Eng Int Syst (2020) 4: 229–240

Engineering
Intelligent Systems

Intelligent Indexing Algorithm for the Association Rules of a Multi-Layer Distributed Database

Hao Peng,* Shun Yang, Qiong Liu, Qiong Peng and Qiao Li

School of Information and Mechanical & Electrical Engineering, Hunan International Economics University, Changsha 410202, China

In order to improve the retrieval ability of multi-layer distributed database association rules, a multi-layer distributed database association rule intelligent indexing method based on fuzzy correlation cluster analysis is proposed. The feature reconstruction of multi-layer distributed database association rules is carried out by the phase-space reconstruction method, the discrete fusion of multi-layer distributed database association rules is carried out by the rough set feature matching method, the association rule feature quantity mining and the statistical regression analysis are carried out in the multi-layer distributed database, the multi-layer distributed database association rules are extracted by the multi-grid partitioning method, and the key information index model of multi-layer distributed database association rules is established. The simulation results show that the multi-layer distributed database association rule intelligent index has a high fusion degree and high accuracy.

Keywords: Multi-tier Distributed Database; Association Rules; Intelligent Index; Information Fusion; Clustering

1. INTRODUCTION

In the network database, there are a large number of multi-tier distributed databases. The cloud data information of the network is stored in multi-tier distributed databases. Users use remote scheduling and access methods to locate and query remote databases. In order to improve the access and retrieval ability to the multi-tier distributed databases, the multi-tier distributed databases need to be source integrated. The integration model of association rules for multi-tier distributed databases is built [1]. Combined with feature extraction and big data mining of association rules for multi-tier distributed databases, the statistical analysis model of association rules for multi-tier distributed databases is built. The optimization mining and combination control of association rules for multi-tier distributed databases are carried out by using association rule scheduling and semantic ontology model design methods. The retrieval and adaptive mining capabilities of association rules for multi-tier distributed databases are improved. The research on intelligent index methods of association rules for related multi-tier distributed databases has attracted great attention [2].

The intelligent index of multi-layer distributed database association rules is based on the feature extraction of heterogeneous data. By mining the effective feature values of multi-layer distributed database association rules and adopting the distributed topology structure design method, the optimal topology design of multi-layer distributed database association rules is carried out to realize the source integration of multi-layer distributed database association rules. In the traditional method, the intelligent index method for association rules of multi-layer distributed database mainly includes the association rule mining method, statistical feature analysis method, autocorrelation feature detection method, and the statistical analysis model of association rules of multi-layer distributed database is constructed to mine source data [3]. A multi-layer distributed database association rule intelligent index method based on phase space reconstruction is proposed in Document [4]. The depth learning method

*Email of Corresponding Author: penghpengh@126.com

is used for adaptive optimization in the intelligent indexing process of multi-layer distributed database association rules, and the phase space reconstruction technology is used to realize the source combination of multi-layer distributed database association rules. However, this method has a large computational cost and poor real-time performance for data integration. Document [5] proposes an intelligent index method for association rules of multi-layer distributed databases based on a graph model structure design. Feature extraction technology is used to extract the average mutual information feature quantity of association rules of multi-layer distributed databases, and virtual data integration is carried out by combining it with the association rule mining method. However, this method has poor anti-interference and adaptive control ability in data integration [6].

To solve the above problems, this paper proposes an intelligent index method for association rules of multi-layer distributed database based on fuzzy correlation fusion clustering analysis. The main content of the article is:

(1) Firstly, the phase space reconstruction method is used to reconstruct the characteristics of association rules of multi-layer distributed databases.

(2) The rough set feature matching method is combined to carry out discrete fusion of association rules of multi-layer distributed databases.

(3) The key information index model of association rules of multi-layer distributed databases is established.

(4) The cluster clustering and fuzzy C-means fusion method are used to realize an intelligent index of association rules of multi-layer distributed databases.

The simulation experiment is used to test the performance, which shows the superior performance of the method in realizing intelligent index optimization of association rules in a multi-layer distributed database.

2. RELATED CONCEPTS AND ALGORITHMS

2.1 Multi-Tier Distributed Database Project Set

Mining association rules are used to produce all strong rules that meet the minimum support threshold (minsup) and minimum confidence (minconf) in a given set of transactions. The mining of association rules is a two-step process:

(1) Find all frequent itemsets satisfying minsup;

(2) Produce strong association rules from frequent itemsets.

Step (2) is relatively simple, and the overall performance of mining association rules is determined by Step (1).

There are n sites $S_1, S_2, \ldots, S_n (n > 2)$. The corresponding member databases are $DB_1, DB_2, \ldots, DB_n$ respectively. X X_{count} and X X_{count_i} represent the number of transactions with X in DB and DB_i, respectively.X_{sup} is the global support of X, and $X_{\text{sup}l}$ is the local support of X at site S_i. If X_{sup_i} > minsup, then X is the local frequent item set of site S_i. If X_{sup_i} > minsup, X is a global frequent item.

2.2 Related Algorithm

2.2.1 PPDMA Algorithm

The main idea behind the PPMDA algorithm is that constrained subtrees and other information transmitted between sites are encrypted with RSA algorithms, and some "interference" itemsets are added to them to hide the reality of the internationally supported constrained subtrees. The information from each site is then passed on to other sites and the information is decrypted.

Let E_i be the Encryption Key used to encrypt information on site i; D_i means decrypting the ciphertext with the decryption key D_i on site i; T represents data that can be used as interference itemsets; $|S_{Te_i}(m)|$ represents the set of encryptions on site i for $S_{T_i}(m)$, $(m \in F)$. The algorithm is divided into the following steps:

(1) The prime number P_i and Q_i are selected to obtain E_i and D_i; each site gets $|S_{Te_i}(m)|$. All sites are capable of interacting with the set $|S_{Te_i}(m)|$ with a standard random number generator. The steps are the same for each cycle in order.

