Network Dynamic Bad Information Security Filtering Algorithms Based on Large Data Analysis

Wenchao Zheng[1(✉)], Yin-zhu Cheng[2], Ze-yu Zhang[1], and Yong-qing Miao[1]

[1] Institute of Data Science, City University of Macau, Macao 999078, China
yanghongl9ll@l63.com
[2] School of Economics and Management, Xinjiang University, Xinjiang 830046, China

Abstract. In view of the low filtering accuracy of traditional bad information in the massive data environment, the security filtering algorithm of network dynamic bad information is innovated and improved in the big data environment. Combining the data set analysis algorithm with the grey statistics theory, this paper evaluates the dynamic information security status of the network structure, extracts the information security features in the evaluation results, compares the data features in the network structure, detects the dynamic time domain range of bad information, and filters and corrects the information in the time domain by nodes and channels, so as to realize the security of the dynamic bad information of the network The experimental results show that the dynamic bad network information security filtering algorithm based on big data analysis is more accurate and effective than the traditional algorithm, with high accuracy and the shortest time, and can be used in the network dynamic bad information security effectively, which meets the research requirements.

Keywords: Big data · Network dynamics · Bad information · Filtering

1 Introduction

With the arrival of the era of big data, the rapid development of network technology has gradually penetrated into various fields, and become an indispensable part of people's daily life and work. With the continuous expansion of network scale, the increasing number of network equipment, applications and services, the effective management of network dynamic bad information is gradually increasing difficulty. The information age brings not only science and technology, but also technology security issues such as science and technology. Network is the most potential new technology product in the 21st century. With its rapid development, network has penetrated into every field of people's daily life. Therefore, how to effectively diagnose the hidden dangers of network operation, effectively filter the bad network information and reduce the loss of information to the network has become the focus of current research [1, 2]. For this reason, researchers in related fields have done a lot of research.

S. Liu and L. Xia (Eds.): ADHIP 2020, LNICST 348, pp. 115–126, 2021.
https://doi.org/10.1007/978-3-030-67874-6_11

In reference [3], an intelligent prediction method of network abnormal failure rate based on big data analysis and fault spectrum feature extraction is proposed. The correlation spectrum feature detection method is used to collect network abnormal failure data, the collected network fault information feature is matched and filtered, and the adaptive beamforming method is used to focus the network fault big data, The extracted network fault big data is classified and identified by fuzzy clustering method, and the intelligent prediction of network abnormal fault rate is realized under the big data, so as to ensure the safe operation of network information. This method can effectively reduce the security of network information, but it focuses on the study of network fault, and seldom considers the network security information. In reference [4], a network information security monitoring system is designed to solve the problems of weak de joint analysis ability, response, processing ability and insufficient monitoring breadth and depth. The data layer of the system uses web service and Oracle10g database to collect and store network logs, and carries out data exchange, communication and encryption through soap and SSL; uses J2EE platform to realize data analysis service; uses flex to realize cross platform display. The proposed system can quickly respond to all kinds of network security attacks and improve the work efficiency and work level of network information security monitoring. However, the lack of consideration on the filtering of complete network information leads to poor filtering effect of bad security information. In reference [5], aiming at the deficiency of existing information security evaluation algorithms in dealing with subjective attitude deviation of multi experts, and the traditional sequential machine learning model method to deal with the problem of deviation accumulation in time period, a deep sequential information security evaluation algorithm based on depth fuzzy correction is proposed. Firstly, the expert fuzzy evaluation index is constructed by trigonometric fuzzy function, and then the modified weighted DS evidence reasoning correction index is used, then the loss and possibility matrix features are created, and finally the information security is evaluated by deep time series network. The simulation experiments are carried out on MIT data set. The experiments analyze whether the features can cope with multi expert conflicts, and evaluate the accuracy, robustness and time efficiency of the algorithm. The proposed algorithm has stronger fuzzy evaluation ability, stronger ability to deal with conflict opinions among experts, and more accurate information security evaluation in time sequence, but the efficiency of the algorithm is low.

