

Retraction

Retracted: Efficient Detection of Large-Scale Multimedia Network Information Data Anomalies Based on the Rule-Extracting Matrix Algorithm

Advances in Multimedia

Received 15 August 2023; Accepted 15 August 2023; Published 16 August 2023

Copyright © 2023 Advances in Multimedia. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their

agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] J. Zhao, "Efficient Detection of Large-Scale Multimedia Network Information Data Anomalies Based on the Rule-Extracting Matrix Algorithm," *Advances in Multimedia*, vol. 2021, Article ID 3299891, 7 pages, 2021.

Research Article

Efficient Detection of Large-Scale Multimedia Network Information Data Anomalies Based on the Rule-Extracting Matrix Algorithm

Jie Zhao 

College of Electronic Information Science, Fujian Jiangxia University, Fuzhou, Fujian 350108, China

Correspondence should be addressed to Jie Zhao; zhaojie@fjxu.edu.cn

Received 15 August 2021; Accepted 22 October 2021; Published 30 November 2021

Academic Editor: Zhendong Mu

Copyright © 2021 Jie Zhao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous development of multimedia social networks, online public opinion information is becoming more and more popular. The rule extraction matrix algorithm can effectively improve the probability of information data to be tested. The network information data abnormality detection is realized through the probability calculation, and the prior probability is calculated, to realize the detection of abnormally high network data. Practical results show that the rule-extracting matrix algorithm can effectively control the false positive rate of sample data, the detection accuracy is improved, and it has efficient detection performance.

1. Introduction

With the continuous development of multimedia social networks, online public opinion information has become increasingly popular. The dissemination of information has become a new trend with the rapid increase of hot topics among the general public, such as Baidu Hot Search and Weibo Hot Search [1–3]. As we all know, this kind of multimedia network is a double-edged sword. It can convey both positive and negative information. It can guide the public to judge something and also affect the harmonious and stable development of society seriously. Therefore, how to effectively detect such abnormal network information has become an urgent problem to be solved. There have been many explorations in the industry, such as the use of the KNN method for network abnormal information data detection, the confidence method for training, and the comparison of abnormal samples to detect abnormal information data in the network. However, there are certain limitations in terms of this method, and the missed detection

rate is relatively high [4, 5]. Therefore, this is still one of the important research directions, and a variety of detection operators have appeared one after another, such as Forstner operator, SUSAN operator, Harris operator, and so on. In addition, some scholars used singular value decomposition methods for analysis, but they produced visual and auditory errors, and the effect is not obvious [6].

The rule-extracting matrix algorithm extracts implicit and useful related knowledge and rules, such as corresponding rules, knowledge, and so on, from the data warehouse. However, due to different complexities of the data, it is difficult to ensure complete mining of the data [7, 8]. The detection of network format data is realized through filtering, and the detection of network information data anomaly is realized through probability calculation based on the rule-extracting matrix algorithm. The value of the prior probability is calculated in an attempt, and the network data abnormality is detected, aiming to explore the detection of abnormal network data performance.

2. Rule-Extracting Matrix Algorithm

In multimedia network information detection, the most obvious distinction is words with emotional color. Therefore, it is effective to identify, extract, and detect network abnormal information based on emotional words. Therefore, during the whole process, firstly the multimedia network information is grabbed, secondly, the corresponding data samples are analyzed and processed through Chinese word segmentation technology, and finally, the word segmentation results are matched with the existing abnormal information database, and abnormal data are detected in the multimedia network according to the corresponding abnormal information threshold [9].

$$A(t) = \sqrt{e} \cdot B^H(t) \cdot x_n, \quad (1)$$

$$H_B(z) = \frac{(1 + \sin \vartheta_2)}{\cos \vartheta_2} \cdot \frac{\cos \vartheta_1 \cdot z^{-1}}{\sin \vartheta_2} \cdot G(z) \cdot A(t), \quad (2)$$

where $A(t)$ represents the network information data traffic oscillation attenuation estimated value obtained after interference filtering and represents the variance coefficient in multimedia network data, and $B^H(t)$ represents the negative information feature detection-related function in the network information data traffic.