(2) Site 0 gets a complete set of encrypted itemsets from the even-numbered sites and Site 1 gets the complete set of encrypted itemsets from the odd-numbered sites. This separate collection is intended to enhance the security of the algorithm, so that Site 0 gets less privacy information from other sites.

(3) Site 1 transfers the collected constraint subtree set to Site 0, and then Site 0 merges the two sets.

(4) The merged bound subtree set is decrypted on each site in turn; the interference item is then removed to obtain the merged decryption set. Output the results on the basis of mining. The encryption process is shown in Fig. 1.

2.2.2 FDMA Algorithm

For the convenience of narration, the corresponding FP tree of DB is recorded as F tree, and the corresponding FP tree for DB_i is recorded as F_i tree $(1 \leq i \leq n)$. The collection of all global frequent items is recorded as F. The set of global frequent itemsets obtained from DB mining is L. The task of mining the FDMA algorithm can be divided into two parts:

a. Setting up F_i tree and the corresponding constrained subtree $S_{T_i(j)}$;

b. Mining the constrained subtree $S_{T_i(j)}$ of each F_i tree to get the set.

By transforming the mining of the project set of distributed database into the mining based on each constraint subtree, the global frequent project set of F tree $S_{T(a)}$ is mined while the

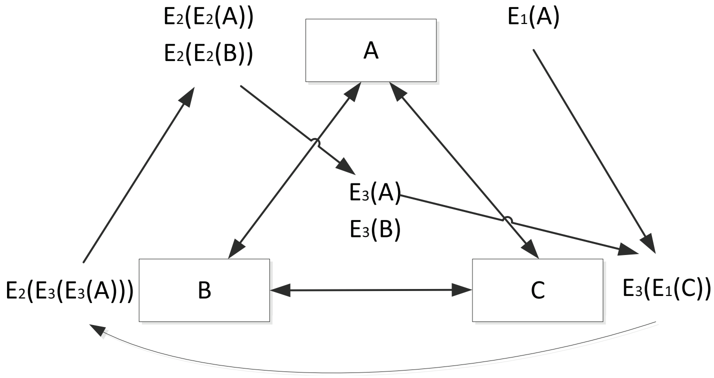

Figure 1 Encrypted information transfer process between sites.

Table 1 Database data types under different sites.

Database	Quantity	Type
DB_1	10	f,a,c,d,g,b,o,h
	20	a,co,h
DB_2	30	b,f,c,d,h
	40	b,c,k,s,a
DB_3	50	a,f,g,h
	60	k,s

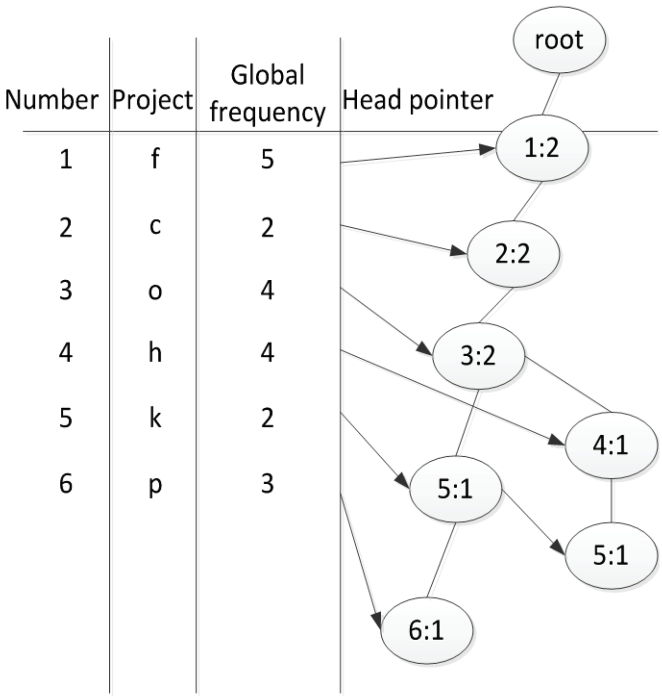

Figure 2 FP-tree 1.

mining of the item set of F tree $S_{T(a)}$ can be obtained from the local frequent item set of F tree $S_{T(a)}$.

As the FDMA algorithm uses the transfer of constrained subtrees (which can be represented by three very small arrays) to mine the item set, the transmission of a large number of candidate sets or frequent pattern trees is avoided, thus significantly reducing the network communication cost and improving the mining efficiency of the project set.

The transaction databases on the three sites S_1, S_2 and S_3 are DB_1, DB_2 and DB_3, respectively, as shown in Table 1.

For supmin $=$ 0.5, the F_1 tree and F_3 tree established by the algorithm FDMA are shown in Fig. 2, and Fig. 3 respectively. To facilitate the maintenance of the item set, the global frequency of all items is retained in the header table on each local frequent pattern tree. The FDMA algorithm is illustrated by finding F tree|ST (6). ST 3(6) in the F_3 tree is shown in Fig. 4.

Transmitted according to the communication strategy of the FDMA algorithm - Go to site S_1 and dig on S_1, project set {f,c,o,h,k,p} and subset of items.

