Detection Method of Abnormal Behavior of Network Public Opinion Data Based on Artificial Intelligence

Ying-jian Kang(✉), Lei Ma, and Yan-ning Zhang

Beijing Polytechnic, Beijing 100176, China
kangyingjian343@163.com

Abstract. In order to improve the effect of network public opinion data abnormal behavior detection, an artificial intelligence-based network public opinion data abnormal behavior detection method is proposed. By constructing the network public opinion data model, recognizing the evolution rule of network public opinion data, locating the abnormal data area according to the behavior detection algorithm, and using the probability neural network under artificial intelligence to detect the abnormal data behavior. The experimental results show that the detection method proposed this time is 28.12% and 84.37% higher than the two traditional methods when detecting large-scale public opinion abnormal behavior data. It can be seen that the detection method based on artificial intelligence is not restricted by the volume of network data, and the detection effect is better.

Keywords: Artificial intelligence · Network public opinion · Data abnormal behavior · Detection method

1 Introduction

Internet public opinion refers to the popular Internet public opinion with different views on social issues, which is a form of social public opinion. Through the Internet, it can spread the public opinions and opinions with strong influence and tendentiousness on some hot spots and focus issues in real life. The network public opinion takes the network as the carrier and the event as the core. It gathers the expression, dissemination and interaction of the emotions, attitudes, opinions and opinions of the majority of netizens. However, as a public use platform, the network has abnormal data behavior. Due to the single detection technology, the traditional detection methods will lose some abnormal data when facing massive network information [1]. Therefore, based on artificial intelligence technology, a new detection method for abnormal behavior of network public opinion data is proposed. By constructing the network public opinion data model, the evolution of network public opinion is divided into three stages: public opinion generation period, public opinion diffusion period and public opinion reduction period. The evolution law of network public opinion data in different stages is identified. The abnormal data area is located by using behavior detection algorithm, and abnormal data behavior is detected by artificial intelligence technology. The

S. Liu and L. Xia (Eds.): ADHIP 2020, LNICST 347, pp. 332–344, 2021.
https://doi.org/10.1007/978-3-030-67871-5_30

experimental results show that the method proposed in this paper has a good effect on abnormal behavior detection of network public opinion data.

2 Artificial Intelligence-Based Network Public Opinion Data Abnormal Behavior Detection Method

2.1 Building the Data Model of Network Public Opinion

To construct a network public opinion data model, first assume that the interaction behavior of individual opinions at the micro level is restricted by the trust threshold, that is to say, the interaction behavior can only occur when the distance between the two parties' views is less than the trust threshold, otherwise the two parties will avoid contact and insist on their own Original point of view. Although the interaction rule of this view is simple, it vividly reveals the phenomenon of "different ways do not conspire" in the process of interpersonal communication. Therefore, through the continuous iteration of this interaction mode, we can explore some rules of the aggregation process of group views from a macro perspective. The model breaks through the limitation of the individual point of view binary or can only take values within a limited range of values. It is believed that the individual point of view can be any real number within a given range of values. Among the categories of models. In addition, the model assumes that any two individual groups randomly selected from the group may have viewpoint interaction behavior, as long as their viewpoint distance is less than the trust threshold [2].

Let the group size be W, i, and j are two random individuals in the group, and the viewpoint values at time t are expressed as $u_i(t)$ and $u_j(t)$, respectively, and $u_i(t), u_j(t) \in [0, 1]$, given a trust threshold φ, is a constant between the range $[0, 1]$, If $\left|u_i(t) - u_j(t)\right| \leq \varphi$, then:

$$\begin{cases} u_i(t+1) = u_i(t) + \lambda\big(u_j(t) - u_i(t)\big) \\ u_j(t+1) = u_j(t) + \lambda\big(u_i(t) - u_j(t)\big) \end{cases} \tag{1}$$