According to the abnormal information data provided by the multimedia network information database as a benchmark, the normal information and abnormal information in the subdatabase are separately counted, and the frequency of normal information and abnormal information of a single word is calculated. The specific calculation formulas are shown in formulas (3) and (4).

$$A_{c_i} = \frac{fP_{c_i}}{\sum_m fP_{c_i}}, \quad (3)$$

$$B_{c_i} = \frac{fN_{c_i}}{\sum_n fN_{c_i}}, \quad (4)$$

where w_1 is the data weight of the character or word segmentation; w_2 is the normal word frequency of w_3 ; $B = b(i, j)$ is the weight of the word segmentation in the abnormal data; and $1 \leq i, j \leq m$ is the frequency of the abnormal words of $S_{i,j}$. In addition, m and n are the corresponding numbers of different characters in the normal information database and the abnormal information database.

From formulas (3) and (4), it can be seen that the abnormal information data in the word segmentation and word database are relatively similar, and the specific calculation formulas are shown in formulas (5) and (6):

$$G(z) = \frac{1 - \sin \vartheta_2}{2} \cdot \frac{1 - z^{-1}}{\sin \vartheta_1 (1 + \sin \vartheta_2)}, \quad (5)$$

$$S_{c_i} = A_{c_i} - B_{c_i}, \quad (6)$$

where $Aa_{i,j}$ is the threshold for measuring whether the word is normal or abnormal information. Assuming $S_{c_i} \geq 0$ is the abnormal frequency corresponding to the correct information data, $S_{c_i} < 0$ represents the abnormal frequency in the network information data.

According to formula (3), the threshold of normal and abnormal information is calculated, and the emotional tendency of words can be expressed by the following formula:

$$S_w = \frac{1}{n} \cdot \sum S_{c_i}. \quad (7)$$

Among them, the calculated result is the detection result of the abnormal information data of the multimedia network.

During the process of detecting abnormal data in multimedia networks, in addition to a lot of effective information, there are also many kinds of redundant interference information. These kinds of interference information will affect the information detection results of abnormal data and cause a certain missed detection rate [10–14].

3. Model Construction and Problem Description

3.1. Large-Scale Visual Multimedia Network Model. As we all know, the current multimedia network model uses a topological network to serve the Internet for service integration. This text builds the network structure of the large-scale visual multimedia network simulation based on the computer graphics and image processing by constructing the multimedia visual network model, as shown in Figure 1.

As shown in Figure 1, the request for each multimedia storing at the same time may be different, but only one storage can be performed. Set $|LFN|$ to the size of the data file request, as shown in the following formula:

$$RC_r = \beta \times \sum |LFN|, \quad (8)$$

where β is a constant used to express the degree of parallelism and the copy of the abnormal data of the multimedia network can be calculated by formula (2), as shown in the following formula:

$$RV = w_1 \times \frac{1}{S} + w_2 \times NA + w_3 \times \frac{1}{CT - LA}, \quad (9)$$

where the values of w_1 , w_2 , and w_3 can be set by themselves and K is the output set of the multiplexer, and the details are shown in the following formula:

$$k = \{1, 2, 3, \dots, |K|\}. \quad (10)$$

For large-scale visual multimedia network topology network, there are abnormal data mutation data, such as error codes, and there are mutation abnormal data in custom data [15]. For the inspection of the abnormal data of the large-scale visual multimedia network, the abnormal information of the large-scale visual multimedia network is required, as shown in formulas (11) and (12):

