Method for Quickly Obtaining User Browsing Behavior Data Under Cloud Computing

Jinbao Shan(✉) and Haitao Guo

College of Information Technology & Art Design,
Shandong Institute of Commerce and Technology, Jinan 250103, China
maojunl980@126.com

Abstract. Under cloud computing, traditional user browsing behavior data acquisition method cannot optimize data classification, which results in slow and low accuracy of data acquisition. For this reason, a fast method to obtain user browsing behavior data under cloud computing is proposed. Using node processing user browsing behavior data, complete the query the user browsing behavior data collection, provide the conditions for data classification optimization, the data to calculate the similar characteristics after multiple iterations data peak, peak according to complete the user browsing behavior data classification, the classification of output data integration, realize the cloud user browsing behavior fast data acquisition. Compared with the traditional data acquisition method, the data acquisition speed of the design method is increased by 20 min and the accuracy is increased by 45%. The experimental data show that the overall performance of the proposed method is better than the traditional method, and it has strong practicability and high reference value.

Keywords: Cloud computing · User browsing behavior data · Fast data acquisition · Acquisition speed

1 Introduction

In the era of big data, with the rapid development of the Internet, the rapid growth of the number of users browsing will generate a large amount of user browsing behavior data. The traditional method cannot achieve rapid acquisition of massive data. Therefore, the rapid acquisition of user browsing behavior data has become an urgent need to be solved [1].

Reference [2] proposed a mobile Internet big data user behavior analysis engine solution, designed the key modules of user behavior data acquisition system, big data warehousing and preprocessing components and big data user behavior analysis model. It is verified that this method can effectively collect user behavior information, but there is a defect that the collection efficiency is not high when faced with massive data. In the reference [3], an improved hierarchical clustering algorithm CURE is proposed. The algorithm uses the Map Meduce function to parallelize the way of sampling and processing data in the original algorithm. At the same time, combined with the concept of interval number, the mobile user data is represented by an interval, calculate the interval distance to adapt to the uncertainty of mobile user data, thereby improving the

S. Liu and G. Yang (Eds.): ADHIP 2018, LNICST 279, pp. 365–371, 2019.
https://doi.org/10.1007/978-3-030-19086-6_41

accuracy of clustering. Although the algorithm can guarantee the accuracy of clustering, the implementation process of the algorithm is more complicated and not simple enough.

Due to the fact that the traditional data acquisition method has a slower acquisition rate, lower acquisition accuracy, and lower acquisition accuracy. This article designed a user browsing behavior data query and data classification optimization algorithm, combine the two algorithms to generate a data acquisition algorithm, and enumerate examples. Because the traditional data acquisition method has the disadvantages of slow acquisition speed, low acquisition rate, and low acquisition accuracy, this article designs a user browsing behavior data query and data classification optimization algorithm that combines two algorithms to generate data acquisition. Experimental results show that the proposed method improves the data acquisition accuracy by 45%, which shows that the proposed method is more suitable for acquiring user browsing behavior data than the traditional method.

2 Cloud Computing User Browse Data Acquisition Method Design

2.1 Collect Query User Browsing Behavior Data

When querying the user's browsing behavior data, it is necessary to first determine the query node. When receiving a query request, each node needs to sense and simultaneously collect the data source of the area covered by the node, and perform operations such as calculation and processing on it. Get result sets that tend to request, the energy required to set the above stages is e_i. Each node then transmits data to the query node along its own path, in the process of transmission, each node conducts fusion processing on the received data.

When the node i Transfer information to j, node i energy required to transmit unit information $e_l(i)$ can be described as:

$$e_l(i) = (e_i + e_d r_{ij}^n)B \tag{1}$$

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{2}$$

In the formula, e_l used to describe the energy required to deliver each bit of information, unit is J/b; e_d used to describe the unit energy loss during mass data transmission in a distributed cloud computing environment, e_l and e_d Mainly depends on the query node itself; Parameter n Used to describe the channel loss index, mainly depends on distributed cloud computing network environment; The greater the noise, n the greater the value, usually take 2 and 4; r_{ij} used to describe the transmission distance; B is a fixed value. x and y denotes the data's independent variables [4].

Node j slave node i the energy required to receive the unit information $e_r(j)$ can be described as:

$$e_r(j) = e_r B \tag{3}$$

In the formula, e_r used to describe the energy required to receive each bit of information, unit is J/b.

Then the information of the K layer node needs to be transmitted through the $K-1$ node, assume that the distance between the i layer node and the $i-1$ layer is d_i, the energy consumption of a message transmission can be calculated by the following formula [5–7]:

$$\begin{aligned} e_k &= \sum_{i=1}^{k} (e_l + e_d d_i^n) B \\ &= (k e_l + e_d \sum_{i}^{k} d_i^n) B \end{aligned} \tag{4}$$

However, not all node information on the M layer is passed to the root node, after the information on the M layer is sent to the $M-1$ layer, data fusion processing is required, there is less information to reach the root node, therefore, e_k is the ultimate value of the quantity of information.