Otherwise:

$$\begin{cases} u_i(t+1) = u_i(t) \\ u_j(t+1) = u_j(t) \end{cases} \tag{2}$$

In formula (1): Parameter λ is the convergence coefficient of the model, which will affect the speed of system convergence. Adjust the value of the convergence coefficient λ to define the nature of the group. Let λ take the value in the interval $[0, 0.5]$. When $\lambda = 0$, all individuals always adhere to their own views without any change; when $\lambda = 0.5$, both sides of the view interaction will get the average of the two views. The above two cases correspond to the interaction groups in two extreme cases respectively. Generally speaking, When the value of table λ is small, it is corresponding to the individuals with stronger strategies. They are not easy to change their own views, and when they are large, they tend to adopt a compromise view interaction strategy. When $\lambda > 0.5$, it means that after two individuals interact with each other, they respectively choose the views that are more inclined to each other, that is, their views have been transposed. At this time, the model constructed believes that this kind of situation rarely occurs in real life, so it is necessary to make $\lambda \in [0, 0.5]$ [3]. In order to simplify the model, the value of convergence coefficient is usually fixed, that is, the rules of opinion interaction are as follows:

$$\begin{aligned} &|u_i(t) - u_j(t)| \le \varphi, \begin{cases} u_i(t+1) = u_i(t) + 0.5\big(u_j(t) - u_i(t)\big) \\ u_j(t+1) = u_j(t) + 0.5\big(u_i(t) - u_j(t)\big) \end{cases} \\ &|u_i(t) - u_j(t)| > \varphi, \begin{cases} u_i(t+1) = u_i(t) \\ u_j(t+1) = u_j(t) \end{cases} \end{aligned} \tag{3}$$

In the network public opinion data model, trust threshold φ has an important influence on the aggregation process of group opinion. Through the model simulation, it can be found that when $\varphi \ge 0.5$, the group tends to form a consensus, that is, all individuals in the group ultimately hold the same views on a given issue. With the decrease of φ value, the group gradually divides into two or more opinion groups, and members of each opinion group share the same views.

2.2 Identify the Evolution of Internet Public Opinion Data

From its generation to its final demise, Internet public opinion is in a complete dynamic change process, and presents or follows some internal rules to run. Therefore, on the basis of the constructed public opinion data model, the evolution of network public opinion is divided into three stages: public opinion generation period, public opinion diffusion stage and public opinion restoration stage, and the evolution law of network public opinion data in different stages is found out. Figure 1 is the life cycle curve of network public opinion obtained from the model analysis.

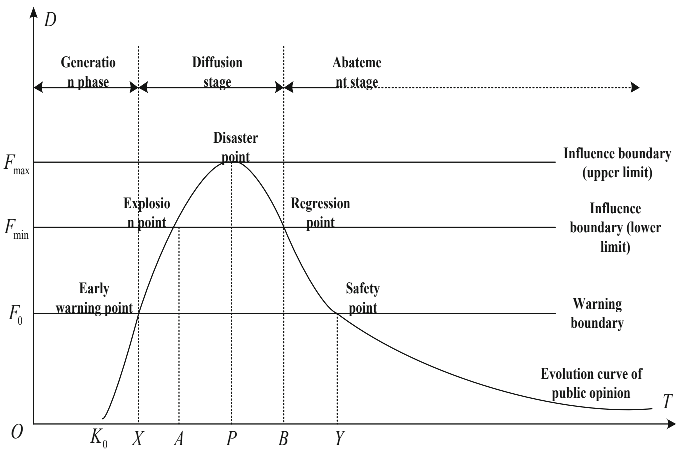


Fig. 1. Life cycle curve of online public opinion

In the figure, *OT* represents the time axis; *OP* represents the evolution degree of network public opinion. Based on the analysis of the existing public opinion life cycle of the emergency network, combined with the key characteristics of "five points", the public opinion of the emergency network can be divided into three stages: public opinion generation stage, public opinion diffusion stage and public opinion reduction stage.