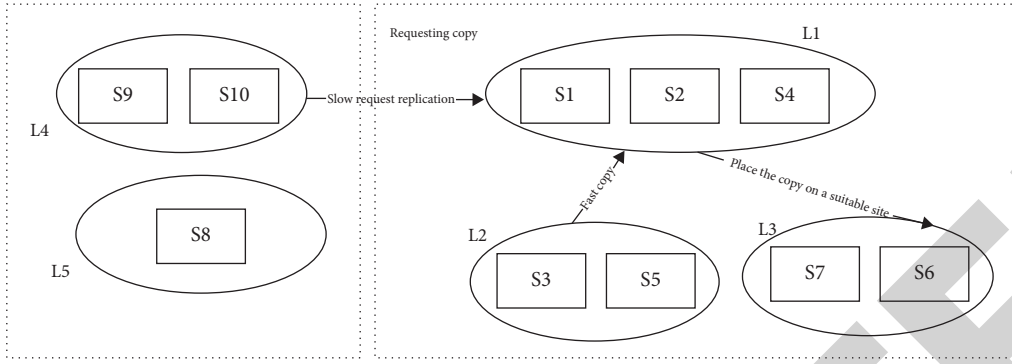


FIGURE 1: Large-scale visual multimedia network simulation platform.

$$T = T_1 + T_2, \quad (11)$$

$$F(x, y) = w_1 \times T \times w_2 \times D(x, y). \quad (12)$$

This function can be tailored because it is defined as a weighted combination of three indicators. Predict the parallel computing time required for calculation of the same chain length on different numbers of cores and obtain the numerical simulation of the CPU running time of the attack data captured by web firewalls of different chain lengths. This realizes the construction of a large-scale visual multimedia network model to provide a model basis for efficient detection of big data.

3.2. Feature Extraction of Big Data in Large-Scale Visual Multimedia Network. If the watermark image in the large-scale visual multimedia network information is represented by $B = b(i, j)$, $1 \leq i, j \leq m$, and m is the dimension of the matrix, the singular value is obtained as shown in formula (10) by solving the singular value of the matrix:

$$Aa_{i,j} = U_{i,j}^* S_{i,j}^* V_{i,j}. \quad (13)$$

On the basis of formula (13), a small matrix unit $Aa_{i,j}$ is obtained. According to a certain decomposition of singular values, a matrix of singular values is obtained, as shown in the following formula:

$$Aa_{i,j} = U_{i,j}^* S_{i,j}^T V_{i,j}. \quad (14)$$

The undirected graph analysis method is used to design the network model. The finite queuing model of each channel NAV task is used to represent the state space, as shown in the following formula:

$$S = \{(k, n), 0 \leq k \leq K, 0 \leq n \leq N\}. \quad (15)$$

Suppose that in the multimedia network database, the predictive performance of data packet information is $\{\lambda_i; 1 \leq i \leq S\}$, and the criterion is $\{R_j; 1 \leq j \leq L\}$. Among them, S represents the amount of input data and L represents the total bandwidth ratio used in the channel. The channel allocation packet conversion waiting time is

$$W_q = W - \bar{X} = \frac{1}{\gamma} \sum_{k=1}^K \sum_{n=1}^N k p_{k,n} - \frac{(N-1)\mu + r}{\mu r}. \quad (16)$$

In the D-dimensional search space established during the mobile learning process, the population size is set to N , the current position of the particle is $X_i = (x_{i1}, x_{i2}, \dots, x_{iD})$, the speed is $V_i = (v_{i1}, v_{i2}, \dots, v_{iD})$, the power consumption of the data stream is C received from all adjacent nodes, and the accurate prediction of the results is $B^H(t)$, and the semisupervised control prediction result of mobile learning is obtained as

$$\hat{q} = \text{arc min}_{q=1} \left(\sum_{q=1}^q q(q|e) - \sum_{q=\bar{q}}^q q(q|e) \right). \quad (17)$$

