Research on Big Data Information Big Model Processing System of IoT Under Computer Artificial Intelligence Technology

Ren Qiong, Xi Hu, and Junming Chang(✉)

School of Artificial Intelligence, Jianghan University, Wuhan 430056, Hubei, China
qiongren@jhun.edu.cn, cjm72@163.com

Abstract. This paper intends to use cloud computing technology combined with multi-source heterogeneous data to study extensive data analysis and modeling methods for dense environments. This paper aims to improve the mining and detection of massive Internet of Things (IoT) big data in the cloud environment. Firstly, relevant statistical characteristics and correlation rules are extracted from massive IoT big data. Secondly, a multi-source heterogeneous network model based on block is proposed. This paper uses a multi-source isomer model to process the collected data, and it proposes a semantic ontology decomposition method for big data in dense IoT scenarios in a cloud environment, and establishes its association rule knowledge base. Meanwhile, the multi-source heterogeneous information transmission mechanism, and then the dense IoT in the cloud environment is analyzed and mined with big data. Experiments show the proposed algorithm performs better anti-jamming when applied in dense IoT environments. This method has better mining accuracy and less time cost.

Keywords: Cloud Computing · Internet of Things · Intensive Scene · Big Data Mining

1 Introduction

Due to the wide use of IoT technology in industry, agriculture and other fields, considerable data compression and anomaly detection have become the key technologies. Researchers use efficient data processing and analysis technology to make the real-time accuracy of IoT applications, reduce the system's consumption of resources and improve its application effect. it causes a lot of resource waste and data processing overhead [1–5]. It cannot solve the problem of effective and accurate analysis and processing of big data. In recent years, some researchers have used the idea of machine learning to propose an incremental data mining algorithm to obtain more profound knowledge. Some researchers use the information of GRNN-DBSCAN to establish the corresponding detection data analysis model to improve detection accuracy. Some researchers take the listed company of a blockchain in a particular region as a case, based on the principal component cluster analysis method, to carry out a specific analysis of the company's

X. Chen et al. (Eds.): IoTaaS 2023, LNICST 585, pp. 304–314, 2025.
https://doi.org/10.1007/978-3-031-70507-6_23

data. After all kinds of data are processed by clustering and classified, the required data can be quickly found from a large amount of data. This improves work efficiency. Studies have used blockchain technology to identify IoT devices and ensure that the devices are immutable [6]. It uses the hash table model based on "allowlist" to encrypt the device to achieve the security of the device. Previous studies have selected the privacy data of IoT with the most significant amount of information as the learning sample, then classified it by using differential privacy technology, and finally established the security protection model of the privacy data of IoT by using linear regression and other algorithms to achieve the security protection of IoT. Some researchers have studied large-scale data acquisition and analysis technology for large-scale IoT to improve system performance, but there are some problems, such as poor anti-multi-path interference performance. This paper will analyze big data and IoT-dense scenarios in cloud computing environments. It mainly includes.

(1) Extracting statistical feature quantity and association rule feature quantity from massive data;
(2) Establishing a knowledge base of association rules for massive data;
(3) Giving the correlation degree transmission and transmission mechanism of massive data;
(4) Carry of extensive data analysis and processing of data, the effectiveness of the proposed algorithm to improve the efficiency of dense IoT extensive data mining in the cloud environment is verified.

2 Design of an Data Processing Systems in IoT

As a service computing architecture for IoT, cloud computing has the characteristics of "quasi-virtual" between cloud and individual computing [7]. It dramatically improves computing speed by moving some of the cloud's work tasks to the cloud. A general IoT data processing architecture is shown in Fig. 1 (image cited in Data Flow: From the Edge to the Server/Cloud).