2.2.1 Generating Law of Network Public Opinion

The generation of network public opinion is a special risk factor and risk influence formed in the network due to the interference of a "nature economy society" system. Previous studies have shown that analyzing and grasping the generation law of public opinion is the basic premise and key to effectively guide and manage public opinion. The generation rule of online public opinion believes that during the generation of network public opinion, it mainly undergoes four triggers: "trigger-agglomeration-hot discussion-burst", and each of them presents corresponding evolutionary rules: shape mutation rule, superposition focus law, resonance convergence law, group polarization law. The state generation of network public opinion can also be regarded as "state break" or "state break", and the control factor of this "state break" belongs to a dependent variable, and the "state break" condition can be regarded as a fixed critical limit value. Therefore, based on this value, we can build a catastrophe simple function generated by public opinion of emergency network [4].

Assuming that the damage status and social impact of a network event are a risk and crisis independent variable function x, the event stakeholders are a risk dependent variable function y, and the calculation formula of public opinion risk value is:

$$\mu = f(x, y) \tag{4}$$

Where: μ is the public opinion risk index; $f(x, y)$ is the risk function of x and y. If the risk value of public opinion does not exceed the risk critical limit value F_0, i.e. $\mu \le F_0$, it is a sub stable balance, i.e. it belongs to the incubation period of network public opinion. During this period, Internet public opinion was basically stable or under control, and Internet public opinion was relatively calm. If the control factor energy continues to increase due to some internal or external forces, the risk value of public opinion will reach or exceed the risk critical value Limit value F_0, i.e. $\mu \ge F_0$, Then it will release potential energy and seriously destroy the inherent state of public opinion. The state of public opinion balance will suddenly "interrupt" or "break", forming a real public opinion risk. Figure 2 below is a schematic diagram of the generation of Internet public opinion.

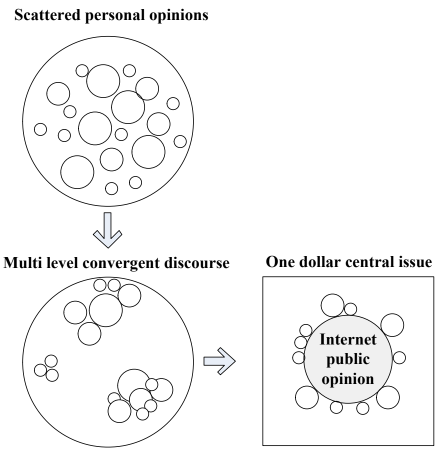


Fig. 2. Schematic diagram of network public opinion generation rule

2.2.2 The Law of Diffusion of Internet Public Opinion

Due to the characteristics of freedom, interaction and complexity of Internet public opinion, its diffusion and development is not a simple linear rise or decline, but a complex process. But generally speaking, the diffusion of network public opinion mainly refers to the process of transmission and change from small to large, from weak to strong, from recessive to dominant after the generation of network public opinion. In essence, it is a process of "strengthening and magnifying" after the generation of online public opinion. Generally speaking, in the stage of network public opinion diffusion, along the logical structure of "strengthening and amplifying", its diffusion is going through four stages of "popularity—reinforcement—repetition—sublimation". It presents certain evolutionary laws: linear asymptotic law, ripple divergence law, interference interaction law and spiral ascent law.

The dissemination effect of public opinion and the amount of information transmission show a linear positive relationship. The linear diffusion evolution of public opinion is a structural description of the process of public opinion communication. That is to say, every link and process of public opinion dissemination are closely linked after the emergence of public opinion in emergency network. Suppose that in the online public opinion, communicator a may spread to communicator b, while communicator b develops to communicator c, etc. The propagation process can be expressed as $W_a \rightarrow W_b \rightarrow W_c \rightarrow W_d \rightarrow \cdots$, and the public opinion is amplified in a linear mode to form large-scale public opinion, and the scale Public opinion is the development and value of communicator a to communicator n. At this time, the scale public opinion function can be expressed as:

$$W = \sum_{i=1}^{n} \sigma_i \tag{5}$$

Where: σ_i represents the scale index of i data. When the scale of public opinion is consistent with the analysis scale in the first section, that is, when W reaches a certain extreme value, the network public opinion will then spread, forming a gradual mode [5].

