Research on Hierarchical Mining Algorithm of Spatial Big Data Set Association Rules

Yue Wang(✉) and Wei Song

Software College and Nanyang Institute of Technology, Nanyang 473000, China
wangyue6651@163.com

Abstract. Aiming to improve the security of large database in cloud storage space, a hierarchical mining algorithm of spatial big data set association rules based on association dimension feature detection is proposed. The statistical characteristic quantity of large spatial data set is constructed by means of group sample regression analysis, and the sampling and sample recognition of spatial big data set are carried out by using fuzzy rough set mapping method. The association rule distribution model of large spatial datasets is constructed by using the hierarchical mining method of association rules, and the feature quantities of association rules are extracted from large spatial datasets. The correlation dimension feature extraction algorithm is used to optimize the extraction process of large spatial data sets adaptively, so as to realize the hierarchical mining optimization of spatial big data set association rules. The simulation results show that the proposed method has higher accuracy, higher mining accuracy and better feature matching ability, which improves the mining ability of association rules in large database in cloud storage space.

Keywords: Cloud storage · Database · Spatial big data · Association rules · Hierarchical mining

1 Introduction

With the development of communication technology of large database of cloud storage space, cloud computing is used to control the transmission of large database of cloud storage space, which can improve the bandwidth and capacity of the output of large database of cloud storage space. However, in the large database of cloud storage space, because of the random distribution and self-organizing network of large database nodes in cloud storage space, the large database of cloud storage space is easy to be tiered by association rules, so the active mining of association rules in large database of cloud storage space is needed. Combining the association rule tiering mining of cloud storage space large database and the association rule analysis of association rule tiered data, mining the feature quantity of spatial big data set of cloud storage space large database [1]. The security management and information storage of cloud storage space large database are realized, and the security of cloud storage space large database is improved. The hierarchical mining method of spatial big data set association rules for cloud storage space large database is studied. It is important to ensure the security of large database of cloud storage space [2].

G. Gui and L. Yun (Eds.): ADHIP 2019, LNICST 302, pp. 200–208, 2019.
https://doi.org/10.1007/978-3-030-36405-2_21

The research on hierarchical mining of spatial big data set association rules is based on feature extraction and information scheduling of hierarchical data of association rules of large database in cloud storage space. According to the statistical features of large spatial data sets, the hierarchical mining of association rules for large database in cloud storage space is carried out [3]. The main coded features and time-frequency correlation features of spectral features are extracted from the spatial big data set. By scheduling and characteristic decomposition of association rules in large spatial data sets, the security of large database in cloud storage space is improved. In Ref. [4], a spatial big data set extraction technique based on genetic algorithm is proposed. Feature extraction and blind separation of spatial big data sets are carried out to construct genetic optimization control for spatial big data set mining. However, the computation cost of this method for hierarchical mining of spatial big data sets association rules is large and its self-adaptability is not good. In Ref. [5], a hierarchical data detection algorithm based on symbol envelope amplitude extraction for large database of cloud storage space is proposed, and the distribution model of symbol transmission channel for large database of cloud storage space is constructed. The feature value of symbol envelope amplitude is extracted from large database traffic sequence in cloud storage space for mining association rules. However, the anti-interference ability of this method is not good, and the accurate probability of mining association rules is not high.

In order to solve these problems, a hierarchical mining algorithm of spatial big data set association rules based on association dimension feature detection is proposed in this paper, and the fuzzy rough set mapping method is used to sample and identify the spatial big data sets. The association rule feature quantity of spatial large data set is extracted, and the extraction process of spatial large data set is optimized adaptively by using correlation dimension feature extraction algorithm, and the hierarchical mining optimization of spatial big data set association rule is realized. Finally, simulation experiments are carried out to show the superior performance of this method in improving the hierarchical mining ability of spatial big data set association rules.