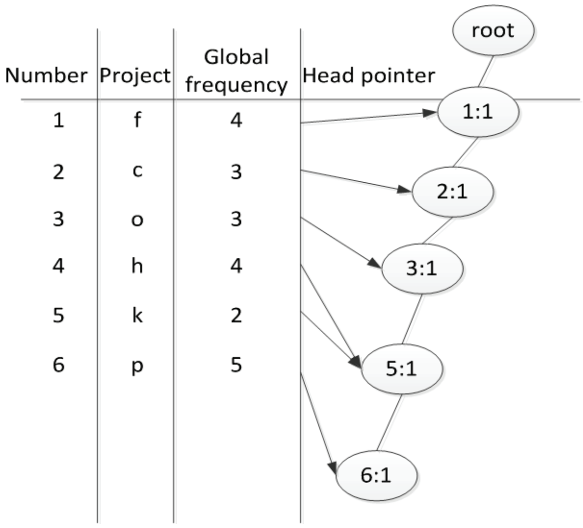

Figure 3 FP-tree 3.

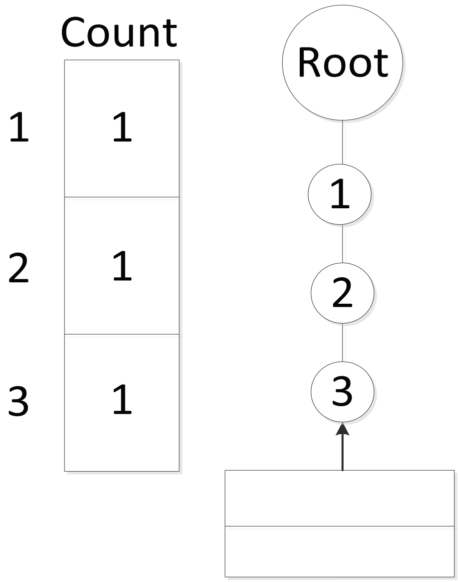

Figure 4 Bundles of FP-tree.

Based on the above algorithms, this paper will carry out the intelligent index of multi-layer distributed database association rules.

3. BASIC DEFINITIONS

3.1 Distributed Architecture of Association Rules for Multi-layer Distributed Database

In order to realize the optimal design of the multi-tier distributed database association rule intelligent index algorithm, the multi-tier distributed database association rule intelligent index is carried out by combining it with the distributed structure reorganization method of the multi-tier distributed database association rule storage nodes. The main flow of this paper is shown in Fig. 5.

The multi-tier distributed database heterogeneous storage structure model is established [7–8], and the information fusion model of the multi-tier distributed database association rule source based on the multi-tier distributed database is established according to the data architecture. The Distributed Architecture of Association Rules is shown in Fig. 6.

Under the fuzzy grid region clustering environment, the edge feature distribution set of multi-layer distributed database association rules under the source combination mode is obtained, and a binary directed graph is used to represent the graph model structure of the multi-layer distributed database association rules, wherein the graph model structure is the vertex set of distribution nodes of the multi-layer distributed database. It is a directed edge combination of association rules for multi-tier distributed databases [9–11]. Assuming that the multi-tier distributed database is a Sink storage node located in different places, the correlation statistical features of the association rules of the multi-tier distributed database are extracted by the multiple regression analysis method,

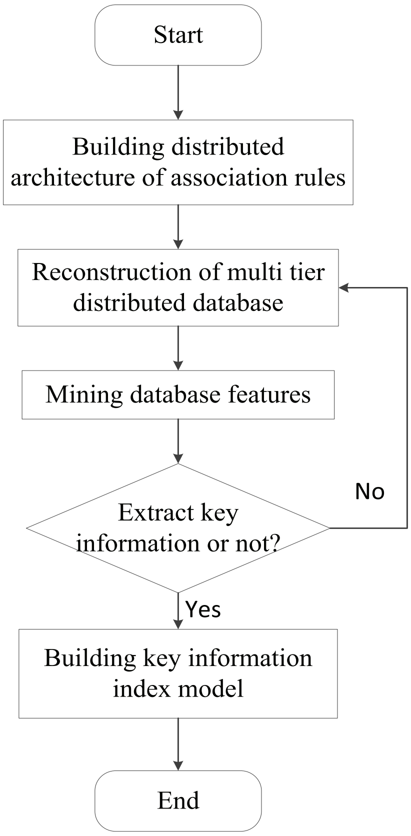

Figure 5 Multi-tier distributed database association rules intelligent indexing flow.

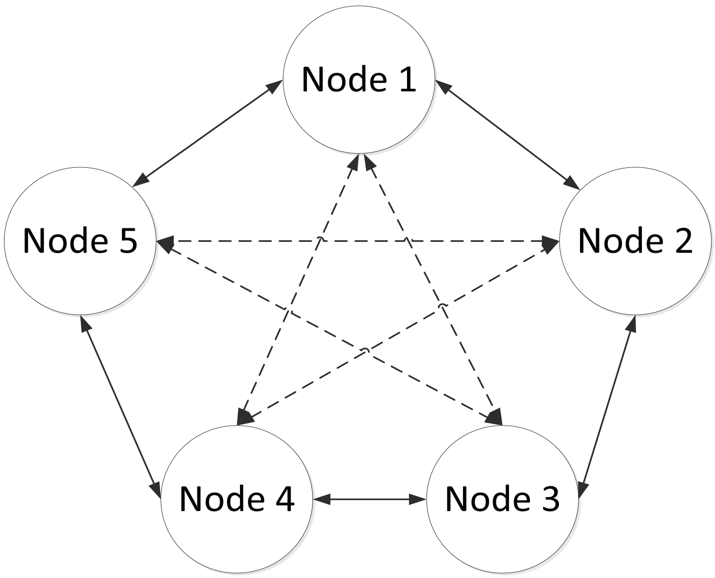

Figure 6 Distributed architecture of association rules.

and the distributed topology structure model of the multi-tier distributed database is obtained as shown in Fig. 7.