2.2.3 Law of Reducing Public Opinion on the Internet

Every material movement in the world is a process of alternate development of prosperity and decline. Prosperity is a kind of development of material movement, and decay is also a kind of development. Similarly, the network public opinion also develops along a rise and fall. The reduction of network public opinion mainly refers to the process of network public opinion gradually decreasing and declining from large to small, strong to weak, and changing from hot-spot events to common events after the generation and spread of network public opinion. With the proper handling and resolution of the event, the social resources driven by the network public opinion are gradually exhausted, the public's attention to the event shows fatigue, the network media and traditional media pay less attention to and report, and the development of the network public opinion lacks a new power mechanism, so the network public opinion begins to gradually enter a slowly subsided reduction stage. In the process of reducing public opinion on the Internet, some have disappeared in a broken manner, and some have repeatedly weakened. But generally speaking, it is the process of progressing from prosperity to decline along the evolution path of "conflict-order change-fading-dying". In this process, each public opinion reduction node accordingly exhibits certain laws: conflict blocking law, substitution transfer law, defocusing and fragmentation law and natural dissipation law [6]. Figure 3 is the conflict resistance function diagram of network public opinion reduction.

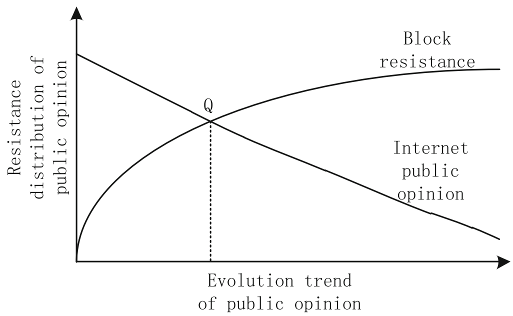


Fig. 3. Network public opinion reduces conflict blocking function

In the figure, point Q represents the intersection point between public opinion reduction and conflict resistance. From this point on, public opinion drops rapidly. At this point, it realizes the recognition of evolution rule of network public opinion data.

2.3 Behavior Detection Algorithm Locates Abnormal Data Area

In the actual environment, the data is not immutable, it changes with the time environment, and new abnormal behavior public opinion data appears, so the behavior detection algorithm is used to divide the information categories. Most of the time, the network is in a normal state, and the network data is normal or abnormal data. With the change of time and environment, the progress of network technology will be accompanied by the continuous change of network abnormal behavior. At this time, new network data will appear, whose characteristics are completely different from the previous data characteristics, that is, the abnormal behavior data will also be different [7].

In the behavior detection algorithm, comparing the distance $d(z, h_i)$ between the test sample and the center point of the nearest anomaly class H_i and the maximum distance D_i, $d(z, h_i) > D_i$ between the center point of the anomaly class and other points in the class, it is determined that a new anomaly class appears, The point is defined as a new anomaly point. When there are other test points that determine the new anomaly class, D_i is the distance between the two, and then the anomaly classification model is updated; otherwise, the test sample is determined as an i type anomaly. Suppose that in the test sample, the object is z, H_i is the center of mass of cluster i, h is the center of mass of all points, and the abnormal node is located. Figure 4 is the positioning result [8].

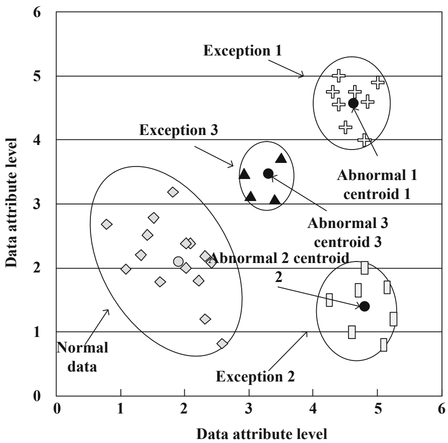


Fig. 4. Location of abnormal data nodes

According to the selected behavior detection algorithm, we can locate the abnormal behavior data in the network public opinion data.