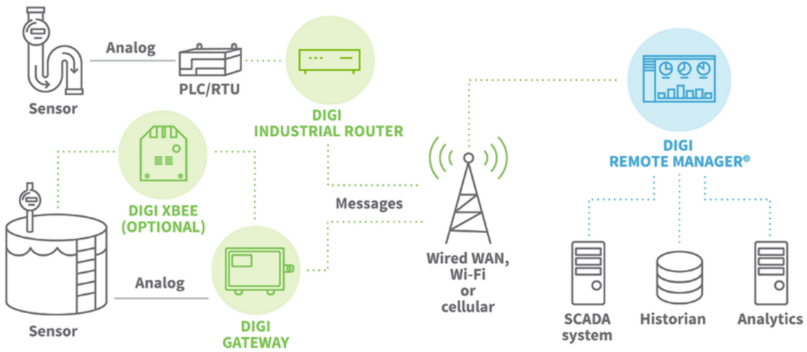

Fig.1. Basic data processing architecture of IoT.

As the amount of data increases, the data processing architecture in Fig. 1 causes the number of resources on the device to increase accordingly, which leads to lower data computation speed and higher resource consumption costs [8]. This paper intends to propose A Data analysis architecture combining offline and real-time analysis based on cloud Computing theory (Fig. 2 is quoted in A Novel Fog Computing Enabled Temporal Data Reduction Scheme in IoT Systems).

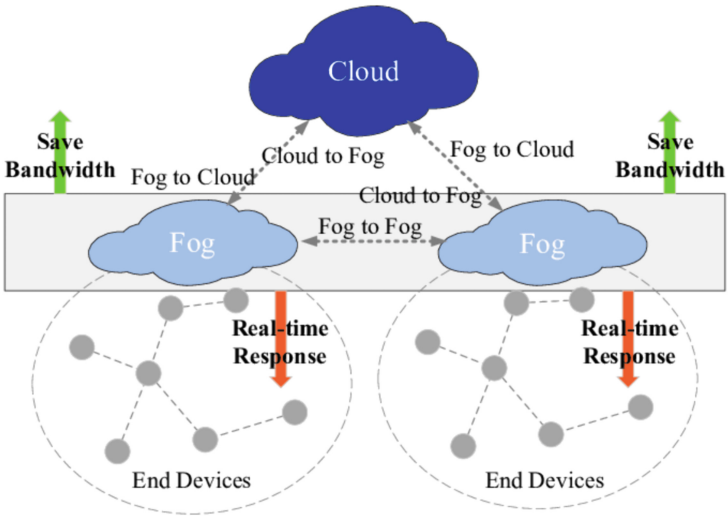

Fig. 2. Architecture of offline and real-time analysis based on fog computing.

As you can see from Fig. 2, this data analysis architecture is divided into two types: offline and just-in-time. The so-called offline analysis uses offline learning and correcting the obtained data [9]. After repeated learning, a revised model is finally obtained. Input the new model into the real-time analysis system to get the real-time data, and then the data comprehensive analysis and the final output results can be.

3 Mass Data Acquisition and Preprocessing in a Dense Environment

3.1 Extensive Data Sampling for Iot Intensive Scenarios

This paper takes multi-source heterogeneous information as the research object and wireless RFID and multi-source information fusion as the core to study the optimal sampling and feature classification of massive information for the large-scale IoT. Build a big data acquisition model in a dense environment. It uses WSN technology to construct RFID information collection tags. This paper intends to research statistics and large-scale IOT big data modeling based on the ZigBee network. The fusion scheduling method of considerable data feature extraction and data analysis for dense environments is studied. Firstly, $Y = \{y_1, y_2, \cdots, y_n\}$ two-valued semantic decision model for large-scale extensive data

collection oriented to IoT is established, and its association map is established by taking the ZigBee network node as the networking network [10]. Secondly, the multi-source information transfer mechanism based on Fuzzy is established in the large-scale ample data distribution space, and finally, the multi-source information fusion mechanism of extensive data application oriented to cloud computing is obtained. The associated map will appear $\psi : P \rightarrow R^{2d+1}$. $\psi(c) = (g(c), g(\nu_1(c)), \cdots, g(\nu_{2d}(c)))^T$ is the node distribution group of a sample. For large-scale iot extensive data collection, its label distribution group is as follows:

$$\begin{aligned} y_i^{(t+1)} &= (1-\lambda)y_i^{(t)} + \frac{\lambda}{\delta_{ni}}(\zeta_i - \sum_{j=1}^{i-1}\delta_{ij}y_j^{(t+1)} - \sum_{j=i+1}^{n}\delta_{ij}y_j^{(t)} \\ i &= 1, 2, \cdots, n \\ t &= 1, 2, \cdots, n \end{aligned} \tag{1}$$

