A Real-Time Detection Algorithm for Abnormal Users in Multi Relationship Social Networks Based on Deep Neural Network

Ai-ping Zhang[1] and Ying Chen[2(✉)]

[1] School of Software, East China JiaoTong University, Nanchang 330013, China
[2] College of Art, Xinyu University, Xinyu 338000, China

Abstract. In view of the imbalance of abnormal data in social networks, the bandwidth value of detection algorithm is high. To solve this problem, a real-time anomaly detection algorithm based on deep neural network is designed. A multi relationship social gathering model was set up, and random forest was used to process tag data. The deep neural network is used to create a set of suspicious network abnormal nodes, and the time-varying component of abnormal data is set. The wavelet transform square integral is used to deal with the abnormal data acquisition process, and the real-time detection algorithm is finally constructed. Prepare the environment parameters required by the algorithm, build the algorithm running environment, prepare two kinds of traditional detection algorithm and design detection algorithm for experiments, the results show that the designed detection algorithm has the largest bandwidth and the best performance.

Keywords: Deep neural network · Multi relationship social network · Abnormal user · Real time detection

1 Introduction

With the popularity of social networks, the security of users' social networks is becoming more and more important. The huge number of users of social network platform has attracted the attention of many attackers for profit. Attackers can create a large number of false accounts and embezzle normal accounts to send users false advertising, phishing, pornography, fraud and other bad information. This kind of malicious behavior seriously affects the user's online experience and the user's information property security [1]. Because social network platform provides users with many social functions, users can establish friend relationship, and the information released is public, which leads to the bad information in social network is more threatening than traditional spam information. As far as Sina Weibo social platform is concerned, attackers gain benefits through malicious likes, concerns, comments and sending malicious links.

Some bad users gain benefits by providing malicious access, attention and other abnormal microblog services. Nowadays, shopping websites can't directly search the

S. Liu and X. Ma (Eds.): ADHIP 2021, LNICST 416, pp. 179–190, 2022.
https://doi.org/10.1007/978-3-030-94551-0_15

shops that provide this service, but they can still purchase this service through the underground market, which seriously affects the social relationship of users and the credit system of the website [2]. Abnormal users generally do not engage in social interaction, but will send a large number of friend requests, and publish spam content in some popular microblogs to attract the attention of normal users. For the current social network abnormal account detection technology, the attacker will constantly update the attack method, by hiding identity and other means to avoid being detected. This phenomenon brings great challenges to the detection technology of abnormal users in social networks. The academia and industry need to constantly improve the detection technology of abnormal users to deal with different attack modes of abnormal accounts. Therefore, the abnormal user detection technology of social network needs to be innovated with the change of attacker attack mode, which has been a technical problem in the field of information security.

2 Real Time Detection Algorithm for Abnormal Users in Multi Relational Social Networks Based on Deep Neural Network

2.1 Setting up a Multi Relationship Social Gathering Model

In the social network environment, the number of abnormal users is less than that of normal users, so in the process of data collection, the collection of abnormal users is relatively troublesome. When setting the multi relationship social gathering model, we first mark the abnormal users in the social network and collect the data from the abnormal users. Compared with the marked normal users, the number of abnormal users is less [3]. Therefore, in the process of data modeling, we should fully consider the imbalance of data. In order to eliminate the imbalance of this part, the random forest is used to process the labeled data, and the labeling process can be expressed as the following.

$$C_R = \frac{O_c}{O_t} \tag{1}$$

In formula (1), C_R denotes tag parameters, O_c denotes exception data sets, and O_t denotes all social network data sets. After label processing, a classifier trained by a deep neural network is used to weight the labeled data set into a relation collection model, which can be expressed as a relation collection model.

$$V = \frac{\sum_{a=1}^{T} T_a}{C_R} \tag{2}$$

In formula (2), V is the relation parameter, T_a is the data set after labeling, and the meaning of other parameters remains unchanged. Using the model constructed above, the collected data sets are continuously generalized. In order to improve the fitting speed of multiple relationships, the trees dealing with neural network are independent from each other [4]. Based on cross validation, the abnormal user detection model is constructed. Random forest algorithm determines the category of the final sample by

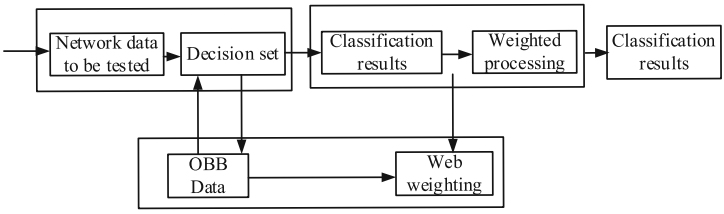

Fig. 1. Processing flow of acquisition model

simply voting the classification results of multiple trees. After verification, the structure of the acquisition model is shown in Fig. 1.