2.4 Artificial Intelligence Technology Detects Abnormal Data Behavior

Known probability neural network is one of the artificial intelligence indexes, which is suitable for data separation. The intelligent network originates from radial basis function network, which can make use of simple linear structure to expand the nonlinear learning algorithm. The main principle of probabilistic neural network is to combine density function estimation and Bayesian decision theory. Under certain conditions, the network can detect abnormal behavior of public opinion data.

The behavior area of abnormal data is E, the prior probability is $V_1 = p(E_1)$, $V_2 = p(E_2)$, and $V_1 + V_2 = 1$. Given the input vector $\gamma = [\gamma_1, \gamma_2, \ldots, \gamma_n]$, in order to get a set of observation results, the classification basis is as follows:

$$E = \begin{cases} E_1, p(E_1|\gamma) > p(E_2|\gamma) \\ E_2, otherwise \end{cases} \tag{6}$$

Where: $p(E_1|\gamma)$ is the probability of occurrence of γ, the posterior probability of category E_1 [9, 10]. According to the Bayesian formula, the posterior probability is equal to:

$$p(E_1|\gamma) = \frac{p(E_1)p(\gamma|E_1)}{p(\gamma)} \tag{7}$$

In classification decision-making, input vectors should be classified into categories with higher posterior probability. In practical application, risk and loss should be considered. For example, the loss caused by the wrong classification of the E_1 sample into the E_2 category, or the wrong classification of the E_2 sample into the E_1 category is often very different, so the classification rules need to be adjusted. Define behavior β_i as the behavior of assigning the input vector to h_i, and ε_{ij} as the loss caused by taking behavior β_i when the input vector belongs to h_j, then the expected risk of behavior β_i is:

$$R(E_1|\gamma) = \sum_{j=1}^{n} \varepsilon_{ij} p(E_1) \delta_i \tag{8}$$

In the formula: δ_i represents the probability density function of E_1. So far, the probability network in artificial intelligence technology is used to detect the abnormal behavior of network public opinion data [11, 12].

3 Experiment and Analysis

In order to verify the reliability of the proposed detection method, two traditional detection methods are selected and applied to the detection of abnormal behavior of network data. The proposed method is used as experimental group A, and the two traditional methods are used as experimental group B and experimental group C respectively. Choose a website as the experimental test background condition. It is known that there is abnormal behavior data on this website. The three detection methods are used to locate the abnormal behavior data in the network public opinion data database, and the positioning differences of the three detection methods are compared.

3.1 Experiment Preparation

Three sets of network public opinion data sequences with different data volumes are selected, and the abnormal data are hidden in the three data sequences respectively. Figure 5 below is a statistical diagram of the data volume of different data sequences.

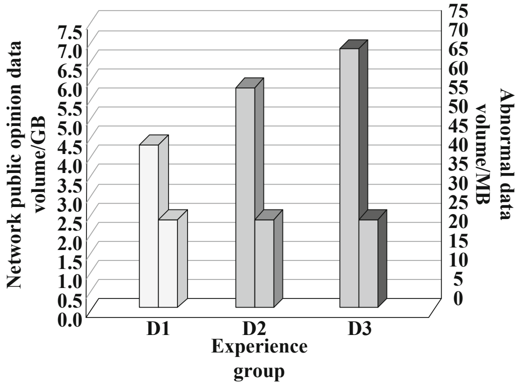


Fig. 5. Statistical chart of data series

In the figure, D1 is a public opinion data sequence with a small volume; D2 is a public opinion data sequence with a moderate volume; D3 represents a public opinion data sequence with a large volume. The two statistical results in each group represent the total data volume and abnormal data volume of the sequence. It is known that under the premise of different data sequence volume, the data volume of abnormal behavior of network public opinion is the same, both of which are 22.5 MB. D1 group and D3 group were used as test variables to detect three groups of public opinion data series. Get and analyze the experimental test results.