The feature classification is carried out based on the type attribute $\sigma_i (i = 1, 2, \cdots, n)$ of the big data in the dense scene of IoT, Hence $H = \{Y_1, Y_2, \cdots, Y_n\}$ is a vector distribution set on U. Extensive data collection in dense big data processing based on the node distribution model of extensive data sampling in dense scenarios. The networking model of its wireless sensing network nodes is shown in Fig. 3 below:

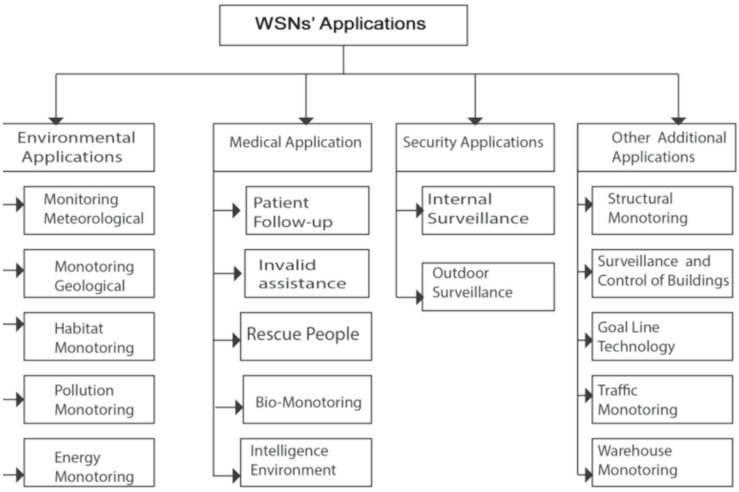

Fig. 3. Networking model of IoT WSN nodes for large-scale data collection.

The data is collected and extracted based on the wireless sensor network nodes for big data acquisition in the dense scene of IoT, as shown in Fig. 3.

3.2 Feature Extraction

This paper intends to study the multi-source heterogeneous information processing technology based on block, feature classification, and adaptive scheduling for the mass IoT [11]. In the process of using feature vectors to extract big data for analysis, this paper combines the use of IoT tools such as data mining, which plays a crucial role in processing a large amount of data.

$$\begin{aligned} s(t) &= \sum_{i} \sum_{j=0}^{N_f-1} \sum_{k=0}^{K-1} \zeta_i \delta_k q(t - iT_s - jT_f - \sigma_j T_\sigma - \chi_k) + \\ \lambda(t) &= \sum_{i} \sum_{j=0}^{N_f-1} \zeta_i q_n(t - iT_s - jT_f - \sigma_j T_\sigma - \chi_0) + \lambda(t) \end{aligned} \tag{2}$$

Among them,

$$q_g(t) = \sum_{k=0}^{K-1} \delta_k q(t - \chi_{k,0}) \tag{3}$$

$\lambda(t)$ is a fuzzy weight extracted from the number of statistical features in the massive IoT big data. This paper intends to conduct distributed fusion research on the statistical feature quantity of large-scale iot big data based on fuzzy correlation degree constraints. This paper intends to study the adaptive fusion method. The regular statistical characteristics of mass data in IoT environment of large data sets are:

$$D(\lambda) = \frac{1}{T} \sum_{t=0}^{T-1} D_t(\lambda) \tag{4}$$

To establish a semantic similarity representation method for massive IoT big data in dense environments [12]. The finite data relation point is denoted as $[\delta_j, \zeta_j]$, and the relation point satisfies $y_j \in [\delta_j, \zeta_j]$. The objectives of routing and forwarding control of this integration node are:

$$y(n) = \frac{1}{\sqrt{N}} \Delta \sum_{t=0}^{N-1} Y(t) \exp(j2\pi kn/N), n = 0, 1 \cdots N - 1 \tag{5}$$

where, Δ is the extent of intelligent collection of massive information in a large-scale IoT environment; $\Delta = \{\delta_{i,j}, 0 < i, j < N\}$. $\delta_{i,j}$ is the transport node in the environment. This paper analyzes big data in IoT based on this model, and then conducts data processing on relevant feature vectors.

