Research on Computer Simulation Big Data Intelligent Collection and Analysis System

Hongying Liu*

Department of Computer Science and Engineering, Guangzhou College of Technology and Business, Guangzhou 510850, China

*Corresponding author email: hongyingliu@gzgs.edu.cn

Abstract. As a characteristic of big data, the individual data in it is no longer isolated, and the data and its underlying mechanisms have complex associations, which make all data into an indivisible whole. The dynamic generation and disappearance of data will change its original relationship and affect the overall characteristics of the data. This feature of big data makes the subject-oriented analysis methods such as data mining present limitations: the presupposition of the subject and the analysis by subject split the interaction relationship between the subjects, leading to the loss of the implicit mechanism in these relationships. Aiming at the problems of traditional network big data multi-resolution acquisition methods such as high acquisition cost, long completion time and low acquisition accuracy, a JA-va3D-based big data network multi-resolution acquisition method is proposed, and average interactive data is introduced. The extraction method estimates the power spectral density of the network data multi-resolution acquisition, and uses the ADASYN algorithm to remove the invalid multi-resolution data, and realizes the large data multi-resolution accurate acquisition. Experimental results show that the proposed method has lower acquisition cost, shorter completion time, and higher acquisition accuracy; it has certain practical value and can be widely used in various fields.

Keywords. Computer, simulation data, intelligent algorithm, collection and analysis system.

1. Introduction

The emergence of technologies such as cloud computing, mobile communications and big data has promoted the rapid development of various application software, and has been widely used in logistics and warehousing, power communications and smart tourism. The database in cloud computing is the most critical part of application software, and it is also the starting point and end point of application software data processing and processing. Therefore, database design is an important factor affecting the use and popularization of application software. At present, in the process of using the database, as the number of visits by users gradually increases, the scale of big data has also become huge [1]. The traditional database design model is prone to load phenomena, which is not conducive to improving the efficiency of database extraction, so a new intelligent storage method needs to be added.
At present, many experts and scholars are conducting research on network multi-resolution big data collection. For example, the multi-resolution collection method of network data based on linear regression, through the application of linear regression analysis method to construct a sensing data model, maintain the characteristics of the sensing data, so that the node only transmits the parameter information of the regression model, instead of transmitting the actual monitored sensing data information. The multi-resolution feature of network data is guaranteed to complete the collection, but this method has many imperfect legacy problems. Multi-resolution acquisition method based on network big data migration. Setting up an experimental environment and performing full data migration can remove invalid data to improve the accuracy of network data multi-resolution acquisition, but this method takes a lot of time in the acquisition process. This paper proposes a method for rapid information acquisition based on Ja-va3D. This method analyses the composition of the information by studying the operating rules and characteristics of the information, and encodes and obtains the information, and adds depth to the information [2]. In order to standardize the information, in order to complete the quantitative extraction of information data and complete the rapid acquisition of information, the process of this method is more complicated and the efficiency is low.

2. Principles of Big Data Intelligent Fusion

In the big data fusion principle, the data set is obtained first, and the data set is divided into several subsets. Then cluster several subsets at the same time to get several cluster centres. Judge whether the sum of the cluster centres is less than the threshold of the data fusion problem scale, if the result is yes, cluster the data whose number is the value obtained by the sum of the data cluster centres, merge the obtained categories, and complete the data fusion [3].

To better distinguish the two sequences of abnormal big data and normal big data. The big data in cloud computing needs to be divided into two segments, and then stored in two buffers of different sizes [4]. The short buffer area can store approximately 23-63 large data sequences, while the long buffer area can store approximately 243 large data sequences. In order to strengthen the storage capacity of the research method, the sequences in the long and short buffers are selected in the appropriate state for the buffer sample set. When new big data is searched, it is stored in the short buffer area. If the short buffer area is full, the first data stored in the short buffer area is stored in the long buffer area. If the long buffer area is full, delete the big data stored in the long buffer area first. The processing process is shown in Figure 1.

![Figure 1. Long and short cache structure](image-url)
2.1. Related theorems

Assume that for the same target, the initial state estimation and covariance matrix of sensors \(i\) and \(j\) are \(\hat{x}_{0|i}^m\) and \(P_{0|i}^m\), \(m = i, j\) respectively. The dynamic equation of target movement is:

\[
x_{k+1} = \Phi_k x_k + \Gamma_k w_k
\]

(1)

Among them, the process noise \(w_k\) is white noise with zero mean value, the covariance matrix is \(Q_k\), and the measurement equations of the two sensors are

\[
z_{k}^m = H_{k}^m x_k + v_k^m, m = i, j
\]

(2)

Among them, the measured noise is white noise with zero mean, the covariance is \(R_k^m\), and the cross-covariance matrix is \(R_i^j\).