In the process flow shown in Fig. 1, the multi relationship data collected by the model is summarized. The collected data correspond to different relationship types, and are sorted into different data sets, then the deep neural network is used to create the suspicious network abnormal node set.

2.2 Using Deep Neural Network to Create Abnormal Node Set of Suspicious Network

Using the data set obtained from the above collation, because the overall structure of most real social media is gradually evolving, there will be no significant structural change in a short time [5]. This means that successive snapshots can be very similar to each other. Therefore, we can only focus on some important components that change over time, so as to find out the users who may be abnormal. This kind of component is called time-varying component, and the meaning and time parameters of this part of variables are set, as shown in Table 1.

Table 1. Time-varying components and time parameters set

Serial number	Time varying component	Time changing process
1	Node	New node insertion
2		Old node deletion
3	Edge	New edge generation
4		Old edge delete
5	Weight	Weight increase
6		Weight reduction

Through Table 1, we can see that the time-varying components include nodes, edges and weights, which are important indicators reflecting the structural changes of the graph. The exception usually occurs in the place where the structure changes, so the collection of suspicious abnormal nodes can be constructed by analyzing the range of nodes affected by time-varying components. Through the above description, the suspicious abnormal

node set is defined as formula (3) quantity relationship.

$$S(n) = \frac{E(v) \cup E(v^{-})}{u} \tag{3}$$

In formula (3), $S(n)$ is the set of suspicious nodes, $E(v)$ is the number of new suspicious nodes, $E(v^{-})$ is the number of new edge endpoints, and u is the network node parameters. From the connectivity of the structure, we can use them to build a local substructure which has a close relationship with the node [5]. But in real social media, the close relationship between users usually follows the network node structure as shown in Fig. 2.

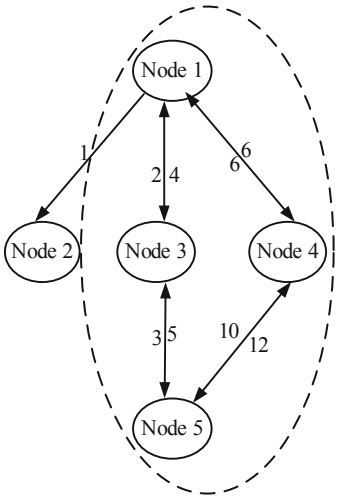

Fig. 2. Structure of social tight network

From the network structure shown in Fig. 2, it can be seen that the self network of nodes only focuses on structural connectivity. The former completely ignores the weight information of edges, while the latter ignores the risk of reliability reduction after continuous transmission [6]. In order to control the risk of data transmission, a weight threshold is set, which can be expressed as.

$$d = \frac{1}{k} \sum_{i=1}^{k} T_i \tag{4}$$

In formula (4), k is the number of network nodes and T_i is the transfer function. According to the calculation formula, only the nodes in the super self network of the node to be processed need to be calculated the intimacy with the node to be processed when constructing the core neighborhood of the node more accurately. Moreover, the maximum size of the core network is the neighbor within two hops, and the core network considers both structural connectivity and intimacy transfer [7]. To process the passed nodes as a collection is to create a collection of suspicious network abnormal nodes. Using this node set, a real-time detection algorithm is constructed.

2.3 Complete the Construction of Detection Algorithm

Using the node and data set [8] obtained from the above processing, the calculation formula (2) and calculation formula (3) of wavelet transform processing are defined, and the square integrable function $\varphi(t)$ of the two formulas satisfies the conditional wavelet mother condition. $\int_{-\infty}^{+\infty} \left|\hat{\varphi}(\omega)\right|^2 |\omega|^{-1} d\omega < +\infty$, where ω is the wavelet coefficient [12]. Set formula (2) as letter a and formula (3) matrix as letter b. set a and B to meet formula (5) at the same time.

$$\varphi_{a,b} = \frac{1}{\sqrt{|a|}} \varphi[\frac{t-b}{a}] \tag{5}$$

In formula (5), t represents the pulse coefficient between two formulas a and B. So the final abnormal user real-time detection expression (6) is obtained.

$$f(t) = \frac{1}{C} W_f(a,b) \varphi(\frac{t-b}{a}) \tag{6}$$

In formula (6), C is a constant between 0.5 and 0.7, W_f is the band coefficient and φ is the covered signal band. In order to maintain the bandwidth value of the detection algorithm in real-time detection [9]. Set the processing steps of the detection algorithm, and the set processing steps are shown in Fig. 3.