2 Statistical Analysis and Feature Extraction of Large Spatial Data Sets

2.1 Statistical Feature Monitoring of Large Spatial Data Sets

In order to realize the hierarchical mining of association rules in spatial big data sets, the statistical features of large spatial data sets are constructed by the method of group sample regression analysis, and the data detection of association rules tiering in large database in cloud storage space is carried out. An undirected graph model $G = (V, E)$ is used to represent the sensor network structure model of spatial big data set monitoring, and in the transmission link model of hierarchical association rules of large database in cloud storage space. The node sensing point v is the root node of the large database in cloud storage space [4]. The on-line monitoring of large spatial data sets is carried out at the output link layer. For any node satisfied with $e \in E$, the Sink link set of the monitoring node is $v \in V$, in the 3D spatial scattering cluster. Scattered cluster cloud storage space large database topology edge structure satisfies SF, assumption transmission link data set $X = \{x_1, x_2, \cdots, x_n\}$, of receiver antenna and transmitter antenna, the statistical analysis model of association rule tiering monitoring in large database of cloud storage space is constructed. The model is described by directed graph model $G(A)$, $G(B)$ and the statistical characteristic point <*x, y*>, under the tiering of association rules in large database of cloud storage space is described by A, B. The statistical feature monitoring model of large spatial data set is obtained as shown in Fig. 1.

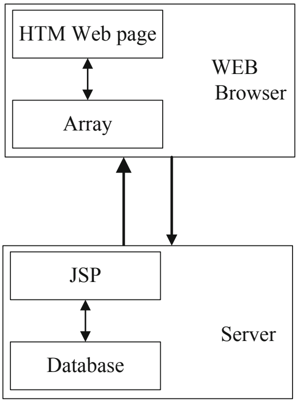

Fig. 1. Statistical feature monitoring node model for large spatial data sets

According to the statistical characteristic distribution model of large spatial data set shown in Fig. 1, the big data sampling discrete characteristic value of constructing the association rule tiering of cloud storage space large database is expressed as follows:

$$X_1(k) = FFT[x_1, x_1(k+1), \ldots\ldots, x_1(k+N-1)]^T \tag{1}$$

$$X_2(k) = FFT[x_2, x_2(k+1), \ldots\ldots, x_2(k+N-1)]^T \tag{2}$$

Where, $\tilde{X}_1(k)$, $\tilde{X}_2(k)$ are a large spatial data set composed of $X_1(k)$, $X_2(k)$ items before $N/2+1$, from which the association rule set of large spatial data set is extracted, and N-dimensional vector x(t) is used to represent the vector of tiered data of association rules in large database in cloud storage space, then:

$$x(t) = As(t) + n(t) \tag{3}$$

Where

$$\boldsymbol{x}(t) = [x_{-P+1}(t), x_{-P+2}(t), \cdots, x_P(t)]^T_{N\times 1} \tag{4}$$

$$\boldsymbol{s}(\boldsymbol{t}) = [s_1(t), s_2(t), \cdots s_I(t)]^T_{I\times 1} \tag{5}$$

$$\boldsymbol{n}(\boldsymbol{t}) = [n_{-P+1}(t), n_{-P+2}(t), \cdots, n_P(t)]_{N\times 1}^T \tag{6}$$

$$\mathbf{A} = [a(\theta_1, r_1), a(\theta_2, r_2), \cdots, a(\theta_I, r_I)]_{N\times I} \tag{7}$$

According to the above analysis, the spatial big data set monitoring model is constructed. According to the monitoring results of the original data, the hierarchical mining of association rules and self-adaptive scheduling are carried out, and the information fusion processing is carried out with the result of feature extraction. Improve the statistical feature analysis ability of large spatial data sets [7].

2.2 Feature Analysis of Large Spatial Data Set

The sub-carrier modulation method is used to describe the hierarchical node distribution characteristics of association rules, and the principal component feature information of the hierarchical data of large database in cloud storage space is obtained as:

$$C_1(m,n) = \sum_{i=1}^{L} c_{4s_i} e^{j2\phi_i} \tag{8}$$

In this formula, $c_{4s_i} = cum\left\{|s_i(t)|^4\right\}$ denotes the energy spectral density of the spatial big data set at node s_i. ϕ_i is the distribution coefficient of characteristic information. C_{4S} is used to represent the information intensity of spatial big data sets in the aggregation link layer:

$$C_{4S} = diag[c_{4s_1}, c_{4s_2}, \cdots, c_{4s_L}] \tag{9}$$

It is known that $a(t) \geq |s(t)|$, it represents the energy spectral density of the hierarchical data of association rules at node s_i. The maximum envelope amplitude of the large spatial data set is $|s(t)|$, and the first-order statistic of the hierarchical data of association rules is $a(t)$. The following $4P \times 4P$ matrix is constructed to represent the statistical characteristics of large spatial data sets:

$$C = \begin{bmatrix} C_1 & C_2 & C_5 & C_4 \\ C_2^H & C_1 & C_6 & C_7 \\ C_5^H & C_6^H & C_1 & C_3^H \\ C_4^H & C_7^H & C_3 & C_1 \end{bmatrix} = \overline{A} C_{4s} \overline{A}^H \tag{10}$$

Where $\overline{A} = \left[A^H, (A\Lambda)^H, (A\Omega)^H, (A\Phi)^H\right]^H$, the fuzzy rough set mapping method is used for sampling and statistical feature distributed description of spatial big data sets in order to improve the detection ability of spatial big data sets [8].

