning classifiers | Hard | (1) Various features are selected: content free and content specific<br>(2) Best classifier: SVM with RBF (radial basis function) kernel<br>(3) $F1$-score: 94.57%<br>(4) In the era of big data, it was inappropriate to rely on statistical and manual selection of features |
| HoID | Content features | Hurst of Interest Distribution | Machine-learning classification (LDA) and statistical analysis (Hurst) | Hard | (1) Feature selection is novel and precise, with personal uniqueness<br>(2) Detection process does not need investment of human resources, which greatly improves algorithm efficiency and accuracy<br>(3) Best classifier: KNN<br>(4) $F1$-score: 95.90% |

and hobbies to detect abnormal users was then analyzed. Experimental results show that the modeling based on the distribution of user interests can improve the effect of detecting abnormal accounts.

It is found that HoID performs better than COMPA, SAHP, and TB-CoAuth on a similar dataset. It is evident from Figure 15 that the four HoID indicators are all above 95%. Each classifier's precision is similar, while the accuracy, recall, and $F1$-measure of HoID are more the 10% higher than those of COMPA and SAHP. In addition, HoID's performance also increased by approximately 1% compared with that of TB-CoAuth.

## 5. Conclusions

In this paper, the detection methods based on user characteristics in social platforms are simply classified and summarized, and the potential hidden dangers are identified. HoID, an abnormal detection algorithm, is proposed to quantify the distribution of user hobbies over a period of time through the LDA model, and the stability of user interests and hobbies is quantified by the Hurst index. Experiments prove that the proposed method has a good effect in abnormal-account detection, which is an improvement over previous research in which the recall rate of abnormal accounts reaches up to 97.93%. It is concluded that with increasing tweet size $S$, decreasing tweet-block overlap $O$, and increasing topic number $N$, the classification effects become better.

While periodic research results are obtained in this paper, several areas for improvement remain. First, the LDA model can be trained with more theme-specific text than just tweets from Twitter users. Because the length of tweets is limited and the topic is not clear enough, the topic classification effect of the LDA model is limited. Second, in terms of the selection of datasets, due to the lack of datasets in the detection of compromised accounts, the method of cross-construction is used to generate abnormal accounts, which are not very close to negative samples used in actual situations.

## Data Availability

Previously reported varol-2017 data were used to support this study and are available at https://botometer.osome.iu.edu/bot-repository/datasets.html. These prior studies and datasets are cited at relevant places within the text as references [36, 37].

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This research was supported by the National Key R&D Program of China (grant no. 2017YFB0802803), Beijing Natural Science Foundation (grant no. 4202002), and Research Project of the Department of Computer Science in Beijing University of Technology (BJUT) (grant no. 2019JSJKY004).

## References

[1] "Digital 2021: global overview report," 2021, http://datareportal.com/reports/digital-2021-global-overview-report.

[2] "Spam and phishing in 2020," 2021, http://securitylist.com/spam-and-phishing-in-2020/100512.

[3] R. Björn, P. Laura, C. Benjamin et al., "Are social bots a real threat? An agent-based model of the spiral of silence to analyze the impact of manipulative actors in social networks," *European Journal of Information Systems*, vol. 28, no. 4, pp. 394–412, 2019.

[4] H. Brian, D. P. Joseph, and M. K. Taghi, "The impact of malicious accounts on political tweet sentiment," in *Proceedings of the 4th IEEE International Conference on Collaboration and Internet Computing*, pp. 197–202, Philadelphia, PA, USA, October 2018.

[5] O. Varol, E. Ferrara, C. A. Davis et al., "Online human-bot interactions: detection, estimation, and characterization," in *Proceedings of the 11th International Conference on Web and Social Media, ICWSM 2017*, pp. 280–289, Montréal, Canada, May 2017.

[6] K. C. Yang, O. Varol, C. A. Davis et al., "Arming the public with artificial intelligence to counter social bots," *Human Behavior & Emerging Technologies*, vol. 115, 2019.

[7] L. Luceri, A. Deb, S. Giordano et al., "Evolution of bot and human behavior during elections," *First Monday*, vol. 24, no. 9, 2019.

[8] L. Luceri, S. Giordano, and E. Ferrara, "Detecting troll behavior via inverse reinforcement learning: a case study of russian trolls in the 2016 US election," in *Proceedings Of the International AAAI Conference On Web And Social Media*, pp. 417–427, Atlanta, GA, USA, June 2020.

[9] M. Kaur, D. Singh, and R. S. Uppal, "Parallel strength pareto evolutionary algorithm-II based image encryption," *IET Image Processing*, vol. 14, no. 6, pp. 1015–1026, 2020.

[10] M. Kaur and D. Singh, "Multiobjective evolutionary optimization techniques based hyperchaotic map and their applications in image encryption," *Multidimensional Systems and Signal Processing*, vol. 32, no. 1, pp. 208–301, 2021.

[11] M. Kaur, D. Singh, and V. Kumar, "Color image encryption using minimax differential evolution-based 7D hyper-chaotic map," *Applied Physics B: Lasers and Optics*, vol. 126, no. 9, 2020.

[12] E. Alothali, N. Zaki, E. A. Mohamed et al., "Detecting social bots on twitter: a literature review," in *Proceedings of the International Conference on Innovations in Information Technology*, pp. 175–180, Al Ain, UAE, November 2018.

[13] A. H. Wang, "Detecting spam bots in online social networking sites: a machine learning approach," *Lecture Notes in Computer Science*, vol. 6166, pp. 335–342, 2010.

[14] P. Efthimion, P. Scott, and P. Nicholas, "Supervised machine learning bot detection techniques to identify social twitter bots," *SMU Data Science Review*, vol. 1, 2018.