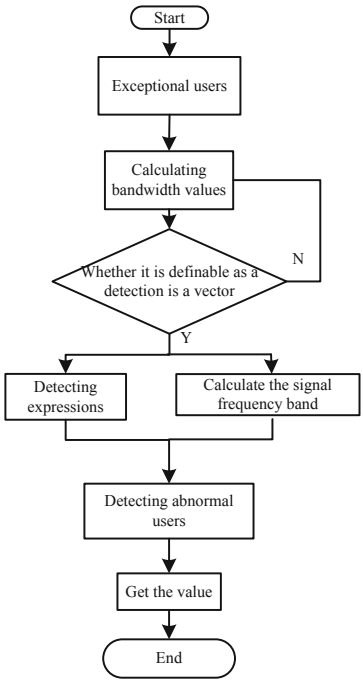

Fig. 3. Steps of electronic music type detection algorithm

In the calculation process of the algorithm shown in Fig. 3, use the formula (4) to calculate the network abnormal data in a variety of data transmission environments [10–12]. Control $f(t)$ to keep the minimum in the detection process to realize the real-time detection of abnormal users in multi relationship social network. The real-time anomaly detection algorithm for social network users studied in this paper, the basic knowledge of the multi-relational social model constructed in this paper, uses the random forest algorithm to weight the labeled data. Use wavelet transform to detect abnormal user data in real time.

3 Simulation Experiment

3.1 Experimental Data Set Preparation

The experimental preparation is shown in Table 2, and the processor parameters are shown in Table 2.

Table 2. Computer parameters prepared

Serial number	Environment	Explain
1	Computer model	Dell XPS 8910
2	Hardware environment	CPU: Intel i76700 Memory: 8G
3	Operating system	Windows 8.1
4	Development language and environment	J2EE, Struct2, Spring,
5	Web server	Tomcat
6	Database	MySQL 5.6
7	Language technology tools	Lucene3.50 ICTCLAS2013

Under the above computer parameters, the social network in the regional network environment is taken as the processing object. Randomly select a wireless sensor network as the communication information network of the experiment, take the sensor node data as the collection object, collect the experimental sample data. The sample data obtained are shown in Table 3.

Table 3. Collected data of abnormal network users

Serial number	Data group name	Number of sample data
1	Data group 1	200
2	Data group 2	400
3	Data group 3	600
4	Data group 4	800
5	Data group 5	1000
6	Data group 6	1200
7	Data group 7	1400
8	Data group 8	1600
9	Data group 9	1800
10	Data group 10	2000
11	Data group 11	2200
12	Data group 12	2400
13	Data group 13	2600
14	Data group 14	2800
15	Data group 15	3000
16	Data group 16	3200
17	Data group 17	3400
18	Data group 18	3600
19	Data group 19	3800
20	Data group 20	4000

Taking the sample data collected in Table 3 as the exception data processing object, the database structure of social network exception users is built. The database structure is shown in Fig. 4.

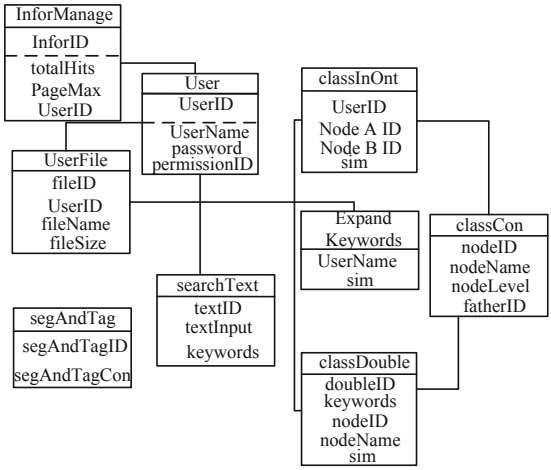

Fig. 4. Structure of exception database constructed

In the above abnormal database structure, the traditional detection algorithm 1, traditional detection algorithm 2 and the detection algorithm designed in this paper are used to experiment. The performance of the three detection algorithms is compared.

3.2 Results and Analysis

Based on the above experimental preparation, the error detection rate of three detection algorithms is defined as the evaluation result of the detection algorithm. The formula (7) can be expressed as.

$$F = \frac{FN}{FN + FP} \times 100\% \tag{7}$$

In formula (7), F is the false detection rate, FN is the number of abnormal data detected as normal, and FP is the number of correct data detected. Under the control of the above numerical relationship, the false detection rate results of the three detection algorithms are shown in Table 4.