3 Spatial Big Data Association Rules Hierarchical Mining Optimization

3.1 Hierarchical Mining of Association Rules

On the basis of the statistical feature quantity of large spatial data set constructed by grouping sample regression analysis method, the data mining optimization design is carried out. In this paper, a hierarchical mining algorithm of spatial big data set association rules based on association dimension feature detection is proposed [9]. Combined with the hierarchical mining method of association rules, the distribution model of association rules in large spatial data sets is constructed. Combined with the correlation detection method, the statistical probability distribution of large spatial data sets is obtained as follows:

$$Vt(k) = \{a_{s+t}...a_{t+1}a_t...a_1 1 | \overline{a_{s+t}...a_{t+1}} = k, a_i \in \{0,1\}, 0 \le k < 2^s\} \tag{11}$$

Assuming that $a_0, a_1 \in V$, machine learning algorithm is applied to adaptive optimization for Sink nodes of spatial big data set distribution, the correlation feature of spatial big data set detection is obtained as follows:

$$T_{l1} = \sqrt{F_{p1}^2 + F_{q1}^2} \tag{12}$$

The quantitative feature distribution set is calculated as $F_{i1} = \frac{1}{P_{i1}}$, $i = p, q$, in the source distribution domain of spatial big data set. Therefore, a hierarchical mining model of association rules for large database of cloud storage space is constructed, and the big data transmission link structure of large database of cloud storage space is established. It is expressed that $W(p) = G_T p^2 - Cp + \alpha T$, $W(p)$ is a quadratic function of spatial big data set link set p. Combined with the result of association rules hierarchical mining, the hierarchical mining of association rules is carried out [10].

3.2 Association Rule Hierarchical Data Association Dimension Feature Detection

The nonlinear feature combined with the hierarchical mining method of association rules is used to construct the association rule distribution model of spatial large data sets [11]. The decision statistics for hierarchical mining of association rules are as follows:

$$\mu(n) = \begin{cases} \beta_1\left[1 - \exp(-\alpha_1 |e_{MCMA}(n)|^2)\right], & E\left[(|e_{MCMA}(n)|^2)\right] > K \\ \beta_2\left[1 - \exp(-\alpha_2 |e_{MCMA}(n)|^2)\right], & \text{else} \end{cases} \tag{13}$$

By using the correlation dimension feature detection method, when the maximum root mean square error is satisfied with $MSE = E[(|e(n)|^2)] > K$, the smaller α_2 and β_2

are selected to mine the large spatial data set of cloud storage space, and the optimized mining algorithm is obtained as follows:

$$\text{ROUTE\_2}(\text{Route } u = u_{s+1}\ldots u_{t+1}u_t\ldots u_1 0, \quad v = v_{s+t}\ldots v_{t+1}u_t\ldots u_1 0)$$
$$x = u_{s+1}\ldots u_{t+1}; \quad y = v_{s+t}\ldots v_{t+1};$$
$$\boldsymbol{I}(x,y) = \emptyset;$$
$$\text{For each } e_i, \text{if}(u_i \neq v_i)\boldsymbol{I}(x,y) = \boldsymbol{I}(x,y) + e_i;$$
$$\text{While}(\boldsymbol{I}(x,y) \neq \emptyset)$$
$$\{e_i = \mathit{firstselect}(\boldsymbol{I}(x,y)); //$$
$$\text{form } x \text{ to } x + e_i; \quad x = x + e_i; \quad \boldsymbol{I}(x,y) = \boldsymbol{I}(x,y) - e_i; \}$$
$$)$$

According to the improved correlation dimension feature extraction algorithm, the adaptive iteration is carried out until the convergence criterion is satisfied, and the spatial big data set extraction is realized according to the coverage [12].