According to the distributed topology structure model of the multi-layer distributed database shown in Fig. 7, the measurement information of association rule detection of the multi-layer distributed database is calculated [12–14]. In the multi-layer distributed database association rule storage network structure model, the weighted coefficient of the directed graph vector of the source combination model is $W = \{u, w_1, w_2, \ldots, w_k\}$. In the information coverage area M of the association rule of the multi-layer distributed database, assuming m transmission link layers, the discrete distribution form of statistical data is $x_s = x(\eta_1), \ldots, x(\eta_N)^T$, the fuzzy node set of the association rule of the multi-layer distributed database is $T = \{t_1, t_2, \ldots, t_n\}$, and the estimated value of the difference:

$$\hat{x}_s = W_s^T y \tag{1}$$

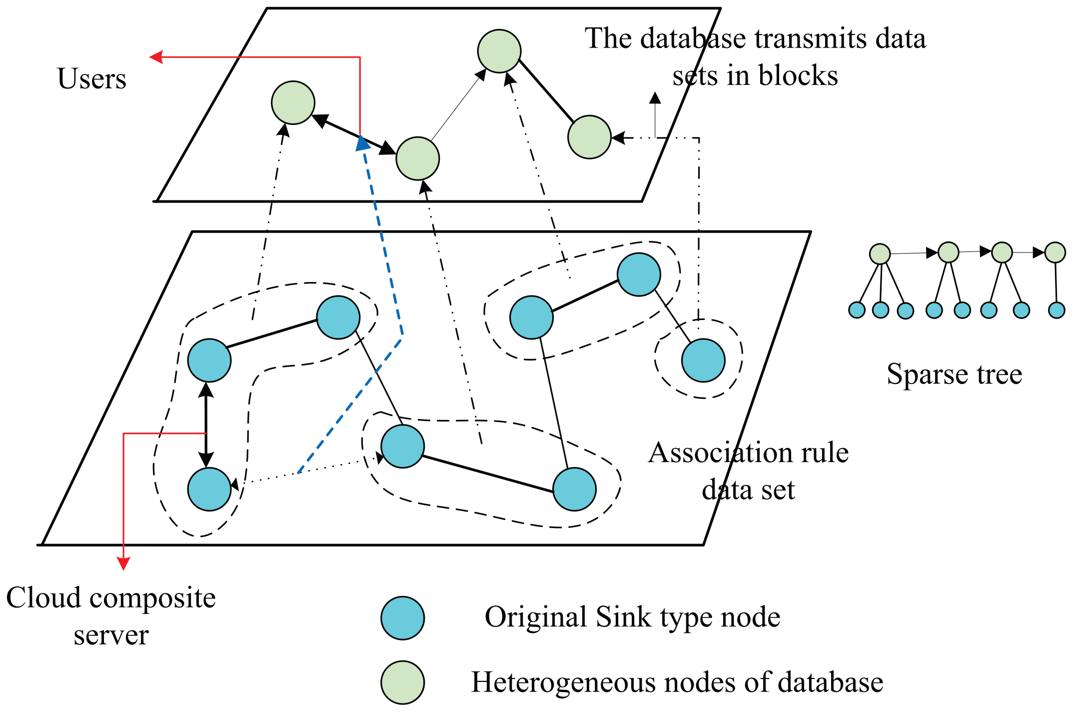

Figure 7 Distributed topology model of the multi-layer distributed database.

Based on the priority division method, the transmission load model for building association rules of multi-layer distributed databases is as follows:

$$r(t) = \sum_{i}\sum_{j=0}^{N_f-1}\sum_{l=0}^{L-1} b_i \alpha_l p\left(t - iT_s - jT_f - c_j T_c - \tau_l\right) + \omega(t) = \sum_{i}\sum_{j=0}^{N_f-1} b_i p_h\left(t - iT_s - jT_f - c_j T_c - \tau_0\right) + \omega(t) \tag{2}$$

Where

$$p_h(t) = \sum_{l=0}^{L-1} \alpha_l p\left(t - \tau_{l,0}\right) \tag{3}$$

In addition, $\omega(t)$ is the data dimension of virtual nodes of multi-tier distributed database, and $p_h(t)$ is the distance between Source and Sink nodes of multi-tier distributed database association rules. According to the above analysis, a distributed architecture model of association rules for multi-tier distributed databases is constructed to index association rules for multi-tier distributed databases [15–17].

3.2 Reorganization of Storage Structure

The phase space reconstruction method is adopted to reconstruct the characteristics of the association rules of the multi-layer distributed database, the rough set feature matching method is combined to carry out discrete fusion processing of the association rules of the multi-layer distributed database, the statistical feature quantity of the distribution feature set of the association rules of the multi-layer distributed database in the dense scene is calculated [18–19], and the time slot feature distribution value of the association rules of the multi-layer distributed database is obtained as follows:

$$\mathrm{ITrust}_{a\to c} = \frac{\sum\limits_{b\in adj(a,c)} \mathrm{DTrust}_{a\to b} \times (\mathrm{DTrust}_{b\to c} \times \beta_d)}{\sum\limits_{b\in adj(a,c)} DT\mathrm{trust}_{a\to b}} \tag{4}$$