Table 4. Results of false detection rate of three detection algorithms

Dataset name	False detection rate results/%		
	Traditional detection algorithm 1	Traditional detection algorithm 2	Design detection algorithm
Data group 1	29.1	15.7	8.5
Data group 2	22.4	15.7	8.7

(*continued*)

Table 4. (*continued*)

Dataset name	False detection rate results/%		
	Traditional detection algorithm 1	Traditional detection algorithm 2	Design detection algorithm
Data group 3	20.9	16.8	8.1
Data group 4	25.5	17.5	8.2
Data group 5	21.4	17.9	8.5
Data group 6	26.4	17.9	8.9
Data group 7	24.7	17.9	8.4
Data group 8	20.5	15.9	8.7
Data group 9	28.4	17.4	8.2
Data group 10	25.9	17.3	8.7
Data group 11	20.1	16.5	8.3
Data group 12	26.5	16.1	8.1
Data group 13	22.6	15.6	8.1
Data group 14	20.6	16.8	8.5
Data group 15	29.5	17.5	8.2
Data group 16	27.8	16.5	8.3
Data group 17	23.9	15.9	8.5
Data group 18	29.7	15.5	8.5
Data group 19	22.2	16.7	8.2
Data group 20	25.8	15.5	8.9

The three detection algorithms are controlled to process the data set prepared for the experiment. According to the calculation formula of the false detection rate defined above, the false detection rate value of the traditional detection algorithm 1 is about 25%, and the false detection rate value is the largest. The false detection rate of the traditional detection algorithm 2 is about 16%, and the actual false detection rate of the algorithm is large. Compared with the two traditional detection algorithms, the designed detection algorithm has the lowest false detection rate and the best detection effect.

Keep the above experimental environment unchanged, corresponding to the prepared experimental data set, repeat 100 iterations to process the data set processed in the preparation stage. Taking the average time generated by the iterative detection data set as the final experimental result, the detection time of the three detection algorithms is measured and calculated. The detection results are shown in Table 5.

Table. 5 Detection time results of three detection algorithms

Dataset name	Test time/s		
	Traditional detection algorithm 1	Traditional detection algorithm 2	Design detection algorithm
Data group 1	11.5	8.5	6.1
Data group 2	11.4	9.2	5.8
Data group 3	14.6	8.1	5.3
Data group 4	11.8	9.4	5.2
Data group 5	13.2	8.1	5.4
Data group 6	14.8	8.6	5.3
Data group 7	13.8	9.4	5.2
Data group 8	12.9	9.9	5.3
Data group 9	12.6	8.8	6.2
Data group 10	13.7	9.7	5.1
Data group 11	14.5	8.9	6.4
Data group 12	14.8	9.9	6.2
Data group 13	13.3	8.8	6.3
Data group 14	11.8	9.7	5.5
Data group 15	13.7	8.1	5.2
Data group 16	10.4	9.8	5.8
Data group 17	11.9	9.4	5.7
Data group 18	13.6	9.7	6.9
Data group 19	11.2	9.8	6.3
Data group 20	12.5	9.2	6.4

According to the statistical calculation results, the average detection time of traditional detection algorithm 1 is about 12.9 s, and the actual detection time is the longest. The average detection time of traditional detection algorithm 2 is 9.15 s, which takes a long time. The average detection time of the designed algorithm is about 5.78 s. Compared with the two traditional detection algorithms, the detection time of the designed algorithm is the shortest.

In the above experimental environment, after selecting the test data set prepared for the experiment, set the operation name of the multi relationship social network and the corresponding data size, as shown in Table 6.

Under the operation set in Table 6, the three detection algorithms are controlled to process the above operation 10 times. Take the operation serial number 1 ~ 10 as a processing process, measure the time required for a processing process. Calculate the

Table 6. Operation name of multi relationship social network and corresponding data volume

Operation number	Operation name	Data volume size/MB
1	File write operation	15.7
2	Relation call	42.6
3	Process creation	21.7
4	Process hiding	39.2
5	File hiding	40.9
6	Invalidation attack	21.7
7	Resource consumption	45.5
8	Kernel boot	42.8
9	Module compilation	38.2
10	End of process	46.7

measurement bandwidth of different detection algorithms, and the bandwidth results are shown in Fig. 5.