3.3 Association Rule Knowledge Base of Big Data in IoT Intensive Scenarios

This paper proposes an extensive data mining method for IoT-dense scenes based on cloud computing information fusion and fuzzy clustering and uses radio frequency tag

identification technology to extract intelligent features from big data for IoT-dense scenes [13]. IoT dense scene big data multi-source distribution model is established:

$$E = [x_1, x_2, \cdots, x_d] \tag{6}$$

Dimensionality reduction uses multi-dimensional scaling characteristics to reduce dimension m to dimension d. The big data fusion mode in dense scenarios is obtained:

$$\begin{aligned} &\max H(Y) = (H_1(Y), H_2(Y), \cdots, H_n(Y)) \\ &s.t.\, g_j(Y) \le 0 (j = 1, 2, \cdots, q) \\ &g_t(Y) = 0 (t = 1, 2, \cdots, q) \end{aligned} \tag{7}$$

Realize multi-source heterogeneous information fusion in dense IoT environment. The expression of i in IoT is obtained:

$$Racall(Y, X) = \frac{Q(Y \cap X)}{Q(Y) + Q(X) - Q(Y \cap X)} \tag{8}$$

$$Overload(Y, X) = \frac{Q(Y \cap X)}{\min(Q(Y), Q(X))} \tag{9}$$

$$Time(Y, X) = \frac{2Q(Y \cap X)}{Q(Y) + Q(X)} \tag{10}$$

where, $Q(Y)$, $Q(X)$ represents the degree of correlation and integration between large-scale and high-density iot big data. Y, X is a jointly assigned feature vector set for large-scale iot big data. $Q(Y \cap X)$ is a cross-distributed collection of massive information in a dense environment [14]. It is assumed that the output parameter of the extensive data association knowledge base in IoT-intensive scenario in the cloud computing center is $param = \{G_1, G_2, e, g, g_2, g_3, h, H_1, H_2\}$, and the time interval of collection of the extensive data association rule base is T_f combined with the resource fusion scheduling method [15]. This paper presents a multi-objective optimization algorithm based on $T_s = N_f T_f$ function. The matching degree of large data mining in a dense iot environment is:

$$T_\sigma = ent(T_f / N_\sigma) \tag{11}$$

The feature distribution of the critical index of dense scene big data is set to $Y(y_1, y_2 \cdots y_D)$, and the distribution structure of knowledge of mining association rules of iot dense scene big data obtained in a restricted space makes it meet $\sigma_j T_\sigma < T_f, \forall j \in [0, N_f - 1]$.

3.4 Extensive Data Analysis in the Dense Environment of IoT Based on Cloud Technology

In this paper,a knowledge base of association rules for massive IoT data is established, and an association transmission control mechanism for massive IoT data is proposed based on the fusion analysis of association knowledge [16]. The calculation formula of the edge property distribution matrix $Z_{N\times 1}$ is as follows:

$$Z_{N\times 1} = D_{N\times K} \cdot T_{K\times 1} \tag{12}$$

The exact transfer probability of the ontology resource m is given to determine $e_t \geq 0, \sum_{t=1}^{t} e_t = 1$. The discrete distribution model of extensive data mining in a dense environment is constructed by using fuzzy decision variables

$$s(t) = [s_1(t), s_2(t), \cdots, s_m(t)]^T \tag{13}$$

The feature mapping method of semantic ontology is applied to carry out adaptive mining of massive data in dense environments, and its spatial distribution matrix is as follows:

$$Y' = \begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ y_{(n-1)1} & y_{(n-1)2} & \cdots & y_{(n-1)n} \end{pmatrix} \tag{14}$$

