Design and Implementation of Large Mass Information Management System

Zhang Haiping[1]

[1]Shenzhen Institute of Information Technology, ShenZhen, GuangDong, 518172

[a]22900699@qq.com

Keywords: information management; massive data; weak association mining

Abstract. The design and implementation method of large mass information management system is researched, and the efficiency of information management is improved. In the design process of large mass information management system, a design method of the information management system is proposed based on weak clustering algorithm. The data are decomposed by the features of data in information management system, according to the data features, all data are fused. The data in the massive information management system are divided effectively, the calculation of the feature association probability is completed, classification of property elements is obtained through weak clustering method, so as to realize the normal operation of the large mass information management system. The experiment results show that, this method has high efficiency, it can meet the demand of information management system design.

Introduction

With the rapid development of computer information processing technology, the design of information management system for large mass data has become the core content of comp uter technology information research areas. Therefore, how to mine the target information in the mass information is very important, so as to realize the design of information manage ment system for large mass data, it has great significance in the field of computer informat ion management, and it has caused a lot of attention from experts[1-3]. At present, the desi gn methods of the information management system include as: information management syst em design method based on decision tree algorithm, information management system design method based on association rule algorithm, and design method of the information manage ment system based on fuzzy clustering algorithm. Among them, the most commonly used is the design method of the information management system based on association rule algorit hm. Because of the information management system design method has broad development space in the information field, therefore, it has great potential for development[4,5]

Principle of design method of large massive information management system

A Relationship matching of data in information management system

In the information management system of large mass data, large amounts of data have relation to each other, there is a certain relationship, the relationship is described by the datamessage, the association feature of the data is key part of information management system, feature of the data is mined in this paper, the data mining algorithm is designed, the principle is shown as follows:

First, the relevant feature of data in information management system is obtained, the data feature decomposition is completed, the following feature matrix can be obtained:

$$\begin{bmatrix} A_{11} & A_{12} \cdots & A_{1l} \\ A_{21} & A_{22} \cdots & A_{2q} \\ \cdots & \cdots \cdots & \cdots \\ A_{l1} & A_{k2} \cdots & A_{lm} \end{bmatrix} \tag{1}$$

Wherein, k is the number of data of information management system, l is the data types number, A_{kl} is the l th data in the k th type, the data feature matrix of the information management system is converted, the average value of feature is obtained as:

$$v = l^2 \cdot \sum_{k=2}^{k} A_{kl} \tag{2}$$

According to the data features in information management system, the whole data fusion is completed, obtain the results as:

$$A_{kl} = A_{kl} \cdot v \tag{3}$$

$$U_{kl} = h\left(A_k, A_l\right) \tag{4}$$

Wherein, U_{kl} is the data set composed by the data features in data information management system after interference filtering. $h\left(A_k, A_l\right)$ is the data feature extraction function, it can be described by the following formula:

$$h\left(A_k, A_l\right) = A_k^2 + A_l^2 \tag{5}$$

According to the related decision probability of data information management system, the massive data can be effectively divided by the features, the association probability of all data feature is calculated as:

$$r\left(g_l \cdot d_k\right) = \left(r\left(g_l\right) + r\left(g_l \cdot \delta_w\right)\right) / r\left(d_k\right)$$
$$l = 0,1,\cdots,q \tag{6}$$

Where, $r\left(d_k\right)$ is used to express a priori probability of data, $r\left(g_l \cdot d_k\right)$ is the corresponding conditional probability. Assumed the correlation of the data between the features is poor, then:

$$r\left(d_k\right) = \sum_{l=2}^{q} r\left(g_l\right) / r\left(d_k \cdot g_l\right) \tag{7}$$

According to the data characteristics, the correlation probability of different kinds of data is calculated:

$$r\left(d_k \cdot g_l\right) = U\left(q\left(d_k\right) + g_l\right) / U_\eta^2 \tag{8}$$

B Implementation of correlation data weak clustering processing of information manage ment system

Assumed data elements in the information management system is composed of correlation data, the data set is $U = \left\{u_1, u_2, \cdots, u_p\right\}$, where, $U = \left\{u_1, u_2, \cdots, u_p\right\}$ is used to describe the k element of the data, wherein, the property of the data as above is $J = \left\{j_1, j_2, \cdots, j_n\right\}$, $u_k\left[j_t\right]$ is the value of k element of the data on the property, the attribute elements are cl assified through weak clustering method, the quantitative elements are converted into the cat egory type.