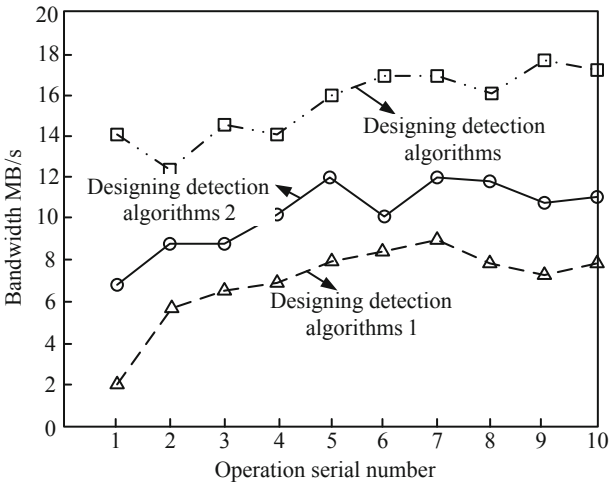

Fig. 5. Bandwidth results of three detection algorithms

According to the experimental results shown in Fig. 5, under the control of the three detection algorithms, corresponding to different operations, the average detection bandwidth generated by the traditional detection algorithm 1 is about 7 MB/s, the processing bandwidth is small, and the detection process is slow. The average detection bandwidth generated by the traditional detection algorithm 2 is about 10 MB/s, and the bandwidth generated by the detection process is larger. The detection bandwidth of the designed

detection algorithm is about 16 MB/s. compared with the two traditional detection algorithms, this detection algorithm has the largest detection bandwidth and can process the most information.

4 Conclusion

In today's information explosion era of big data, while social networks provide people with convenient and diversified services, there are also some hidden information security risks. In order to improve the detection efficiency of abnormal network users, this paper uses deep neural networks to develop a personalized crawling strategy to crawl abnormal social user data, deeply analyze the difference between normal users and abnormal users in user information and behavior characteristics, and extracts that can distinguish between normal users and abnormal users. The important characteristics of the user to realize the real-time detection process. Experiments show that the average detection time of this method is about 5.78 s, and the detection band frame is about 16 MB/s, which has good performance. In the future, we will study more anomaly detection conditions corresponding to different social network scenarios, and consider in-depth study of the design of multiple social network user interaction models as the focus of subsequent work.

References

1. Chen, S., Zhu, G.-S., Qi, X.-Y., et al.: Custom user anomaly behavior detection based on deep neural network. Comput. Sci. **46**(S2), 442–445, 472 (2019)
2. Yin, J., Peng, Y., Lu, Y., et al.: Research on user abnormal behavior prediction of enterprise information system based on deep neural networks. J. Manage. Sci. **33**(01), 30–45 (2020)
3. Li, H., Zhu, M.: A small object detection algorithm based on deep convolutional neural network. Comput. Eng. Sci. **42**(04), 649–657 (2020)
4. Liu, S., Liu, G., Zhou, H.: A robust parallel object tracking method for illumination variations. Mob. Netw. Appl. **24**(1), 5–17 (2019)
5. Zhang, Y., Xu, T., Feng, D., et al.: Research on faster RCNN object detection based on hard example mining. J. Electron. Inform. Technol. **41**(06), 1496–1502 (2019)
6. Yang, X.-X., Li, H.-B., Hu, G.: An abnormal behavior detection algorithm based on imbalanced deep forest. J. China Acad. Electron. Inform. Technol. **14**(09), 935–942 (2019)
7. Xu, H.-J., Zhang, H., He, W.: A cloud user anomaly detection method based on mouse behavior. J. Harbin Univ. Sci. Technol. **24**(04), 127–132 (2019)
8. Li, X., Han, X., Wang, Z., et al.: A deep heterogeneous network embedding algorithm based on twin neural network. Telecommun. Eng. **60**(11), 1271–1277 (2020)
9. Shen, X., Shen, Z., Huang, Y., et al.: Deep convolutional neural network for parking space occupancy detection based on non-local operation. J. Electron. Inform. Technol. **42**(09), 2269–2276 (2020)
10. Li, J., Yun, X., Li, S., et al.: HTTP malicious traffic detection method based on hybrid structure deep neural network. J. Commun. **40**(01), 24–33 (2019)
11. Liu, S., He, T., Dai, J.: A survey of CRF algorithm based knowledge extraction of elementary mathematics in Chinese. Mob. Netw. Appl. **26**(5), 1891–1903 (2021). https://doi.org/10.1007/s11036-020-01725-x
12. Liu, S., Fu, W., He, L., et al.: Distribution of primary additional errors in fractal encoding method. Multimedia Tools Appl. **76**(4), 5787–5802 (2017). https://doi.org/10.1007/s11042-014-2408-1