According to the feature extraction result, the surface information of the association rules of the multi-layer distributed database is reconstructed, and in the high-dimensional phase space structure model, the spectrum z of the association rules source of the multi-layer distributed database is obtained to obey Gaussian distribution with parameters β_d, wherein:

$$\beta_d = (\mathrm{MPDist} - d + 1)/\mathrm{MPDist}, \quad d \in [2, \mathrm{MPDist}] \tag{5}$$

and where $adj(a, c)$ represents the number of reconstruction vectors $a \to c$, and the ontology structure of the multi-layer distributed database association rules sources is reorganized. In the surface distribution structure model of multi-layer distributed database association rules, sparse point expression method is used to analyze the similarity characteristics of multi-layer distributed database association rules, and the mapping relation of multi-layer distributed database association rules is obtained as $A \to B$, $B \to C$. The regression analysis model is derived as follows:

$$MSD_{a\to b} = 1 - \frac{\sum\limits_{i=1}^{|I_{a,b}|} \sqrt{\left(d_{a,i} - \bar{d}_a\right)^2 + \left(d_{b,i} - \bar{d}_b\right)^2}}{\left|I_{a,b}\right| \times \sum\limits_{i=1}^{|I_{a,b}|} \left[\sqrt{\left(d_{a,i} - \bar{d}_a\right)^2} + \sqrt{\left(d_{b,i} - \bar{d}_b\right)^2}\right]} \tag{6}$$

Using feature extraction technology to extract the average mutual information feature quantity of multi-layer distributed database association rules, and outputting the mutual information of attribute distribution of multi-layer distributed database association rules as follows:

$$I(Q, S) = H(Q) - H(Q|S) \tag{7}$$

wherein

$$H(Q \mid s_i) = -\sum_j \left[p_{sq}(s_i, q_j) \Big/ p_s(s_i) \right] \log_2 \left[p_{sq}(s_i, q_j) \sum_j \Big/ p_s(s_i) \right] \tag{8}$$

Combined with feature extraction technology, the average mutual information feature quantity of association rules of multi-layer distributed database is extracted, and the fuzzy correlation feature matching method is adopted to carry out the principal component analysis of association rules of multi-layer distributed database, thus realizing the reorganization of the data storage structure [20].

4. INTELLIGENT INDEX OPTIMIZATION OF ASSOCIATION RULES IN MULTI-LAYER DISTRIBUTED DATABASE

4.1 Sparsity Feature Extraction of Association Rules in Multi-Layer Distributed Databases

On the basis of the above feature reconstruction of multi-layer distributed database association rules by phase space reconstruction method, the optimization design of multi-layer distributed database association rules intelligent index method is carried out. This paper proposes a multi-layer distributed database association rules intelligent index method based on fuzzy correlation fusion clustering analysis. According to the attribute mining results of multi-layer distributed database association rules, the source combination is carried out, and the decision criteria for data source integration are satisfied:

Norm (1):

$$\sqrt{\frac{R^2_{(m+1)n} - R^2_{mn}}{R^2_{(m+1)n}}} = \frac{\left|x_{\eta(n)+m\tau} - x_{n+m\tau}\right|}{R_{(m+1)n}} \geq R_{tol} \tag{9}$$

Norm (2):

$$\frac{R_{(m+1)n}}{\sqrt{\frac{1}{N}\sum_{k=1}^{N}\left[x_k - \frac{1}{N}\sum_{k=1}^{N} x_k\right]^2}} > A_{tol} \tag{10}$$

According to the decision criteria of the intelligent index of multi-layer distributed database association rules, the principal component analysis of multi-layer distributed database association rules is carried out. In the characteristic distribution attribute set of data, $\{u_1, \ldots, u_N\}$ is set to represent the class space distribution set of multi-layer distributed database association rules including several virtual node sets, $v_1, \ldots, v_M$ is set to represent semantic ontology node sets, $R = [R_{u,v}]_{N\times M}$ is set to represent the attribute rule set of multi-layer distributed database association rules, information sampling is carried out in combination with the characteristic coding method of multi-layer distributed database association rules, and diversity scheduling of multi-layer distributed database association rules is carried out by adopting a grouping detection method. The following formula is deduced:

$$p_i^* = \frac{1}{\sum_{j=i}^{N} \frac{2m_j}{\sum_{k=j+1}^{N+1} L_k p_k - \sum_{k=j}^{N} E_k}} - 1, \quad i = 1, \ldots, N+1 \tag{11}$$

$CIntra_i(n)$ is used to represent the optimal interval between access nodes of multi-tier distributed database association rules, $CInter_i(n)$ is used to represent the total time slot of competing nodes i, and the distributed reorganization structure of multi-tier distributed database association rules is obtained as follows:

$$\begin{aligned} \boldsymbol{X}(n) &= \{x(n), x(n+\tau), \cdots, x(n+(m-1)\tau)\} \\ n &= 1, 2, \cdots, N \end{aligned} \tag{12}$$

where τ represents the embedding delay of association rules of multi-layer distributed database in high-dimensional phase space. According to the above analysis, combined with rough set feature matching method, the discrete fusion of association rules in multi-tier distributed database is carried out, and association rule mining and statistical regression analysis are carried out on association rule feature quantities in multi-tier distributed database [21].