In the complex Internet environment, the messages sent and received by the network are affected by bad information, which leads to the impact of network security. An intelligent filtering method of bad information in mobile network environment based on mode matching is proposed. This method analyzes the weight of all the information in the network. Based on this, the network information feature database is designed. According to the ontology element weight of positive information, Bayesian classification algorithm is introduced to classify the network information in the database, and the classified network information data are screened for bad information to realize the filtering of bad information. This method can effectively classify information, but there are many positive information in the filtered information, which is not conducive to universal application. In reference [6], an efficient filtering method for spam information in double buffered communication networks is proposed. In this method, the existing information in the network is processed by dimension, and then

the network information is partitioned according to the different dimensions. The principal component analysis method is introduced to gain the network information of each area. The boosting algorithm is used to construct the spam information filter, and the extracted information features are input into the spam information filter to filter the spam information existing in the double buffer communication network, To achieve efficient filtering of spam information in double buffer communication network. This method can realize the filtering of bad information quickly, but the accuracy of filtering information needs to be improved.

Based on the above problems, this paper proposes a dynamic bad information security filtering algorithm based on large amount of data. By using the nonlinear time series location algorithm, we can quickly extract and filter the bad information features, ensure the normal operation of the network, and avoid the loss caused by the failure of network security information. The experimental results show that the proposed algorithm can effectively filter the network dynamic bad information security and has high efficiency.

2 Network Dynamic Bad Information Security Filtering Algorithms

2.1 Network Information Security Situation Assessment

By investigating a large amount of data, a method of evaluating network security situation based on data set analysis algorithm is proposed. The design of network information security situation assessment algorithm is mainly based on grey statistical theory. Assuming that the characteristic quantity of network information's security behavior is S_0, the acquisition factor's behavior characteristic is $S_i - S_0$, the data security is observed on the network information sequence t, and defined as $S_0(t)$, $(t = 1, 2, \ldots, n)$, the sequence of network security information's characteristic behavior can be deduced by combining grey statistical theory, and the sequence of network security information's characteristic behavior is $S_0(t) = [S_0(1), S_0(2) \ldots S_0(n)]$. From this, we can deduce that the sequence of network bad information characteristic factors is as follows: $S_i(t') = [S_i(1), S_i(2), \ldots, S_i(n)]$. Among them, t' is the indefinite value of information detection time, bad information index, object number, etc. It can be concluded that the correlation coefficients of sequence Si and S0 at t point are defined as:

$$\delta_{0i}(t) = \frac{\min\limits_{i}\min\limits_{t}|S_0(t) - S_i(p)|^k + \mu \max\limits_{i}\max\limits_{t}|S_0(t) + S_i(p)|}{|S_0(t) + S_i(p)| + \mu \max\limits_{i}\max\limits_{t}|S_0(t) - S_i(p)|^a} \tag{1}$$

Among them, μ is the evaluation value of the attack that may occur during the operation of network bad information, k is the operation time of network, and a is the number of kinds of bad information that network structure suffers. If n is the degree of harmfulness of bad information to network security attacks, the above algorithms are

combined to collect and process the characteristics of bad information in large data environment. The algorithms are as follows:

$$\delta(S_0, S_i) = \frac{1}{n}\sum_{i=1}^{t} 10\mu_{0i}(t), \mu \in (0,1) \tag{2}$$

Combining the above formulas, we can conclude that the evaluation algorithm of the t security situation evaluation index of network structure R under the influence of bad information is as follows:

$$Rs_{\mathrm{i}}(t) = \delta(S_0, S_i)\sum_{i=1}^{t} 10^{\mu_{0i}} v \tag{3}$$

Among them, v is the number of network structure information supply types. Combined with the above algorithm, the hidden dangers of bad information security in cloud computing network are extracted, and the situation of network information security is evaluated.