Assume that under a query request, the node i conducts m message sending and n message receiving, the remaining energy of the node i after the query can be described as:

$$\begin{aligned} E_{leave} &= e - \sum_{a=1}^{m} e_l(i, j_a) - \sum_{b=1}^{n} e_r(i, j_b) - e_i \\ and \quad & j_a, j_b \in N_i \end{aligned} \tag{5}$$

In the formula, $\sum_{a=1}^{m} e_l(i, j_a)$ and $\sum_{b=1}^{n} e_r(i, j_b)$ are used to describe the energy consumption of m information transmission and n information reception respectively; e_i is used to describe the energy consumption required by node i to process data.

In summary, after calculating the energy consumption required for each node to process user browsing behavior data, the data can be queried according to the calculation result.

2.2 Data Optimization Classification

After successfully querying users for browsing behavior data, these data are classified and optimized. For databases that contain large amounts of data in a cloud computing environment. The data text is described by $p(t)$, the designed transmission operator is

described by $h_i(t)$, complete the convolution operation, $n_{pi}(t)$ is used to describe classification interference, the genetic iterative classification spread can be described as [8, 9]:

$$p_n(t) = p(t) * h_i(t) + n_{pi}(t) \tag{6}$$

Among them, $h_i(t)$ is used to describe how $p(t)$ performs mass data classification and optimization in a cloud computing environment, the classification response function between data is:

$$S_{ri}(t) = S(t) * h_i(t) + n_{ri}(t) \tag{7}$$

Among them, $S_{ri}(t)$ is used to describe the corresponding function of classification between data; $h_i^{'}(t)$ is used to describe the corresponding channel function of the massive data in the cloud computing environment in the $S(t)$ classification process, then there are:

$$\begin{aligned} r_i(t) &= S_n(t) * p_{ri}(-t) \\ &= S(t) * (-t) * h_i(t) * h_i(-t) + n_{1i}(t) \end{aligned} \tag{8}$$

Among them, $n_{li}(t)$ is used to describe classification interference items of similar characteristic data when classifying, then:

$$\begin{aligned} r(t) &= \sum_{i=1}^{M} r_i(t) * p(t) \\ &= S(t) * p(t) * p(-t) * \sum_{i=1}^{M} h_i(t) * h_i(-t) + \sum_{i=1}^{M} n_i(t) \end{aligned} \tag{9}$$

If the peak of the $p(t)$ autocorrelation amplitude is prominent, there are:

$$p(t) * p(-t) \cong \delta(t) \tag{10}$$

In the formula, δ is the coefficient [10].

In this way, the formula for optimizing the mass data classification in the cloud computing environment is:

$$\begin{aligned} r(t) &\cong S(t) * \delta(t) * \sum_{\mathrm{i=1}}^{\mathrm{M}} \delta(\mathrm{t}) + \sum_{\mathrm{i=1}}^{\mathrm{M}} \mathrm{n_i(t)} \\ &= MS(t) + \sum_{i=1}^{M} n_i(t) \end{aligned} \tag{11}$$

Because nodes that exceed the distribution area mainly use multi-hop to transmit information, it is assumed that there are M layers in the distributed cloud computing environment. Then, the information of the n_i layer node needs to be transmitted through the S layer, and is transmitted layer by layer to the source node. According to formula

(11), we can get the classification of massive data in the cloud computing environment. In the formula, $r(t)$ represents the result of massive data classification.

2.3 Realizing Fast Acquisition of User Data in Cloud Computing

Integrate 1.1 Designed User Browsing Behavior Data Query Algorithms and 1.2 Designed Data Optimized Classification Algorithms, get a quick calculation algorithm for cloud computing user browsing behavior data, namely:

Bring user browsing information into the data acquisition algorithm formula, first, use formula (4) and (5) to calculate the energy consumption required for each node to process user browsing behavior data, the calculated result is the user browsing data on each node; then the calculated query data is brought into formula (11), calculate the peak value of the similar feature data at the time of classification by genetically iterative classification calculations, this completes the optimization of data classification, output the same type of data directly, in this way, the user's browsing behavior data under cloud computing can be quickly obtained.

3 Experimental Analysis

3.1 Experimental Data

In order to verify the validity of the design method of this article, it will be tested. The experimental data comes from the browsing sequence of different types of users on a website. The site has 90 main navigation bars, there is a cross between the navigation bar contents, and the navigation bar nests up to four levels. The personnel involved in the experiment came from people of different age groups, different career backgrounds, and different educational backgrounds. By using the user behavior data to obtain the mathematical model on the client side, the user behavior data of the experimenter browsing the website is obtained, and the duration is 160 min.

3.2 Experimental Results and Analysis

The comparison between the data collected by this method and the traditional method in the same time is shown in Fig. 1.