4.2 Intelligent Index of Association Rules in Multi-layer Distributed Database

Setting a data set x formed by the characteristics of the association rules of the multi-layer distributed database, and establishing a state transition model. The expression of the characteristic evaluation concept set of the association rules of the multi-layer distributed database is as follows:

$$p(y \mid \alpha, \theta) = \sum_{k=1}^{K} \alpha_k p_k \left(y \mid \mu_k, \sum_k \right) \tag{13}$$

Mining attribute association rule feature of multi-layer distributed database association rule:

$$\max_{x_{a,b,d,p}} \sum_{a\in A}\sum_{b\in B}\sum_{d\in D}\sum_{p\in P} x_{a,b,d,p} V_p \tag{14}$$

$$\text{s.t.} \sum_{a\in A}\sum_{d\in D}\sum_{p\in P} x_{a,b,d,p} R_p^{bw} \leq K_b^{bw}(S), b \in B \tag{15}$$

By using the cloud sparse scattered point structure reorganization method, the scattered point set of the I-th

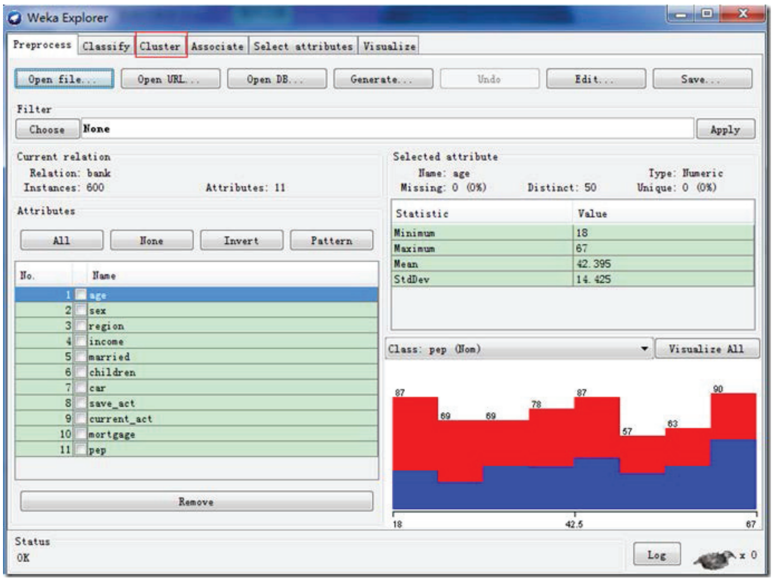

Figure 8 Simulation interface.

multilayer distributed database association rule is $P_i = (p_{i1}, p_{i2}, \ldots, p_{iD})$, wherein:

$$j \in N_i(k),\ N_i(k) = \left\{ \left\| x_j(k) - x_i(k) \right\| < r_d(k) \right\} \tag{16}$$

Multi-dimensional spatial feature extraction and adaptive fusion processing of multi-layer distributed database association rules are carried out by adopting a multi-dimensional grid partition structure rearrangement method, association rule items of multi-layer distributed database association rules are adjusted, and a fuzzy information fusion model of multi-layer distributed database association rules is constructed:

$$\begin{cases} a\left(H_{ac}\right) = 1 - \frac{H_{ac}}{\max(H_{ac})+l} \\ \max\left(H_{ac}\right) = \log_2 k \end{cases} \tag{17}$$

Under strong interference, the boundary value convergence condition of intelligent index of association rules of multi-layer distributed database satisfies the following boundary function:

$$\mathrm{w}_{ji}(k+1) = w_{ji}(k) - \alpha \frac{\partial F}{\partial w_{ji}} \tag{18}$$

$$z_{kj}(k+1) = z_{kj}(k) - \alpha \frac{\partial F}{\partial z_{kj}} \tag{19}$$

The multi-dimensional spatial feature extraction and adaptive fusion of multi-layer distributed database association rules are carried out by adopting a multi-dimensional grid partition structure rearrangement method, the key information index model of multi-layer distributed database association rules is established, and the multi-layer distributed database association rule structure reorganization is carried out by adopting a non-linear statistical sequence analysis method, so that the multi-layer distributed database association rule intelligent index model is obtained as follows:

$$X = [s_1, s_2, \cdots s_K] = \begin{bmatrix} x_1 & x_2 & \cdots & x_K \\ x_{1+\tau} & x_{2+\tau} & \cdots & x_{K+\tau} \\ \cdots & \cdots & \cdots & \cdots \\ x_{1+(m-1)\tau} & x_{2+(m-1)\tau} & \cdots & x_{M+(m-1)\tau} \end{bmatrix} \tag{20}$$

where $K = N-(m-1)\tau$ represents the embedding dimension of multi-layer distributed database association rule intelligent index, τ is the time delay, m is the number of virtual nodes and virtual link layers, and $K = N - (m-1)\tau$ is the spatial distribution feature quantity. To summarise, cluster clustering and the fuzzy C-means fusion method are adopted to realize the intelligent index of multi-layer distributed database association rule.

5. SIMULATION EXPERIMENT AND RESULT ANALYSIS

In order to verify the application performance of the method in realizing the source integration of multi-layer distributed database association rules, a simulation experiment is carried out, and the algorithm design of data source integration is carried out by combining Matlab and C++ programming software. The simulation platform interface is shown in Fig. 8.

The sample database of multi-layer distributed database association rules is from the cloud combination database Pearson Database, wherein the linear correlation coefficient of heterogeneous data in Pearson Database is set to 0.68. Regression analysis coefficient of association data is 1.35, correlation coefficient of grouping detection of association rules of multi-layer distributed database is set to 0.14, dimension of phase space recombination of association rules of multi-layer distributed database is set to $m = 4$, embedding delay is set to 12, sampling sample length of association rules of multi-layer distributed database is set to 1024, training set is set to 200, intelligent index simulation analysis of association rules of multi-layer distributed database is carried