The plaintext block feature quantity of big data in iot intensive scenarios is calculated [17]. In combination with the routing and forwarding control protocol of IoT, the association and fusion factor δ^i_{desira} of the big data model is obtained

$$\delta^i_{desira} = \delta_1 \cdot \frac{Den\,sit\,x_i}{\sum\limits_i Den\,sit\,x_i} + \delta_2 \frac{\Delta Q_i}{\Delta Q_{in\,it}} \tag{15}$$

Among them,

$$\begin{cases} \delta_1 + \delta_2 = 1, \delta_1, \delta_2 \in [0, 1] \\ \delta_2 = \dfrac{\max\limits_i(\Delta Q_i) - \min\limits_i(\Delta Q_i)}{\Delta Q_{in\,it}} \end{cases} \tag{16}$$

Considered the equivalent semantic mapping, the link set of extensive data distribution in IoT-dense scenarios satisfies $Q \in R^{n\times n}$, $R \in R^{m\times m}$ and $H \in R^{m\times m}$, and the feature distribution set satisfies $d \sim q(e, q)$.

Combined with the matching index set $E_t \in E(t = 1, 2, \cdots, t)$ of extensive data integration in IoT intensive scenarios, the optimized integrated graph model is $Q_i \in Q(i = 1, 2, \cdots, m)$. Correlation detection and mining of iot data are carried out according to the fusion results of correlation knowledge.

4 Experimental Results and Analysis

The correctness of IoT data privacy protection model in a heterogeneous cloud computing environment is tested by experiments on information loss degree, data availability, performance and security. The results of the experiment are shown below. The test compares the algorithm in literature [18] and literature [19] and the method in this paper.

4.1 Degree of Information Loss

In protecting private data,the papermust ensure the degree of information loss is shallow. Only in this way can the model have practical application value. For different IoT data, three ways are adopted to protect it from information loss. It is determined that the degree of information loss in this way is the smallest, and the optimal privacy data protection model is obtained. Figure 4 shows the five groups of experimental results. The algorithm can ensure the performance of privacy data protection. The information lost by the other two methods is much higher than that of the proposed methods, so the data protection model will likely become meaningless [20]. The low information loss of the proposed method is due to the preprocessing of the original data before establishing the protection model. At the same time, the algorithm can significantly ensure the integrity of the data, thus reducing the amount of information loss.

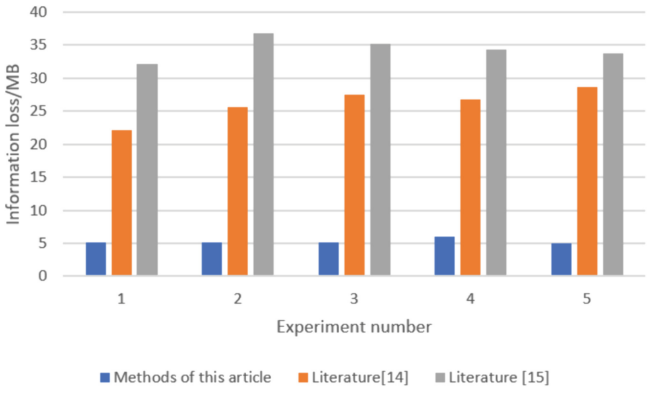

Fig. 4. Amount of information loss under the three algorithms.

4.2 Data Availability

In the privacy data protection model, the amount of data is a direct factor, and as the amount of data increases, data availability will also increase [21]. The comparison results are shown in Fig. 5. It can be seen that the method proposed here is the strongest under what kind of data volume, while the other two methods are not as good as the algorithm proposed here to some extent, so its correctness can be demonstrated.