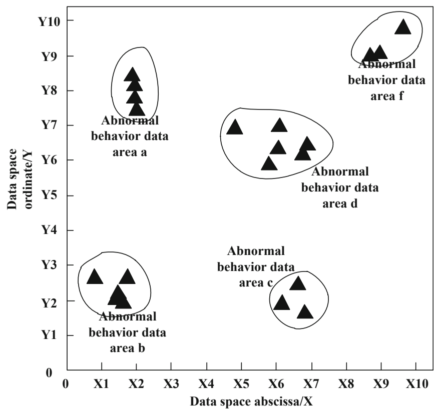
3.2 The First Set of Experimental Results

Group D1 with relatively small data volume is taken as the first group of experimental test objects, and Fig. 6 below is the experimental test results.

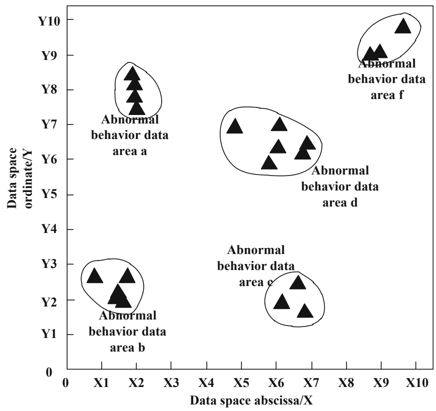
According to the test results shown in the above figure, in the face of small data series, the three detection methods can detect the abnormal behavior data, without missing any abnormal network public opinion data. According to the statistics, the detection rate of three groups of methods is 100%, and the difference between them is also 0. It can be seen that three detection methods are applicable to the detection of abnormal behavior of network public opinion data with small volume.

3.3 The Second Set of experimental results

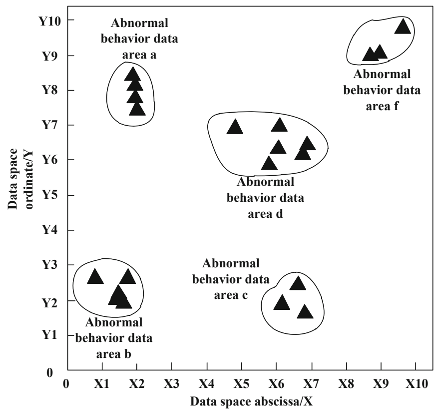
In order to ensure the authenticity and reliability of the experimental test results, group D3 with relatively large data volume is taken as the second group of experimental test objects, as shown in Fig. 7 below.



（a）Experiment A test results

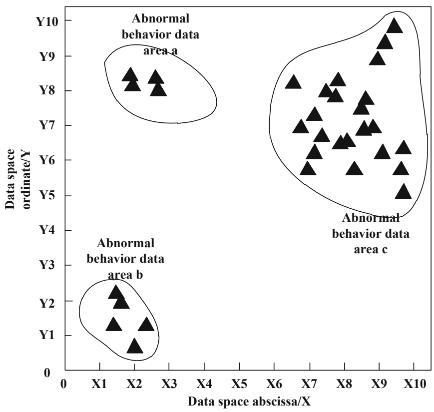


（b）Test results of group B

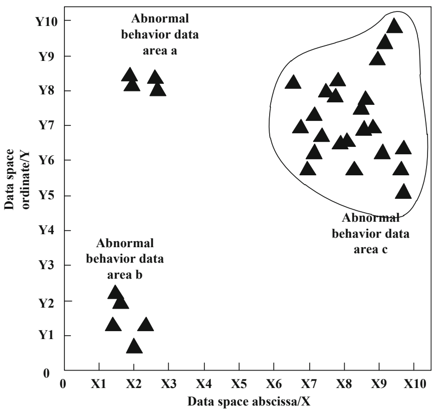


（c）Experiment C test results

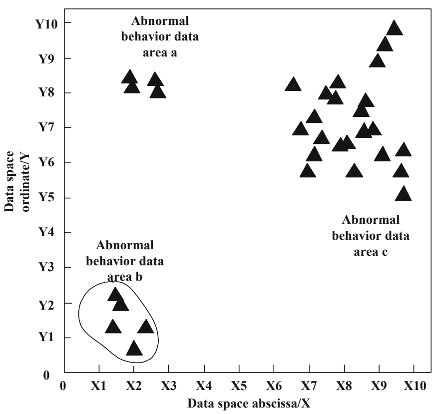
Fig. 6. The first group of experimental test results



（a）Experiment A test results



（b）Test results of group B



（c）Experiment C test results

Fig. 7. The second set of experimental test results

According to Fig. 7, in the face of large-scale data series, experimental group a can also fully detect abnormal behavior data, while experimental group B and experimental group c lose a lot of abnormal behavior information. The detection rate of the three methods is as shown in Table 1 below.