In order to realize learning from other particles in the evolution, the linear superposition detail capture algorithm is used to perform the optimal configuration of the search ability balance state, and the objective function of supervised learning is obtained as

$$L(\bar{q}, q) = \begin{cases} 1|\bar{q} - q| > \frac{\delta}{2}, \\ 0|\bar{q} - q| \leq \frac{\delta}{2}. \end{cases} \quad (18)$$

In the formula, $B^H(t)$ refers to the related parameters of the scheduling task instruction in the M-learning process, and the computer vision features can be effectively extracted, which can be explained by the following formula:

$$D(C_1, C_2) = \begin{cases} \text{true}, & \text{if } Di f(C_1, C_2) > MInt(C_1, C_2), \\ \text{false}, & \text{otherwise,} \end{cases} \quad (19)$$

$$MInt(C_1, C_2) = \min(\text{Int}(C_1) + \tau(C_1), \text{Int}(C_2) + \tau(C_2)), \quad (20)$$

where the Canny edge detection function $H_B(z)$ is a constant and $|C|$ is the size of the area C , which realizes the feature extraction of big data in a large-scale visual multimedia network.

The stability of multimedia networks and the quality of video and audio transmission are improved through the efficient detection of abnormal data. According to the above description, the implementation process of the improved algorithm is described as shown in Figure 2.

4. Detection Method of the Abnormal Information Data of the Multimedia Network

4.1. Multimedia Network Information Data Interference Filtering Processing. For multimedia network information data, the anti-jamming filtering algorithm is detected for network abnormal information detection to improve the accuracy of information data detection and reduce the rate of missed detection. For the collected multimedia network information data center, the filtered input vector is obtained through Fourier transformation, and the input vector is set as a stable vector input to achieve multimedia network interference filtering.

If the data of the multimedia network information are constructed based on linearly correlated time series, interference suppression is achieved through the following filtering model, as shown in the following formula:

$$x_n = a_0 + \sum a_0 x_{n-i} + \sum b_j, \quad (21)$$

where the initial sample used to represent the flow of the multimedia network information data center is a_0 , the multimedia scalar time series is represented by x_{n-i} , and the central flow fluctuation is represented by b_j .

Based on the above analysis, the obtained FIR cascade filter can be expressed as shown in Figure 3.

Among them, the sparse iterative expression of the filter can be expressed by the following formula:

$$\vartheta_1(k+1) = \vartheta_2(k+1) = \vartheta(k) - y(k) \cdot G(z). \quad (22)$$

According to FIR anti-interference filtering, the flow output of multimedia network information data center can be expressed by the following formula:

$$z(t) = x(t) + y(t) = a(t) + \vartheta_1(k+1) + n(t). \quad (23)$$

4.2. Anomaly Information Data Detection Method Based on Rule-Extracting Matrix Algorithm. Based on the various public opinion topics of the sample, the corresponding probability expression is built, and the record to be detected can be expressed by the following formula:

$$P(\text{Category}|z(t)) = \frac{P(\text{Category}) \cdot P(z(t)|\text{Category})}{Pz(t)}, \quad (24)$$

where $P(\text{Category}|z(t))$ is used to represent a certain probability to be detected. For each data record to be detected, the data vector can be converted into a content vector.

$$P(\text{Category}|z(t)) = \frac{[P(\text{Category}) \cdot P(x_{i1}, \dots, x_{in})]}{P(x_{i1}, \dots, x_{in})}, \quad (25)$$

where $P(x_{i1}, \dots, x_{in})$ is a constant sequence. According to the quantitative calculation of formula (25), the corresponding objective function F_n can be constructed, as shown in the following formula:

$$\begin{aligned} F_n &= \arg \max P(\text{Category}|z(t)) \\ &= \arg \max P(\text{Category}) \cdot P(x_{i1}, \dots, x_{in}). \end{aligned} \quad (26)$$

Based on the assumption of the rule-extracting matrix algorithm, each record has an independent relationship between the component vector values, and the joint probability is the product of the component probabilities. In summary, using the maximum value formula, the final objective function expression constructed is

$$F_n = \arg \max P(\text{Category}) \prod P(x_{in}|\text{Category}). \quad (27)$$

5. Experimental Results and Analysis

In order to effectively detect and simulate the abnormal data of the multimedia network, the data required for the experiment are collected through the collection of the corresponding 5000 pieces of information through the hot search on Weibo (October-November 19) for 2 months. The content involves current affairs politics, sports content, science and technology, literary content, etc. These data are vectorized, including normal information and abnormal information at the same time, to form the final experimental dataset.