2.2 Dynamic Time Domain Location of Bad Information

In the context of the current large data network environment, the scale of data is relatively large, resulting in complex bad information, relatively strong perturbation to the network structure, and easy to lead to network information failure. Therefore, combined with data feature mining algorithm to locate the bad information in the network today, in order to mine the bad information accurately and avoid the interference of bad information (Fig. 1).

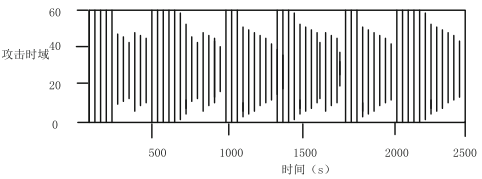

Fig. 1. Time domain analysis model of bad information attacks

Aiming at the attack degree analysis model of bad information in the figure above, the time-domain location analysis and processing are carried out. The scalar time-domain location sequence of network bad information dynamic data is set as follows: $\mathrm{a}(t), t = 0, 1, .., n-1$. The basic screening function algorithm of bad information location signal can be described as:

$$\mathrm{x} = [x_1, x_2, ..., x_N] \in Q^{mN} \tag{4}$$

The time-domain feature analysis method is used to decompose the characteristic data of bad information in the network in two gradient directions in the horizontal and vertical direction. The maximum gradient difference algorithm of data mining can be obtained as follows:

$$\mathrm{f} = \frac{1}{i \times j} \sum_{x=1}^{j} Rs_{\mathrm{i}}(t) \sum_{\mathrm{j}=1}^{i} |\mathrm{f}(x)|\mathrm{a}(t) \tag{5}$$

Formula (5): f is the correlation coefficient of bad light signal feature mining. According to the above formulas, the directional feature range algorithm of bad information location area is as follows:

$$p = \mathrm{a}(t) \frac{Q^{\mathrm{mN}}}{\sqrt{(x)^{m}} \sqrt{\mathrm{f}(x)^{N}}} \tag{6}$$

Assuming that the channel of bad information communication data transmission channel in network structure is continuous, the obtained frequency domain model is:

$$G_n = \zeta_{ij}[a(t_0)]^{\mathrm{i}} + \delta \prod f * Rs_{\mathrm{i}}(\bar{t}) \tag{7}$$

In the formula, $\bar{t}$ is the range of difference of information feature ranking and ζ_{ij} is the error parameter of bad information filtering. Through the above steps, it can provide an effective reference and data basis for the filtering of bad information security by investigating the bad information domain of the network [7].

2.3 Network Dynamic Bad Information Filtering Method

According to the demand of bad information filtering in network, the information security filtering model is constructed. By establishing the data security set, the security data eigenvector is set as: $F=(X, Y)$, where X is the set of limited non-empty security activity information terminal objects e_{n} in the network structure, and Y is the set of all bad information elements in X. If the direction of security information transmission is: $(e_{\mathrm{n}}, e_{\mathrm{m}}) \in Y$. Among them, e_{n} is the impact of bad information on network information e_{m}. Each item e_{n} has a designated security weight of z_n, which fully considers the failure probability of any bad information active terminal object depending on e_{n}, so as to filter bad information effectively. Its filtering algorithm is as follows:

$$L = \prod \mathrm{F}(\mathrm{X}, \mathrm{Y}) * z_n[e_n \rightarrow e_m]' G_n \tag{8}$$

Combining with the above algorithm, the bad information is filtered, and the information transmission path weighting more than z_n is found in the network structure, as well as the shortest path from the bad information feature point to other feature points, so as to avoid the attack of bad information and potential information security risks through the route checking and control [8]. Connect all effective paths in the

process of information transmission in time, and form bad information filtering and early warning location domain whose structure is shown in the Fig. 2.