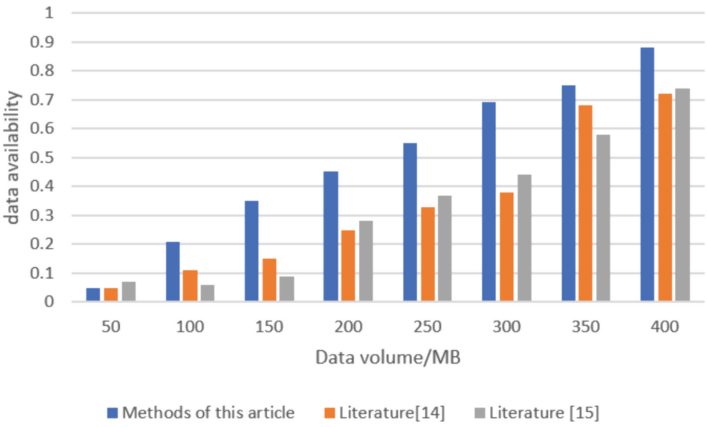

Fig. 5. Data availability for the three methods.

4.3 Model Protection Performance

By comparing the upload rate of the above three schemes with that of the unencrypted one, the scheme that is closest to the unencrypted one is obtained. The results are shown in Fig. 6. From Fig. 6, it can be seen that the method described in this paper has the closest upload speed to unencrypted data, indicating that it is the most effective of the three methods and confirming the overall efficiency of the method described in this paper.

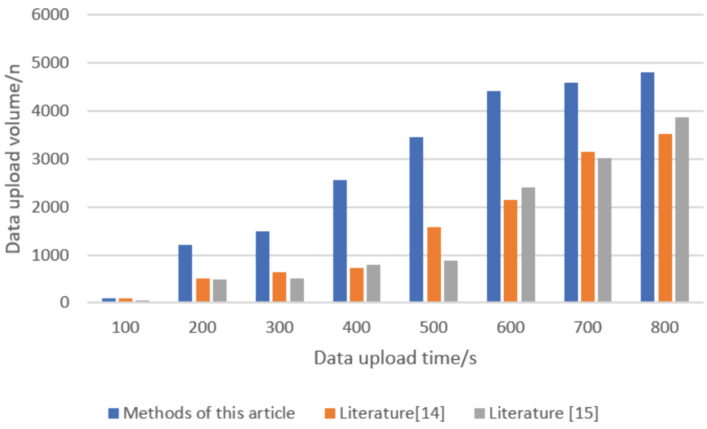

Fig. 6. Model protection performance of the three methods.

4.4 Security

The security problem in cryptography is evaluated by measuring the rate of change of ciphertext in cryptography. Assuming that the number of plaintexts is fixed in a cipher system, if any cipher system changes, the ciphertext change rate under the cipher system can be obtained, and then the propagation characteristics under the cipher system can be obtained. The security of the cryptosystem is judged. Table 1 lists the comparative data of the three experimental methods. The stronger the ciphertext change rate, the stronger the data diffusion.Table 1 can display the following data of the highest average ciphertext change rate, which further proves the security of the proposed method.

Table 1. Diffusivity of the three methods.

Experiment number	The ciphertext change rate of the proposed method	Reference [18] method ciphertext change rate	Reference [19] method ciphertext change rate
1	0.83	0.78	0.66
2	0.87	0.76	0.41
3	0.80	0.66	0.46
4	0.76	0.60	0.42
5	0.48	0.75	0.86
6	0.97	0.86	0.45
7	0.86	0.73	0.19
8	1.00	0.83	0.57
9	0.99	0.48	0.43
10	0.86	0.78	0.44
Mean value	*0.84*	*0.72*	0.49

5 Conclusion

This paper intends to use cloud computing technology combined with multi-source heterogeneous data to study extensive data analysis and modeling methods for dense environments. This paper intends to use WSN technology to build RFID data collection identification. At the same time, based on the ZigBee networking protocol, this paper conducts statistical analysis and fuzzy sampling research on massive IoT big data and studies the feature classification and adaptive scheduling of massive IoT big data through zoning fusion and fuzzy clustering technologies. At the same time, the high-density IoT extensive data mining method based on cloud services is studied. Through analysis, it is found that the algorithm has better accuracy, less time, and better overall performance in the dense, extensive data mining of IoT.