In the information management system, the data sample space $Y = \left\{y_1, y_2, \cdots, y_p\right\}$, on the basis of the weak clustering method, it can be divided into d different categories, for some elements $y_j \in Y$, in clustering process, no longer classified accurate operation of all eleme nts in the information management system, so the demand probability x_{jk} of a particular el ement j belongs to the category k. The following equation is used to describe the weak classification matrix in the data sample space:

$$X = \left(x_{jk}\right) \quad (9)$$

The massive data classification function is obtained in the information management system based on the following formula:

$$K_n\left(X, A\right) = \sum_{j=1}^{p}\sum_{k=1}^{d} x_k^n e_{jk}^2\left(y_i, a_k\right) \quad (10)$$

In the data information management system, the data correlation parameters of massive data are $d, p, n, c = 1$, the weak clustering center is: $A_{(c)} = \left(a_1, a_2, \cdots, a_d\right)$. Through the following formula, the data update operation is completed:

$$x_{jk} = \frac{1}{\sum_{t=1}^{d}\left[\frac{e_{jk}}{e_{jt}}\right]^{\frac{2}{n-1}}} \text{ while } e_{jk} \neq 1$$

$$x_{jk} = 0 \text{ while } d_{jk} = 0, k \neq l$$

$$x_{jk} = 1 \text{ while } d_{jk} = 0 \quad (11)$$

The sample means parameters can be calculated by the following formula:

$$a_k = \frac{\sum_{j=1}^{p} x_{jk}}{\sum_{j=1}^{p} x_{kl}} \quad (12)$$

Values a_c and a_{c+1} are compared, if the values meet relationship as following formula, data clustering operation is realized in information management system, if it is not satisfied, then continue to the cluster analysis:

$$\left|a_c - a_{(c+1)}\right| \leq \phi \quad (13)$$

Through the analysis of the process, the data in the information management system are taken with clustering processing, the design of information management system is obtained.

Simulation experiments

In order to verify the validity of the method in this paper, the simulation is taken, in the process of the experiment, the traditional algorithm and the new algorithm are used for design of large mass information management system, the data query is taken in the management system, and the features are extracted, the data query efficiency is used as the comparison index, the results are obtained as follows, which shown in Figure 1.

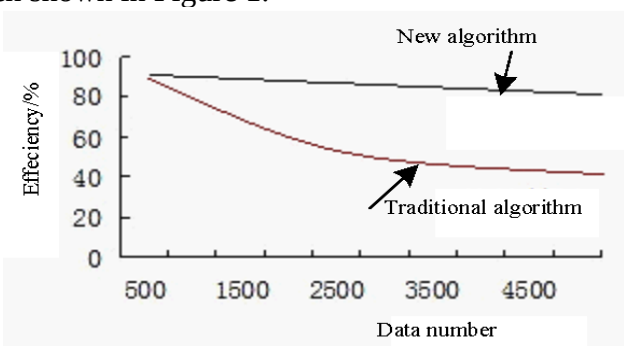

Figure 1 Data query efficiency comparison of different algorithms for information management system design

According to the above results, we can lean that the new method is applied in information management system design, it can quickly query the information of the target, satisfactory results are obtained, the efficiency is better than the traditional method.

Conclusions

In the design process of large mass information management system, a design method of the information management system is proposed based on weak clustering algorithm. The data are decomposed by the features of data in information management system, according to the data features, all data are fused. The data in the massive information management system are divided effectively, the calculation of the feature association probability is completed, and the classification of property elements is obtained through weak clustering method, so as to realize the normal operation of the large mass information management system. The experiment results show that, this method has high efficiency, it can meet the demand of information management system design, it has good application value in practice.

References

[1] Wen T. Unified Tuning of PID Load Frequency Controller for Power Systems via IMC[J]. IEEE Transactions on Power Systems,2010,25(1): 341- 350.

[2] Alfaro V M, Vilanovab R. Robust tuning of 2DoF five-parameter PID controllers for inverse response controlled processes[J]. Journal of Process Control, 2013,23(4): 453-462.

[3] Hu W M, Hu M, Maybank S. Adaboost based algorithm for network intrusion detection [J]. IEEE Transactions on Systems, Man and Cybernetic, Parb B: Cybernetics, 2008, 38: (2): 577-583.

[4] He Yongqiang，Gu Chunying. Distributed Spatial Query Processing and Parallel Schedule Based on Zonal Fragmentation[J]. Bulletin of Science and Technology, 2014,30(1):110-116.

[5] CHRN Hao, YANG Jun-AN, ZHUANG Zhen-Quan. The Core of Attributes and Minimal Attributes reduction in Variable Precision Rough Set[J]. Chinese Journal of Computers, 2012, 35(5): 1011-1017.