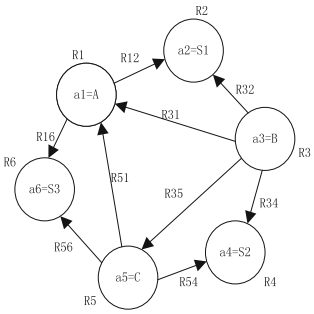

Fig. 2. Time-domain location of bad information filtering

It can be seen from the graph that different kinds of network nodes are universal in network structure. Impression needs to filter bad information in time by modifying network parameters to provide support for fast fault location of network communication [9, 10]. Combined with the above algorithm, the steps of network bad information security filtering in large data environment are improved and perfected. The network dynamic bad information filtering process is shown in the following Fig. 3.

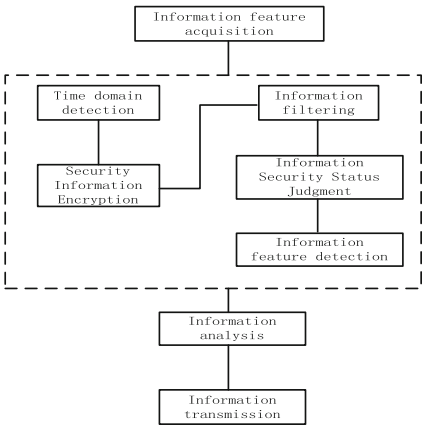

Fig. 3. Network dynamic bad information filtering process

In order to accurately complete the information filtering process and avoid the harmful information influence, damage and other phenomena in the process of information use, the security encryption steps are added. By calculating the time domain range of information security, the network security structure is encrypted and strengthened, and the dynamic bad information of the network is intelligently filtered

and intercepted to effectively protect the network information system. Data security [11, 12].

2.4 Implementation of Dynamic Information Security Filtering

After determining the network topology, the coverage rate of network nodes and the occupancy rate of network information transmission channels are calculated by judging the bad information, so as to search and filter the bad information pertinently. The filtering steps are as follows:

All bad information received within a fixed time interval forms a set of bad information feature sources;

Step 1: Relevant information of all bad information received has N sets of entity objects.
Step 2: For any object in the entity's bad information set, it can accurately locate and screen the bad information.
Step 3: In the process of information filtering, bad information and security information can be divided into two subsets, namely, left subset U (bad information set) and right subset N (new security set). Each information feature in the set has the greatest interdependence.
Step 4: The iteration process is implemented until the iteration result is a single point source set. The common features of left and right subsets of information are sorted from bottom to top or from top to bottom in order to reduce the complexity of filtering the information. By calculating the time-domain location of bad information, we can accurately detect all possible factors of bad information in all locations, and ensure the normal operation of network communication.
Step 5: Randomly extract two common characteristics of bad information in any non-left subset;
Step 6: Find the nodes that can explain the common characteristics of two bad information at the same time;
Step 7: Use these nodes to replace the common characteristics of the original security information, and record and compare them.
Step 8: Repeated acquisition of bad information characteristics until there is no intersection between the two bad information.
Step 9: Determine the propagation node of bad information in the network structure.

In order to accurately obtain the feature interpretation of each bad information, it is necessary to find the bad information transmission nodes in the network structure, and calculate the bad information factor between each node through the cross chain storage method. The maximum common factor method is used to construct the filtering step formula of dynamic bad information in network structure, which is as follows:

$$(U_1 \cap N_1) \cup (U_2 \cap N_2) \cup \ldots \cup (U_{\mathrm{n}-1} \cap N_{n-1}) \cup (U_{\mathrm{n}} \cap N_n) \tag{9}$$

According to the results of the above steps, the best feature interpretation of all bad information in the network structure can be accurately obtained, thus effectively completing the security filtering of dynamic bad information in large data networks.

3 Analysis of Experimental Results

3.1 Experimental Environment

In order to verify the accuracy of network dynamic bad information security filtering algorithm under large data analysis, experimental verification and analysis are carried out. The experiment is based on the Matlab environment. The specific experimental environment is shown in Fig. 4:

Fig. 4. Experimental environment

3.2 Experimental Parameter Setting

Assuming that the network nodes are distributed in 2000 m × 2000 m uniform array area in large data environment, the experimental parameters are set as shown in Table 1.