As can be seen from Fig. 1, under the same operation time, the data scale obtained by the method in this paper is larger than that obtained by the traditional method, and with the increase of time. The larger the gap between the data acquisition scale of this method and the traditional method is, the faster the data acquisition speed and more data are obtained.

As can be seen from Fig. 2, when the data set increases gradually, the acquisition hit ratio of both the present method and the traditional method decreases to a certain extent. In the range of error tolerance, the acquisition hit ratio of this method is obviously higher than that of the traditional method, which proves the effectiveness of this method.

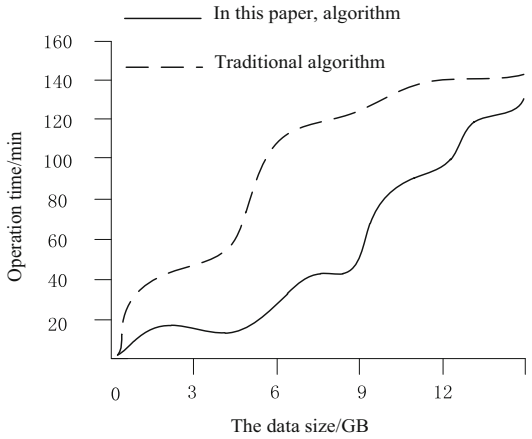

Fig. 1. Data collection

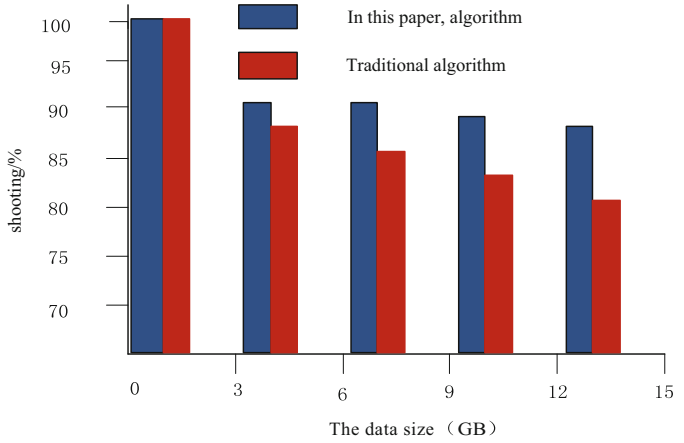

Fig. 2. Shows the hit ratio comparison

Table 1. Data acquisition error

	In this paper, algorithm	Traditional algorithm
Mean absolute error	0.342	4.161
Root mean square error	1.240	10.520

The data acquisition error of the algorithm and the traditional algorithm in this paper is shown in Table 1.

Several experimental results show that the design of data acquisition method compared with traditional methods, data acquisition speed faster, data capture ratio higher, obtain the result accuracy is higher, and speed, shooting and accuracy are increased by about 45%.

4 Conclusion

In this paper, the method of data acquisition of browsing behavior of cloud computing users is designed to improve the speed, hit rate and accuracy of data acquisition through data query and data optimization classification. The test data show that compared with the traditional methods, the method designed in this paper has improved the speed, the hit rate and the accuracy by about 45%. It is hoped that this study can provide useful help for users to obtain browsing behavior data.

References

1. Alharbi, S.: An empirical investigation on the impact of trust mediated determinants and moderating factors on the adoption of cloud computing. Int. J. Inf. Technol. Comput. Sci. **9**(11), 12–22 (2017)
2. Gdaniec, N., Szwargulski, P., Knopp, T.: Fast multi-resolution data acquisition for magnetic particle imaging using adaptive feature detection. Med. Phys. **44**(12), 6456 (2017)
3. Gao, C., Wang, H., Wang, J.: An improved CURE algorithm based on the uncertainty of mobile user data clustering. Comput. Eng. Sci. **38**(4), 768–774 (2016)
4. Namasudra, S., Roy, P.: Time saving protocol for data accessing in cloud computing. IET Commun. **11**(10), 1558–1565 (2017)
5. Zhe, D., Zhen, Q., Zheng, W.T., et al.: A recommendation model based on browsing behaviors of mobile users. J. Univ. Electron. Sci. Technol. China **46**(6), 907–912 (2017)
6. Bo, Z., Jin, L.: Research on the acquisition and application of user data in publishing industry against the background of big data. Editors' Friend **12**(8), 26–31 (2017)
7. Guohua, Y., Jingjing, K., Fang, L.: Data source description approach for deep web based on domain features and user query-based sampling. Libr. Inf. Serv. **1**(15), 138–145 (2017)
8. Taieb, C.: A quick model method for obtaining real-height parameters from routine ionospheric data. Radio Sci. **2**(10), 1263–1267 (2016)
9. Wang, Y.: Accurate detection of user characteristic data in large data network. Comput. Simul. **34**(6), 415–418 (2017)
10. Yang, J., Shi, Z., Liu, Z.: Simulation research on fast acquisition of terminal user information based on big data analysis. Comput. Simul. **1**(2), 110–111 (2018)