The state of sensor \(i\) at time \(k\) is estimated as:

\[
\hat{x}_{k|k}^i = \Phi_{k} \hat{x}_{k-1|k-1}^i + K_k^i (z_k - H \Phi_{k} \hat{x}_{k-1|k-1}^i)
\]

(3)

The estimated error is:

\[
\hat{x}_{k|k}^i = x_k - \hat{x}_{k|k}^i = \Phi_{k} \hat{x}_{k-1|k-1}^i + \Gamma_k w_k - \Phi_{k} \hat{x}_{k-1|k-1}^i - K_k^i (z_k - H \Phi_{k} \hat{x}_{k-1|k-1}^i) = \Phi_{k} \hat{x}_{k-1|k-1}^i + \Gamma_k w_k - K_k^i z_k
\]

(4)

Cross-covariance:

\[
P_k^i = E[\hat{x}_{k|k}^i \hat{x}_{k|k}^i]\]

\[
= P_{k}^i (P_{k|k-1}^i)^{-1} [\Phi P_{k-1}^i \Phi' + \Gamma Q_{k-1} \Gamma'] (P_{k|k-1}^i)^{-1} P_k^i + K_k^i R_i^j K_k^j
\]

(5)

It can be seen from the above formula that the cross-correlation is caused by a priori estimation \(P_{k-1}^j\) process noise \(Q_{k-1}\) and measurement noise \(R_i^j\). The concept of consistency: Suppose that the true state of the target is \(a\), the sensor’s estimated state of the target is \(\hat{a}\), the estimated error covariance is \(P\), and the true error covariance is \(\bar{P} = E(\hat{a} - a)(\hat{a} - a)'\). The so-called consistency is the estimated covariance \(P \geq \bar{P}\).

2.2. Fusion algorithm for correlation estimation

Assuming that the true correlation system between two local estimates is \(\gamma_0\), we can determine a \(\gamma_{\text{min}}\) and \(\gamma_{\text{max}}\) such that \(\gamma_0 \in (\gamma_{\text{min}}, \gamma_{\text{max}})\), and can obtain an estimate \(\hat{P}_{ab}\) of the cross-covariance matrix that satisfies the following conditions, that is, \(P_{ab}, \hat{P}_{ab}\) satisfies:
\[
\gamma_{\min}^2 P_{bb} \leq P_{ab}^{-1} P_{aa}^{-1} P_{ab} \leq \gamma_{\max}^2 P_{bb}
\]

(6)

Under such conditions, making full use of these correlation information can improve the fusion accuracy. Algorithm flow:

After obtaining the estimates of \( \gamma_{\min} \), \( \gamma_{\max} \), and \( \hat{P}_{ab} \), define the joint covariance matrix:

\[
P' = \begin{bmatrix}
P_{aa} & \hat{P}_{ab} \\
\hat{P}_{ab} & P_{bb}'
\end{bmatrix}
\]

(7)

Use formulas (8) and (9) to determine the fusion weight for fusion. In the literature, the estimation method of the correlation coefficient bound is given. The cross-covariance matrix can be estimated by formula (5), and the estimated \( \hat{P}_{ab} \) must satisfy

\[
\gamma_{\max}^2 P_{bb} \leq \hat{P}_{ab}^{-1} P_{aa}^{-1} \hat{P}_{ab} \leq \gamma_{\max}^2 P_{bb}
\]

(8)

3. Experimental simulation analysis

3.1. Experimental environment construction

In order to verify the overall effectiveness of the big data intelligent fusion algorithm based on the estimation mechanism, a simulation experiment is needed. Experimental data: Take a certain city centre as the local area network within 2km of the circle. Experimental environment: 8 machines are used to correspond to the data stored in each cell. The configuration of each machine is: CPU is i5-2400, 3.1GHz; memory is 8G, operating system is Win10 Ultimate. The experimental indicators are the lifetime of network nodes after data fusion and the network energy consumption during the fusion process. Among them, network node lifetime: After data fusion, it will be divided into different types of data nodes \([5]\). As time goes by, a large amount of data will flow into each node. The current data fusion node will be washed away, and the current data fusion time from when a node is established to when it is washed away is the network node lifetime.