References

1. Liu, G., Rong, K., Tang, L., et al.: Research on technology integration and application of Shenzhen urban ecological Big Data intelligent management and service platform. Acta Ecol. Sin. **42**(4), 98–104 (2022)
2. Xiong, X., Xi, H., Guo, H.: A hybrid optimized grey seasonal variation index model improved by whale optimization algorithm for forecasting the residential electricity consumption. Energy **234**, 121127 (2021)
3. Xiong, X., Xi, H., Tian, T., Guo, H., Liao, H.: A novel Optimized initial condition and Seasonal division based Grey Seasonal Variation Index model for hydropower generation. Appl. Energy **328**, 120180 (2022)
4. Li, Z., Xi, H., Guo, H., Xiong, X.: A novel Weighted Average Weakening Buffer Operator based Fractional order accumulation Seasonal Grouping Grey Model for predicting the hydropower generation. Energy **277**, 127568 (2023)
5. Xi, H., Xiong, X., You, W., Shi, M., Wei, P., Ma, C.: A hybrid clustered SFLA-PSO algorithm for optimizing the timely and real-time rumor refutations in online social networks. Expert Syst. Appl. **212**, 118638 (2023)
6. Zheng, P.: Design of high precision ultrasonic gas flow monitoring system based on IoT. Instrum. Technol. Sens. **10**(2), 65–70 (2021)
7. He, S., He, X., Song, D., et al.: Multi-parameter integrated warning model of rock burst and intelligent identification cloud platform. J. China Univ. Min. Technol. **51**(5), 130–133 (2022)
8. Zhang, Y.-L., Wang, S., Ye, Z.: Industrial equipment based on IoT remote PLC control system design. Autom. Technol. Appl. **42**(3), 8–10 (2022)
9. Cheng, X., Lang, G.: Secure centralized storage of medical big data information based on IoT. Inf. Technol. **47**(1), 109–114 (2021)
10. Ding, G., Chen Qihang, X., Chen, et al.: Model sharing for GPU-accelerated DNN inference in big data processing systems. J. Tsinghua Univ.: Nat. Sci. Edit. **62**(9), 78–82 (2022)
11. Li, G.: Design of IoT Access identity security authentication system based on big data. Autom. Technol. Appl. **42**(4), 118–121 (2022)
12. An, Y., Zhu, Y., Wang, J.: Analysis of applicable conditions of distributed hash table in big data scenario of IoT. J. Comput. Sci. **44**(8), 170–177 (2021)
13. Li, Z., Geng, X., Kaiyong, X., et al.: Operation management practice of laboratory animal center based on IoT and big data technology. J. Exp. Anim. Sci. **38**(4), 22–29 (2022)
14. Liu, H.: Application analysis of embedded software based on big data in IoT technology. Microcomput. Appl. **39**(4), 195–198 (2021)
15. Luo, X., Pang, Z., Tan, S., et al.: Design and practice of water big data platform system based on cloud computing. Water Suppl. Drain. **48**(1), 17–26 (2022)
16. Yan, P., Zhou, L., Yan, H.: Research on IoT privacy data protection model under hybrid cloud storage. Comput. Simul. **40**(2), 530–534 (2022)
17. Weimin Guo, H., Zhao, J.Z., et al.: Big data analysis of Henan tobacco curing process based on IoT data acquisition technology. Tob. Sci. Technol. **54**(9), 92–99 (2021)
18. Huang, Z., Yang, J., Zhang, Y., Yin, Z.: Reputation evaluation model of iot data service based on blockchain. Comput. Eng. **48**(1), 33–42 (2022)
19. Rui, H., Rui, Z., Dong, X., et al.: Design and application of intelligent power system based on multi-sensor fusion. Mod. Electr. Technol. **44**(7), 4–8 (2021)
20. Li, Y., Wang, H., Fang, M.: Design and application of IoT electrical safety system in university laboratory. Mod. Electr. Technol. **46**(10), 66–70 (2022)
21. Chao, X., Guo, F., Chuankun, W.: MQTT-SE data encryption transmission algorithm. Appl. Comput. Syst. **31**(12), 169–177 (2022)