[15] C. Cai, L. Li, and D. Zengi, "Behavior enhanced deep bot detection in social media," in *Proceedings of the 2017 IEEE International Conference On Intelligence And Security Informatics (ISI)*, Beijing, China, July 2017.

[16] K. Sneha and F. Emilio, "Deep neural networks for bot detection," *Information Sciences*, vol. 467, pp. 312–322, 2018.

[17] V. S. Subrahmanian, A. Azaria, S. Durst et al., "The DARPA Twitter bot challenge," *Computer*, vol. 49, no. 6, pp. 38–46, 2016.

[18] J. Im, E. Chandrasekharan, J. Sargent et al., "Still out there: Modeling and identifying russian troll accounts on twitter," 2019, https://arxiv.org/abs/1901.11162.

[19] A. Addawood, A. Badawy, K. Lerman et al., "Linguistic cues to deception: identifying political trolls on social media," in *Proceedings Of the International AAAI Conference On Web And Social Media*, Münich, Germany, June 2019.

[20] S. Kumar, R. West, and J. Leskovec, "Disinformation on the web: impact, characteristics, and detection of wikipedia hoaxes," in *Proceedings of the 25th International Conference On World Wide Web*, pp. 591–602, Montréal, Canada, April 2016.

[21] H. Cai, V. W. Zheng, and C. Chang, "A comprehensive survey of graph embedding: problems, techniques, and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 9, pp. 1616–1637, 2018.

[22] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021.

[23] S. Kirill, T. Denis, and Z. Andrey, "Make social networks clean again: graph embedding and stacking classifiers for bot detection," *CEUR Workshop Proceedings*, vol. 2482, 2019.

[24] A. Seyed Ali, N. Pejman, T. Raad Bin et al., "Detect me if you can: spam bot detection using inductive representation learning," in *Proceedings of the the Web Conference 2019-Companion of the World Wide Web Conference*, pp. 148–153, New York, NY, USA, May 2019.

[25] G. Wang, X. Zhang, S. Tang et al., "Unsupervised clickstream clustering for user behavior analysis," in *Proceedings of the Chi Conference. ACM, 2016*, San Jose, CA, USA, May 2016.

[26] D. Kim, T. Graham, Z. Wan, and M.-A. Rizoiu, "Analysing user identity via time-sensitive semantic edit distance (t-sed): a case study of Russian trolls on twitter," *Journal of Computational Social Science*, vol. 2, no. 2, pp. 331–351, 2019.

[27] R. Xin, Z. Wu, H. Wang et al., "Profiling online social behaviors for compromised account detection," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 1, pp. 176–187, 2016.

[28] S. Wu, Q. Liu, Y. Liu et al., "Information credibility evaluation on social media," in *Proceedings of the 30th AAAI Conference On Artificial Intelligence*, pp. 4403-4404, Phoenix, ARI, USA, February 2016.

[29] Z. Yamak, J. Saunier, and L. Vercouter, "Detection of multiple identity manipulation in collaborative projects," in *Proceedings of the 25th International Conference Companion on World Wide Web*, pp. 955–960, Montréal, Canada, April 2016.

[30] N. Chavoshi, H. Hamooni, and A. Mueen, "DeBot: Twitter Bot Detection via Warped Correlation," ICDM," *IEEE Computer Society*, vol. 1, 2016.

[31] Y. Liu, J. Wang, and Y. Jiang, "PT-LDA: a latent variable model to predict personality traits of social network users," *Neurocomputing*, vol. 210, 2016.

[32] P. Zhang, H. Gu, M. Gartrell et al., "Group-based latent dirichlet allocation (group-LDA): effective audience detection for books in online social media," *Knowledge-Based Systems*, vol. 105, 2016.

[33] P. Shinjee, K. Eunhui Kim, and K. Munchurl Kim, "LDA-based unified topic modeling for similar TV user grouping and TV program recommendation," *IEEE Transactions on Cybernetics*, vol. 45, no. 8, pp. 1476–1490, 2015.

[34] Z. Gao, Y. Fan, C. Wu et al., "SeCo-LDA: mining service Co-occurrence topics for composition recommendation," *IEEE Transactions on Services Computing*, vol. 12, no. 3, pp. 446–459, 2019.

[35] R. Yan and S. J. Li, "Document retrieval algorithm based on query intent identification and topic modeling," *computer engineering*, vol. 44, no. 3, pp. 189–194, 2018.

[36] O. Varol, E. Ferrara, C. A. Davis et al., "Online human-bot interactions: detection estimation, and characterization," in *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*, Montréal, Canada, May 2017.

[37] J. Roesslein, "Tweepy," http://www.tweepy.org.

[38] D. Trang, F. Johansson, and M. Rosell, "Evaluating algorithms for detection of compromised social media user accounts," in *Proceedings of the 2015 Second European Network Intelligence Conference*, pp. 75–82, IEEE, Karlskrona, Sweden, September 2015.

[39] N. V. Chawla, K. W. Bowyer, L. O. Hall et al., "SMOTE: Synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, no. 1, pp. 321–357, 2011.

[40] M. Egele, G. Stringhini, C. Kruegel, and G. Vigna, "Towards detecting compromised accounts on social networks," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 4, pp. 447–460, 2017.

[41] H. Tang, X. Wang, K. Zheng et al., "Detection of compromised accounts in osns based on a supervised analytical hierarchy process," *IET Information Security*, vol. 14, 2020.

[42] R. Kaur, S. Singh, K. Harish, and "TB-CoAuth, "Text based continuous authentication for detecting compromised accounts in social networks," *Applied Soft Computing Journal*, vol. 97, 2020.