Table 1. Settings of experimental parameters

Parameter	Remarks
Network frequency band	3 kHz–8 kHz
Information carrier frequency and time domain	3 ms
Initial frequency of information	0.15 Hz
Number of sampling points	256个
Bad information domain scope	−15 dB–15 dB

According to the experimental environment and the results of parameter setting, the experimental contents were analyzed. In the same environment, the accuracy of the network dynamic bad information security filtering algorithm is compared and tested under the large data analysis, and the calculation results are recorded.

3.3 Analysis of Experimental Results

3.3.1 Analysis of Filtering Accuracy of Network Information with Different Algorithms

In order to verify the effectiveness of the proposed method, the experiment analyzes the accuracy of the proposed algorithm and the traditional algorithm for network bad information filtering, and the experimental results are shown in the table below (Tables 2 and 3):

Table 2. Bad information security filtering algorithm in this paper

Practical value	Forecast value	Error value
0.24	0.22	0.02
0.45	0.45	0.00
0.41	0.40	0.01
1.02	1.00	0.02
1.45	1.45	0.00
0.84	0.86	0.02
0.54	0.55	0.01

Table 3. Traditional bad information security filtering algorithms

Practical value	Forecast value	Error value
0.24	0.22	0.12
0.45	0.45	0.08
0.41	0.40	0.23
1.02	1.00	0.14
1.45	1.45	0.32
0.84	0.86	0.12
0.54	0.55	0.23

Comparing with the above results, it is not difficult to find that compared with the traditional bad information security filtering algorithm, the error rate of the proposed dynamic bad information security filtering algorithm is relatively lower, which shows that the accuracy of this method is relatively higher. The experiment analyzes the prediction value, the actual value and the error value. It can be seen from Table 1 that the maximum error value of the proposed algorithm is 0.02, and it is found in the experiment that the difference between the actual value and the predicted value of the proposed algorithm is relatively small; while the minimum error value of the traditional method is 0.08, and the difference between the international value and the predicted value is relatively large. Through the comparison, it can be seen that the proposed algorithm has a high accuracy when filtering the bad information of the network, which is verified The effectiveness of the proposed algorithm.

3.3.2 Gradient Analysis of Network Information Filtering Based on Different Algorithms

In order to verify the practicability of the algorithm, the gradient in the process of information filtering is detected. In the experimental process, the higher the gradient is, the worse the practicability is. The test results are plotted as follows (Fig. 5):

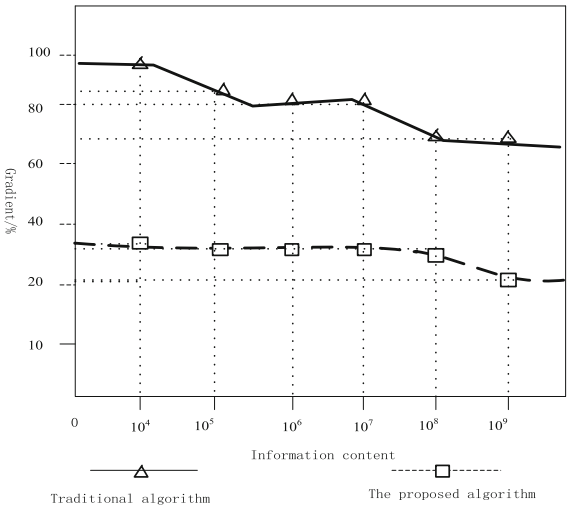

Fig. 5. Comparisons of experimental results

Through the above test results, it is not difficult to find that compared with the traditional bad information filtering algorithm, the proposed algorithm has lower data gradient and significantly improved stability, so it is proved that the proposed algorithm has better practicability and fully meets the research requirements.