- (1) The false positive rate of rule-extracting matrix algorithm is tested.
- (2) The missed detection rate of the rule-extracting matrix algorithm is tested.

The corresponding results are obtained through two simulation experiments, as follows.

5.1. Experiment 1. In Figure 4, the so-called false positive rate is the probability that normal data are judged as abnormal data. Through the analysis of the results, the false positive rates of the detection results of three different detection algorithms under the same number of experimental iterations are quite different. Among them, for the extraction based on rules under the matrix algorithm's multimedia network negative information data detection method, the false positive rate of the detection result is overall low, with the highest one of not more than 5%, and for other information gain network anomaly detection methods, migration technology, and D-S theory data detection methods, the relative false positive rate is more than 25%.

5.2. Experiment 2. In Figure 5, U represents the number of abnormal data in the experimental data subset; 0 represents the number of missed detections of network anomaly

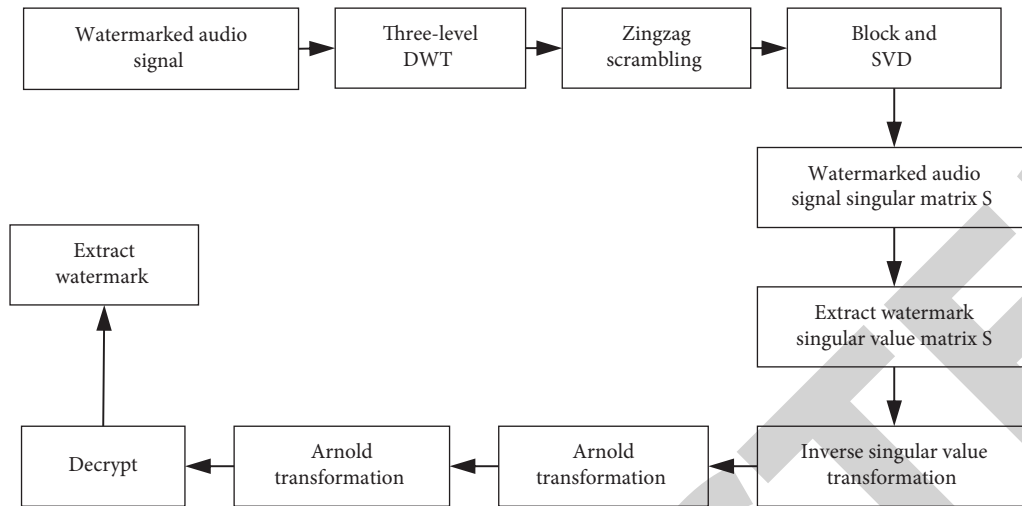


FIGURE 2: Efficient detection of big data in large visual multimedia networks.