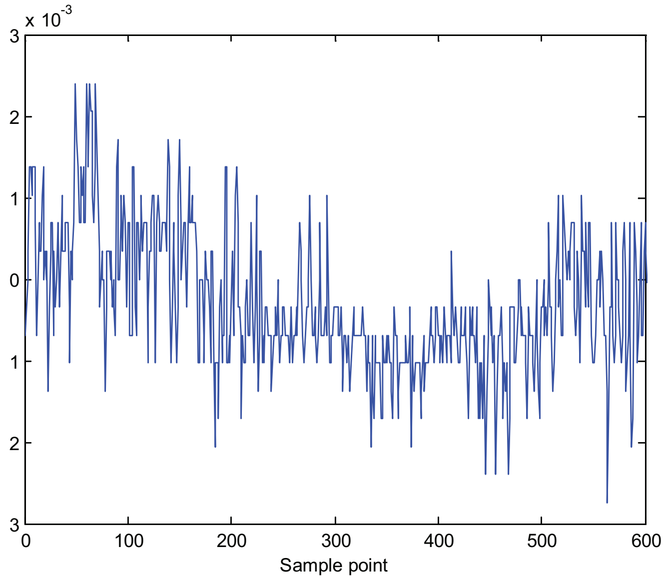

Figure 9 Sample distribution of association rules of multi-layer distributed database.

out according to the above simulation parameter settings, and the sample distribution time domain diagram of association rules of multi-layer distributed database is given as shown in Fig. 9.

Taking the above data as the research sample, and combining the rough set feature matching method, the discrete fusion processing of the association rules of the multi-layer distributed database is carried out, the feature reconstruction of the association rules of the multi-layer distributed database is realized, the source combination of the data is realized, and the intelligent index output of the association rules of the multi-layer distributed database is obtained as shown in Fig. 10.

Analysis of Fig. 10 shows that the method of this paper can effectively realize the source integration of association rules for multi-tier distributed databases with good regularity and strong data regularity integration ability. The performance of different methods for multi-tier distributed database association rules integration is tested, and the precision test results are shown in Fig. 11.

According to the analysis of Fig. 11, the data source integration using this method has a higher accuracy for multi-tier distributed databases, and the time cost of data source integration using different methods is tested. The comparison results are shown in Tables 2–4.

The analysis shows that the intelligent index of association rules for multi-tier distributed databases using this method has a higher integration degree and reduces the time cost.

6. CONCLUSIONS

The statistical analysis model of association rules in multi-tier distributed database is constructed. The optimal mining and combination control of association rules in multi-tier distributed database are carried out by using the methods of association rule scheduling and semantic ontology model design. The retrieval and adaptive mining capabilities of association rules in multi-tier distributed database are improved. The intelligent indexing method of association rules in multi-tier distributed database based on fuzzy correlation fusion clustering analysis is proposed in this paper. Combined with the rough set feature matching method, discrete fusion of association rules in a multi-tier distributed database is carried out, association rule mining and statistical regression analysis are carried out on association rule feature quantities in a multi-tier distributed database, multi-dimensional spatial feature extraction and adaptive fusion of association rules in a multi-tier distributed database are carried out, key information index model of association rules in a multi-tier distributed database is established, and intelligent index of association rules in a multi-tier distributed database is realized. Research shows that the integration of association rules in multi-tier distributed database is better, and the accuracy of a multi-tier distributed database is improved.

The improvement of the proposed algorithm is:

(1) The identity array of frequent itemsets is declared separately at each local site, i.e. each item in the candidate set has a corresponding bit in the array, and the corresponding position is set to 1 when the support technology is larger than minsup when the database is scanned, otherwise the initial value is 0. The array type is set to bit type to reduce traffic and storage;

(2) Whenever a set of items is frequent at a local site, the corresponding bits of the array are 1, effectively avoiding neglecting the counting of a particular site;

(3) After each round of frequent itemsets mining, the central site will broadcast to each local site, thereby avoiding the missing elements of the candidate itemsets;

(4) Each local site always sends an array to the global site for identification before sending the frequent itemsets and their support numbers to the central site, thus effectively reducing the traffic between sites.

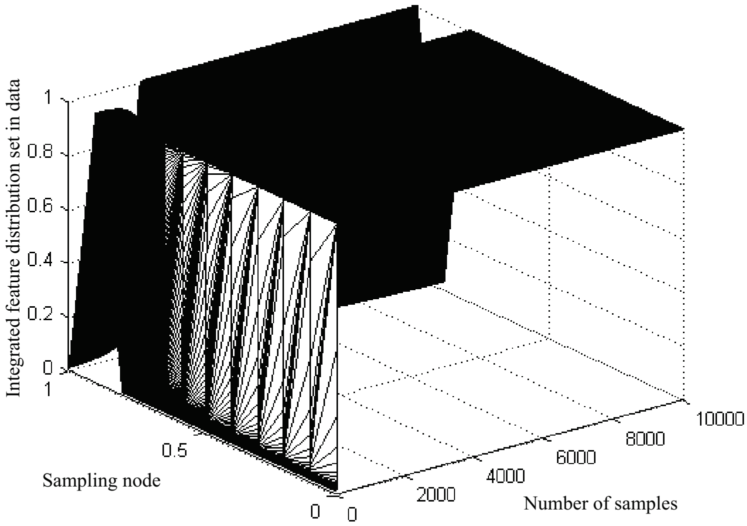

Figure 10 Intelligent index output of association rules of multi-layer distributed database.

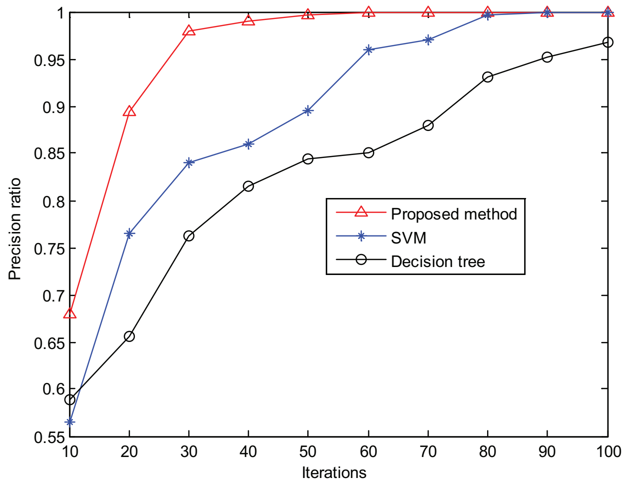

Figure 11 Comparison of database alignment performance.