4 Simulation Experiment and Result Analysis

In order to verify the performance of this method in spatial big data set detection, simulation experiments are carried out. In the experiment, the algorithm is designed with Matlab, and the type of association rules tiering in large database in cloud storage space is DoS. The fundamental frequency of spatial big data collection is 20 kHz, the spatial big data set coverage is 300 × 300, and the modulation frequency of association rule layered data varies between [240 Hz, 1200 Hz]. Three kinds of association rule layered data are divided into two groups: Probe and ipsweep, the fundamental frequency is 20 kHz, the spatial big data set covers 300 × 300, and the modulation frequency is between 240 Hz and 1200 Hz. Under the condition of interference signal-to-noise ratio of −10–−2 dB respectively, the hierarchical mining of association rules for large data sets in cloud storage space and large database space is carried out, and the statistical features of large spatial data sets are constructed by using the method of group sample regression analysis. The original cloud storage space, large database space, large data set time domain distribution is shown in Fig. 2.

Taking the data of Fig. 2 as input, the association rule feature quantity of spatial large data set is extracted, and the extraction process of spatial large data set is optimized adaptively by using correlation dimension feature extraction algorithm, and the mining result of association rule hierarchy is obtained as shown in Fig. 3.

The analysis of Fig. 3 shows that the proposed method has strong anti-interference ability in mining spatial big data set association rules. On the basis of Fig. 3, in order to further prove the anti-interference performance of the proposed method, the traditional method in 2 was used as the contrast experimental group to conduct a comparative anti-interference experiment. The amplitude of the three methods was recorded respectively, with the interference sizes of −10 dB, −6 dB, −4 dB and −2 dB. The comparison results are shown in Table 1.

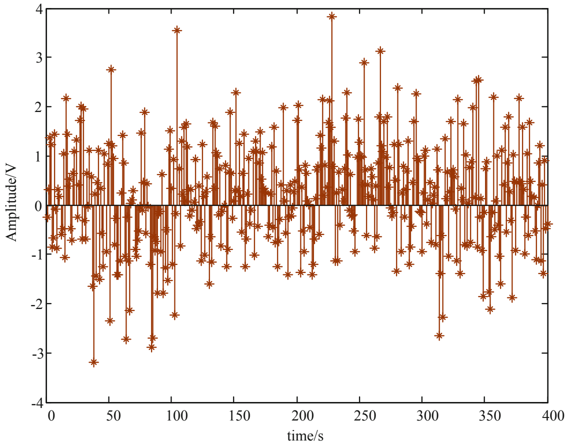

Fig. 2. Time domain distribution of large spatial data sets

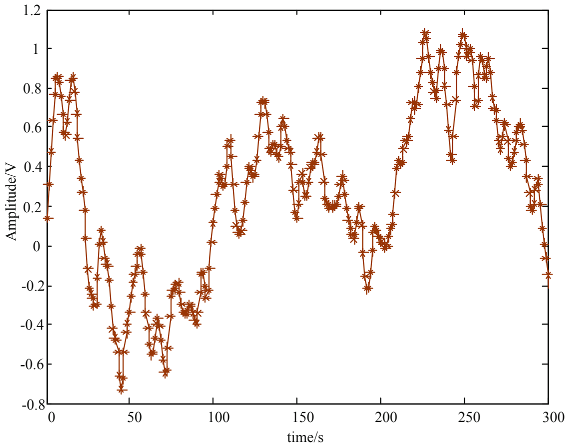

Fig. 3. Mining results of association rule hierarchies for association rule hierarchical data

Table 1. Performance comparison of hierarchical mining of association rule

SNR/dB	Proposed method	Wavelet detection	Time-frequency detection
−10	0.865	0.734	0.798
−6	0.997	0.905	0.876
−4	1	0.967	0.944
−2	1	0.988	0.969

The analysis Table 1 shows that with the increase of input signal-to-noise ratio, the probability of mining association rules layering increases continuously, and the accuracy of spatial big data set mining and detection of association rules is high by using the method proposed in this paper.

In order to further verify the performance of the proposed hierarchical mining algorithm for association rules in spatial large data sets, it is compared with literature [3], literature [4] and literature [5] to obtain the following mining accuracy experimental results.

As can be seen from the experimental results of the Fig. 4 above, compared with the traditional method, the proposed method has higher mining accuracy, and the precision value is more in line with the current application requirements in this field.