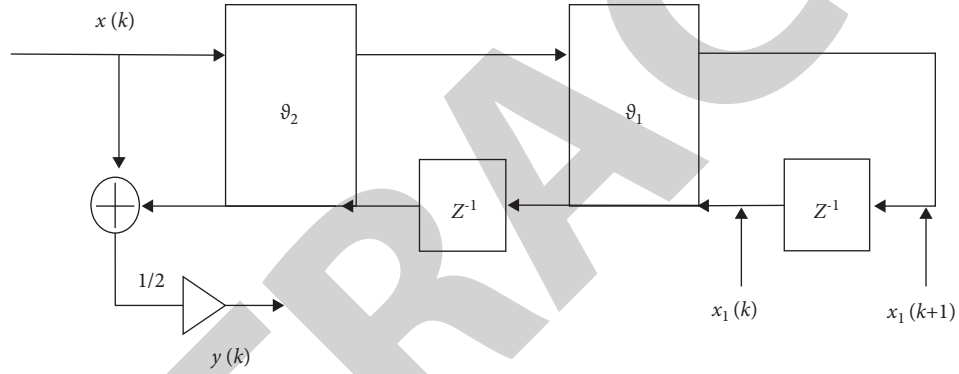
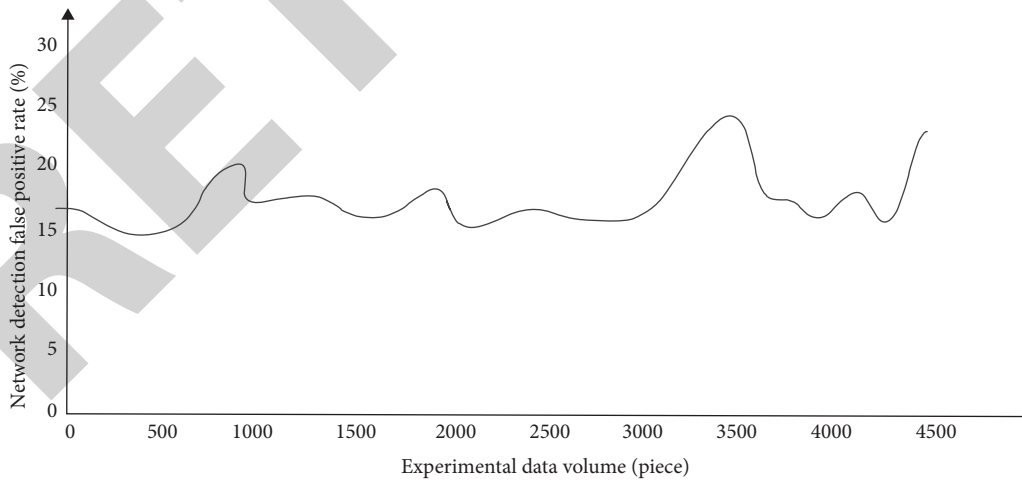


FIGURE 3: FIR filter.



(a)

FIGURE 4: Continued.