Table 2 The proposed method is time-consuming (Unit: S).

	Length of time
Group 1	0.567
Group 2	0.432
Group 3	0.367

Table 3 Length of FCM method (Unit: S).

	Length of time
Group 1	3.325
Group 2	4.323
Group 3	3.467

Table 4 Length of PCA method (Unit: S).

	Length of time
Group 1	1.435
Group 2	2.321
Group 3	5.345

ACKNOWLEDGEMENTS

The study was supported by Hunan Provincial Natural Science Foundation of China (No. 2017JJ2135) and Hunan Provincial the Research Foundation of Education Bureau of China (No.18A481, No. 18B529).

REFERENCES

1. Zhao W., Huang X. J., You R. Y. Removal of White Noise from ECG Signal Based on Morphological Component Analysis, Chinese Journal of Biomedical Engineering, 23(1), (2014), 1–6.
2. Ke S. N., Gong J., Li S. N., *et al.* A Hybrid Spatio-temporal Data Indexing Method for Trajectory Databases, Sensors, 14(7), (2014), 12990–13005.
3. Ma X. T., Luo J. Q., Wu S. L. Joint Sorting and Location Method using TDOA and Multi-parameter of Multi-station, Journal of National University of Defense Technology, 37(6), (2015), 78–83.
4. Nie X. W. Radar Signal Pre-sorting Based on K-means Algorithm, Electronic Science and Technology, 26(11), (2013), 55–58.
5. Ju C. H., Zou J. B. An Incremental Classification Algorithm for Data Stream Based on Information Entropy Diversity Measure, Telecommunications Science, 31(2), (2015), 86–96.
6. Lyu Y. X., Wang C. Y., Wang C., *et al.* Online Classification Algorithm for Uncertain Data Stream in Big Data, Journal of Northeastern University (Natural Science Edition), 37(9), (2016), 1245–1249.
7. Huang S. C., Liu Y. Classification Algorithm for Noisy and Dynamic Data Stream, Journal of Jiangsu University of Science and Technology (Natural Science Edition), 30(3), (2016), 281–285.
8. Sun B., Wang J. D., Chen H. Y., *et al.* Diversity Measures in Ensemble Learning, Control and Decision, 29(3), (2014), 385-395.
9. Metwally A., Faloutsos C. V-SMART-Join, a Scalable MapReduce Framework for All-pair Similarity Joins of Multisets and Vectors, Proceedings of the VLDB Endowment, 5(8), (2012), 704–715.
10. Hashemi S., Yang Y. Flexible Decision Tree for Data Stream Classification in the presence of Concept Change, Noise and Missing Values, Data Mining and Knowledge Discovery, 19(1), (2009), 95-131.
11. Ju C H, Zou J B. An Incremental Classification Algorithm for Data Stream based on Information Entropy Diversity Measure, Telecommunications Science, 31(2), (2015), 86-96.
12. Lyu Y. X., Wang C. Y., Wang C., *et al.* Online Classification Algorithm for Uncertain Data Stream in Big Data, Journal of Northeastern University (Natural Science Edition), 37(9), (2016), 1245-.1249.
13. Chen Y., Li L. J. Very Fast Decision Tree Classification Algorithm based on Red-black Tree for Data Stream with Continuous Attributes, Journal of Nanjing University of Posts and Telecommunications (Natural Science Edition), 37(2), (2017), 86–90.
14. Qiao Z. W., Sun W. X. Two Improvements to C4.5 Algorithm, Journal of Jiangsu Polytechnic University, 20(4), (2008), 56–59.
15. Huang S. C., Liu Y. Classification Algorithm for Noisy and Dynamic Data Stream, Journal of Jiangsu University of Science and Technology (Natural Science Edition), 30(3), (2016), 281–285.
16. Hu W. K., Chen T. W. Discovering Association Rules of Mode-dependent Alarms from Alarm and Event Logs, IEEE Transactions on Control Systems Technology, 3(15), (2017), 1–13.
17. Özgür Cingiz M., Biricik G., Diri B. ARNetMiT R Package: Association Rules Based Gene Co-expression Networks of miRNA Targets, Cellular and molecular biology (Noisy-le-Grand, France), 63(3), (2017), 18–22.
18. Zhang Z. J., Huang J., Hao J. G., *et al.* Extracting Relations of Crime Rates Through Fuzzy Association Rules Mining, Applied Intelligence, 12(4), (2019), 40–48.
19. James H., Lynne D. Talley, John G., *et al.* An Argo Mixed Layer Climatology and Database, Geophysical Research Letters, 44(11), (2017), 120–130.
20. Kenneth P. B., Namratha R. V., Brennan M., *et al.* Multi-database Description of Primary Splenic Diffuse Large B-Cell Lymphoma, Anticancer Research, 37(12), (2017), 6839–6843.
21. Perez C. I., Chansangpetch S., Thai A., *et al.* Normative Database and Color-code Agreement of Peripapillary Retinal Nerve Fiber Layer and Macular Ganglion Cell-Inner Plexiform Layer Thickness in a Vietnamese Population, Journal of Glaucoma, 27(3), (2018), 67–81.