3.2. Experimental analysis

The proposed Java3D-based network multi-resolution large data acquisition method, the optical fibre network communication data resolution acquisition method and the network data resolution acquisition method based on the linked list structure are compared with the results of the completion time of the network data multi-resolution acquisition. The unit of completion time is seconds, which is represented by s. The experimental results are shown in Figure 2. In Figure 2, A represents the proposed method; B represents the data resolution acquisition method based on optical fibre network communication; C represents the network data resolution acquisition method based on the linked list structure \([6]\).
The network data resolution acquisition method based on the linked list structure takes the longest time to complete the network data multi-resolution acquisition process; the network data multi-resolution acquisition based on the optical fibre network communication data resolution acquisition method takes longer to complete than the network data based on the linked list structure. The network data resolution acquisition method of the proposed method takes a short time to complete the acquisition, but with the continuous increase of the amount of data, the acquisition completion time has been above 60s; the proposed method has the shortest completion time of the network data multi-resolution acquisition, and increases with the number of data. It keeps increasing, has been below 60s, and has a high collection efficiency. The proposed method is compared with the acquisition cost (yuan) based on the optical fibre network communication data resolution acquisition method and the network data resolution acquisition method based on the linked list structure [7]. The experimental results are shown in Figure 3.
In Figure 3, A represents the proposed method; B represents the data resolution acquisition method based on optical fibre network communication; C represents the network data resolution acquisition method based on the linked list structure.

Obtaining information according to the characteristics of user information is the basis for the rapid information acquisition method under big data analysis to obtain user information. The parameter $A_k$ represents the characteristics of user information. When the characteristics of user information are larger, user information is easier to obtain, to make the results more accurate. The parameter $A_k$ is used to test the rapid information acquisition method. Table 1 shows the test results of the rapid information acquisition method under big data analysis and the feature-based user information rapid acquisition method.

Table 1. Test results of two different methods

|   | SY | A  | B  |
|---|----|----|----|
| 1 | 5  | 4  | 3  |
| 2 | 4  | 3  | 3  |
| 3 | 5  | 3  | 3  |
| 4 | 5  | 3  | 3  |
| 5 | 5  | 2  | 2  |
| 6 | 4  | 2  | 2  |
| PJ| 4.7| 2.8|    |
4. Conclusion
Aiming at the various problems in the traditional Java3D network big data multi-resolution acquisition process, a Java3D-based network big data multi-resolution acquisition method is proposed. This method has shorter completion time for multi-resolution acquisition of network big data, lower cost, higher acquisition accuracy, certain application performance, and can be widely used in various fields.

References
[1] Wang, L., & Wang, G. Big data in cyber-physical systems, digital manufacturing and industry 4.0. International Journal of Engineering and Manufacturing (IJEM), 6(4) (2016) 1-8.
[2] Zhu, L., Yu, F. R., Wang, Y., Ning, B., & Tang, T. Big data analytics in intelligent transportation systems: A survey. IEEE Transactions on Intelligent Transportation Systems, 20(1) (2018) 383-398.
[3] Jung, D., Tran Tuan, V., Dai Tran, Q., Park, M., & Park, S. Conceptual Framework of an Intelligent Decision Support System for Smart City Disaster Management. Applied Sciences, 10(2) (2020) 666-675.
[4] Zhong, R. Y., Xu, C., Chen, C., & Huang, G. Q. Big data analytics for physical internet-based intelligent manufacturing shop floors. International journal of production research, 55(9) (2017) 2610-2621.
[5] Sumalee, A., & Ho, H. W. Smarter and more connected: Future intelligent transportation system. IATSS Research, 42(2) (2018) 67-71.
[6] Zheng, X., Chen, W., Wang, P., Shen, D., Chen, S., Wang, X., ... & Yang, L. Big data for social transportation. IEEE Transactions on Intelligent Transportation Systems, 17(3) (2015) 620-630.
[7] Chih-Lin, I., Sun, Q., Liu, Z., Zhang, S., & Han, S. The big-data-driven intelligent wireless network: architecture, use cases, solutions, and future trends. IEEE vehicular technology magazine, 12(4) (2017) 20-29.