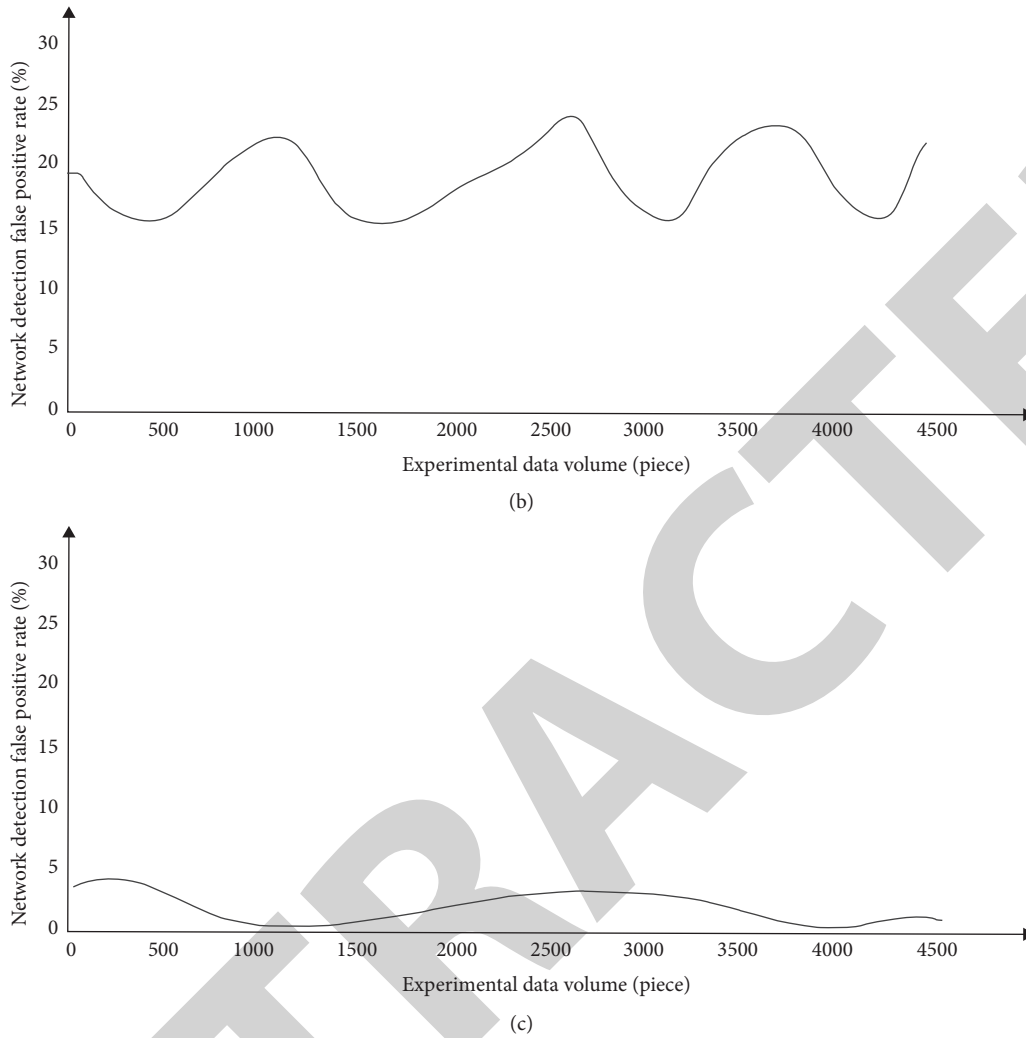


FIGURE 4: Comparison of false positive rates in different methods. (a) False positive rate of network anomaly detection based on information gain. (b) False positive rate of data detection based on migration technology and D-S theory. (c) False positive rate of data detection based on rule-extracting matrix algorithm.

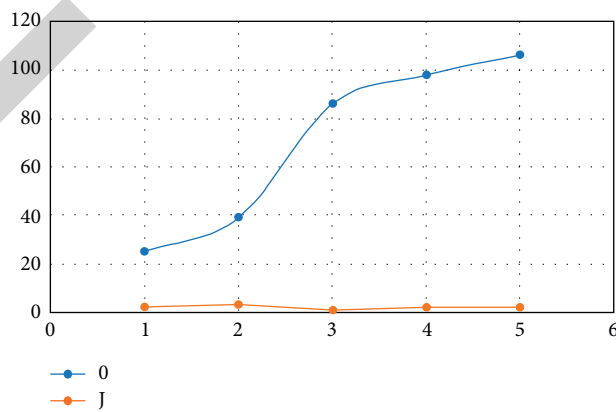


FIGURE 5: Comparison of missed detection by different methods.

detection methods based on data mining and information extraction; and J represents the number of missed detections of the multimedia network negative information data detection method based on the rule-extracting matrix algorithm.

Analyzing Figure 5, the missed detection rate of the network anomaly detection method based on data mining and information extraction is about 23.6%. The missed detection rate of the multimedia network negative

information data detection method based on the rule-extracting matrix algorithm is about 0.67%. The comparison of the above data shows that the proposed method has a low probability of missed detection and is practical.

6. Conclusions

Multimedia network is a double-edged sword. It can convey both positive and negative information. It can guide the public to judge something and also greatly affect the harmonious and stable development of society. Data detection network is used to remove the interference of the network's bad factors, which can improve the detection efficiency and reduce the missed detection rate. Experimental comparison shows that the rule matrix algorithm proposed in this paper is highly reliable and scientific.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The author declares that there are no conflicts of interest.