3.3.3 Time Consuming Analysis of Network Information Filtering with Different Algorithms

In order to further verify the feasibility of the proposed algorithm, the time-consuming of the proposed algorithm and the traditional algorithm in filtering network bad information is analyzed in the experiment, and the experimental results are shown in Fig. 6:

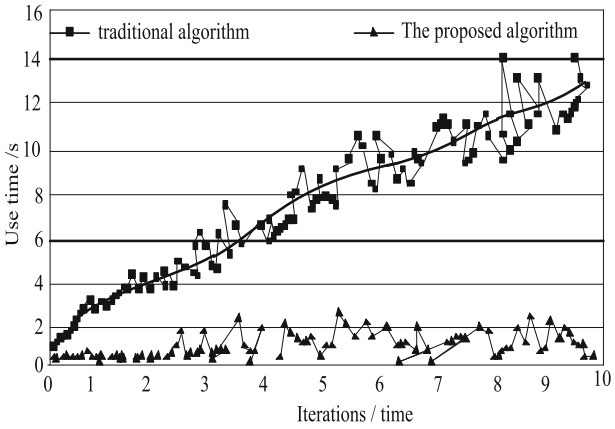

Fig. 6. Comparison of network information filtering time of different algorithms

It can be seen from the analysis of Fig. 6 that there is a certain gap in the time taken by the two methods for filtering the bad network information. Among them, the proposed algorithm takes the shortest time and the traditional method takes longer time. The validity and feasibility of the proposed algorithm are verified.

4 Conclusion

In the complex large data network environment, the accurate perception and prediction of information security situation has become a key issue of current research in various fields. Aiming at the problems of low precision and low efficiency of traditional bad information filtering algorithm, a dynamic bad information filtering algorithm based on data mining and large data information processing is proposed. Experiments show that the accuracy and practicability of the proposed method are significantly improved compared with traditional methods.

References

1. Qiang, J., Wei, Z., Song, L., et al.: Research on Personalized Adaptive Learning - A New Normal of Digital Learning in Big Data Age. China Audio Vis. Educ. **12**(2), 24–32 (2016)
2. Qingzhou, Z., Yong, L., Shiming, T., et al.: State monitoring and fault handling based on big data analysis of intelligent distribution network. Grid Technol. **40**(3), 774–780 (2016)
3. Lei, Y., Bin, L., Zhuoyu, W.: Research on the new model of teachers' information technology ability training - based on "internet plus" and "big data" thinking. China Audio Vis. Educ. **15**(8), 61–66 (2016)
4. Haijuan, Y.: Research on influencing factors of reform and innovation of personnel training mode of information management specialty in the background of big data - taking Hubei university as an example. Libr. Inf. Knowl. **24**(2), 21–29 (2016)

5. Xianmin, Y., Si, T., Jihong, L.: Framework and development trend of educational big data: the overall framework of "research and practice column of educational big data." Modern Educ. Technol. **26**(1), 5–12 (2016)
6. Si, Z., Qingtang, L., Shijie, L., et al.: Research on learner input in online learning space - big data analysis of online learning behavior. China Audio Vis. Educ. **13**(4), 24–30 (2017)
7. Shuai, L., Gelan, Y.: Advanced Hybrid Information Processing, pp. 1–594. Springer, USA
8. Liu, S., Fu, W., He, L., Zhou, J., Ma, M.: Distribution of primary additional errors in fractal encoding method. Multimed. Tools Appl. **76**(4), 5787–5802 (2014). https://doi.org/10.1007/s11042-014-2408-1
9. Liu, S., Lu, M., Li, H., et al.: Prediction of gene expression patterns with generalized linear regression model. Front. Genet. **10**, 120 (2019)
10. Minghua, W., Jingui, Z.: An evaluation algorithm of information security based on fuzzy adjustment feature matrices and deep time sequence model. **46**(5), 464–470 (2018)
11. Nanzhong, W.: Reconstruction of teaching design framework from the perspective of mixed learning - also on the supporting role of educational big data in teaching design. China Audio Vis. Educ. **59**(5), 18–24 (2016)
12. Liu, S., Glowatz, M., Zappatore, M., Gao, H., Gao, B., Bucciero, A.: E-Learning, E-Education, and Online Training, pp. 1–374. Springer International Publishing, USA