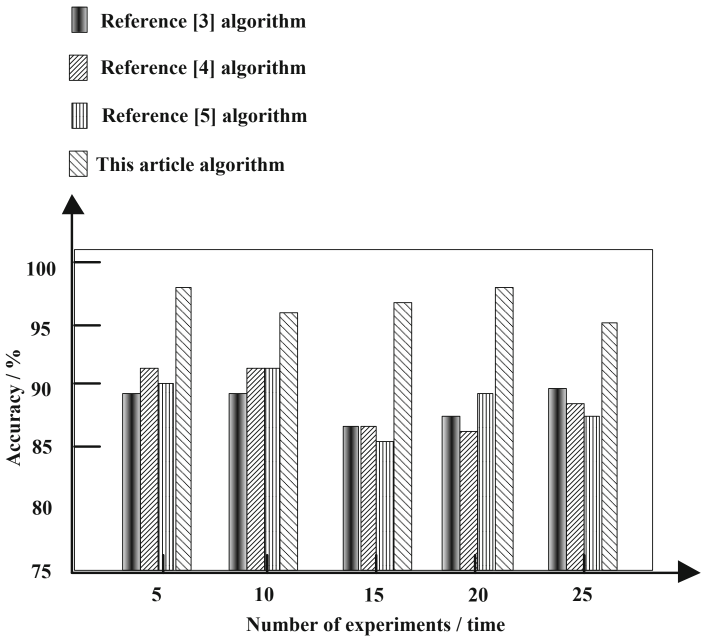

Fig. 4. Comparison of mining accuracy of different algorithms

5 Conclusions

In this paper, a hierarchical mining algorithm of spatial big data set association rules based on association dimension feature detection is proposed. The statistical characteristic quantity of large spatial data set is constructed by means of group sample regression analysis, and the sampling and sample recognition of spatial big data set are carried out by using fuzzy rough set mapping method. The association rule distribution model of large spatial datasets is constructed by using the hierarchical mining method of association rules, and the feature quantities of association rules are extracted from large spatial datasets. The correlation dimension feature extraction algorithm is used to optimize the extraction process of large spatial data sets adaptively, so as to realize the hierarchical mining optimization of spatial big data set association rules. The simulation results show that the proposed method has higher accuracy, higher mining accuracy and better feature matching ability, which improves the mining ability of association rules in large database in cloud storage space. The method has good application value in large data mining.

References

1. Muramatsu, M.: On network simplex method using the primal-dual symmetric pivoting rule. J. Oper. Res. Soc. Jpn. **43**(1), 149–161 (2017)
2. Ma, Z., Chen, W.: Friction torque calculation method of ball bearings based on rolling creepage theory. J. Mech. Eng. **53**(22), 219–224 (2017)
3. Tu, G., Yang, X., Zhou, T.: Efficient identity-based multi-identity fully homomorphic encryption scheme. J. Comput. Appl. **39**(3), 750–755 (2019)
4. Wang, Z., Huang, M., et al.: Integrated algorithm based on density peaks and density-based clustering. J. Comput. Appl. **39**(2), 398–402 (2019)
5. Helmy, A., Hedayat, A., Al-Dhahir, N.: Robust weighted sum-rate maximization for the multi-stream MIMO interference channel with sparse equalization. IEEE Trans. Commun. **60**(10), 3645–3659 (2015)
6. Alfaro, V.M., Vilanovab, R.: Robust tuning of 2DoF five-parameter PID controllers for inverse response controlled processes. J. Process Control **23**(4), 453–462 (2013)
7. Han, D., Chen, X., Lei, Y., et al.: Real-time data analysis system based on spark streaming and its application. J. Comput. Appl. **37**(5), 1263–1269 (2017)
8. Sun, D.W., Zhang, G.Y., Zheng, W.M.: Big data stream computing, technologies and instances. J. Softw. **25**(4), 839–862 (2014)
9. Hao, S.G., Zhang, L., Muhammad, G.: A union authentication protocol of cross-domain based on bilinear pairing. J. Softw. **8**(5), 1094–1100 (2013)
10. Zhu, Y., Zhu, X., Wang, J.: Time series motif discovery algorithm based on subsequence full join and maximum clique. J. Comput. Appl. **39**(2), 414–420 (2019)
11. Wang, Z., Zheng, Q., Chen, C., et al.: Virtual network mapping algorithm based on network simplex. Comput. Eng. **45**(4), 13–17 (2019)
12. Wang, I.L., Lin, S.J.: A network simplex algorithm for solving the minimum distribution cost problem. J. Ind. Manag. Optim. **5**(4), 929–950 (2017)