References

- [1] B. Hou, C. Hou, T. Zhou, Z. Cai, and F. Liu, "Detection and characterization of network anomalies in large-scale RTT time series," *IEEE Transactions on Network and Service Management*, vol. 18, no. 1, 2021.
- [2] C. Benhamed, S. Mekaoui, and K. Ghomid, "Large-scale ip network data analysis for anomalies detection thanks to svm," *International Journal of Design & Nature and Ecodynamics*, vol. 11, no. 3, pp. 376–386, 2016.
- [3] M. L. Shyu, Y. Yan, and J. Chen, "Efficient large-scale stance detection in tweets," *International Journal of Multimedia Data Engineering & Management*, vol. 9, no. 3, pp. 1–16, 2018.
- [4] A. Abrardo, M. Martalò, and G. Ferrari, "Information fusion for efficient target detection in large-scale surveillance wireless sensor networks," *Information Fusion*, vol. 38, pp. 55–64, 2017.
- [5] M. Mardani, "Leveraging sparsity and low rank for large-scale networks and data science," *Dissertations & Theses Gradworks*, vol. 3, no. 2, pp. 1–8, 2015.
- [6] D. C. V. F. Julio, J. Seixas, and G. L. Miotto, "Detection of data taking anomalies for the ATLAS experiment," *China Water & Wastewater*, vol. 3, no. 1, pp. 42–50, 2015.
- [7] S. Kanarachos, S.-R. G. Christopoulos, A. Chroneos, and M. E. Fitzpatrick, "Detecting anomalies in time series data via a deep learning algorithm combining wavelets, neural networks and Hilbert transform," *Expert Systems with Applications*, vol. 85, no. 11, pp. 292–304, 2017.
- [8] S. Papadopoulos, A. Drosou, and D. Tzovaras, "A novel graph-based descriptor for the detection of billing-related anomalies in cellular mobile networks," *IEEE Transactions on Mobile Computing*, vol. 15, no. 11, pp. 2655–2668, 2016.
- [9] V. A. Skazin, A. V. Pavlychev, and S. S. Zotov, "Detection of network anomalies in log files using machine learning methods," *IOP Conference Series: Materials Science and Engineering*, vol. 1069, no. 1, pp. 120–126, 2021.
- [10] Z. Tang, Z. Chen, Y. Bao, and H. Li, "Convolutional neural network-based data anomaly detection method using multiple information for structural health monitoring," *Structural Control and Health Monitoring*, vol. 26, no. 1, pp. 1–12, 2019.
- [11] H. Kasai, W. Kellerer, and M. Kleinstauber, "Network volume Anomaly detection and identification in large-scale networks based on online time-structured traffic tensor tracking," *IEEE transactions on network and service management*, vol. 13, no. 3, pp. 636–650, 2016.
- [12] G. Al Naymat, M. Al Kasassbeh, and E. Al Harwari, "Using machine learning methods for detecting network anomalies within SNMP-MIB dataset," *International Journal of Wireless and Mobile Computing*, vol. 15, no. 1, pp. 67–76, 2018.
- [13] D. Vallejo-Huanga, M. Ambuludi, and P. Morillo, "Empirical exploration of machine learning techniques for detection of anomalies based on NIDS," *IEEE Latin America Transactions*, vol. 19, no. 5, pp. 772–779, 2021.
- [14] N. R. Storozhenko, A. I. Goleva, D. A. Tunkov, and V. I. Potapov, "Modern problems of information systems and data networks: choice of network equipment, monitoring and detecting deviations and faults," *Journal of Physics: Conference Series*, vol. 1546, no. 1, pp. 120–130, 2020.
- [15] L. Dimitriou and C. Antoniou, "Monitoring social network formation and information content analysis of transport anomalies: the case of airline crashes," *Journal of Air Transport Management*, vol. 65, no. 10, pp. 127–141, 2017.