Table 1. Statistical results of detected bit rate

Group	Arrival rate	Difference from the rate of full arrival
Group A	100%	0
Group B	71.88%	28.12%
Group C	15.63%	84.37%

According to Table 1, the detection of group A is completely in place; the detection of group B exists 28.12%. The detection rate of group C was only 15.63%. It can be seen that the traditional two detection methods are not suitable for the detection of abnormal behavior of large-scale network public opinion data. It can be seen from the results of the two sets of experimental tests that the detection method proposed this time can perform data abnormal behavior detection in large-scale network public opinion data.

4 Conclusion

Artificial intelligence, also known as intelligent machinery and machine intelligence, is the intelligence shown by machines made by human beings. Usually, artificial intelligence presents human intelligence through ordinary computer programs. Artificial intelligence is applied to abnormal behavior detection to provide more reasonable technical support for detection methods. The proposed detection method fully realizes the detection of abnormal behavior data. However, the total number of tests in this experiment is less. In future research, we should expand the experimental data, increase the number of experiments, and strengthen the persuasion of test results [13].

References

1. Xiwei, W.A.N.G., Ruonan, J.I.A., Duo, W.A.N.G., et al.: Research on the development trend of artificial intelligence research in library and information field. Libr. Inf. Serv. **63** (01), 70–80 (2019)
2. Longjie, S., Kaijun, Y.: Research on big data analysis model of library user behavior based on internet of things. Comput. Eng. Softw. **40**(06), 113–118 (2019)
3. He, H., Haojiang, D., Jun, C.: Network abnormal behavior detection technologies based on traffic-feature modeling. J. Netw. New Media **8**(04), 11–20 (2019)
4. Lianhai, S., Ying, L.: Multivariate autoregression based algorithm for anomaly detection in rating data. Comput. Eng. Des. **39**(06), 1629–1632 1652 (2018)
5. Huanli, P.A.N.G., Hongyan, L.I.: Intelligent detection simulation for crowded pedestrian abnormal behavior. Comput. Simul. **35**(11), 405–408 (2018)

6. Jia, Y., Xiaonan, L., Yang, Z., et al.: Abnormal behavior detection for campus email systems based on big data analysis. J. Commun. **39**(S1), 116–123 (2018)
7. Xiaolin, W.U., Fuyuan, C.A.O.: An unsupervised outlier detection algorithm for categorical matrix-object data. J. Shenzhen Univ. (Sci. Eng.) **36**(01), 33–42 (2019)
8. Baisheng, G.U.A.N., Chunjiang, B.I.A.N., Shuichun, F.E.N.G., et al.: Abnormal action recognition of interaction research based on neural network. Electron. Des. Eng. **26**(20), 1–5 (2018)
9. Xiao, L., Yi, L., Yefeng, J., et al.: Overall design of artificial intelligence application based on monitoring big data platform. Power Syst. Big Data **22**(04), 37–42 (2019)
10. Shuai, L., Weiling, B., Nianyin, Z., et al.: A fast fractal based compression for MRI images. IEEE Access **7**, 62412–62420 (2019)
11. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks. Entropy **21**(9), 902 (2019)
12. Liu, S., Li, Z., Zhang, Y., et al.: Introduction of key problems in long-distance learning and training. Mob. Netw. Appl. **24**(1), 1–4 (2019)
13. Liu, S., Glowatz, M., Zappatore, M., et al. (eds.).: e-Learning, e-Education, and Online Training, pp. 1–374. Springer International Publishing (2018)