Weak Association Mining Algorithm for Long Distance Wireless Hybrid Transmission Data in Cloud Computing

Simayi Xuelati(✉), Junqiang Jia, Shibai Jiang, Xiaokaiti Maihebubai, and Tao Wang

State Grid Xinjiang Electric Power Co., Ltd. Information and Communication Company, Urumqi 830001, China
1484394089@qq.com

Abstract. Long distance wireless hybrid transmission data is vulnerable to noise, resulting in low data mining accuracy, large mining error and poor mining effect. Therefore, a weak association mining algorithm for remote wireless hybrid transmission data under cloud computing is proposed. The moving average method is used to eliminate noise data, and the attribute values of continuous data are divided into discrete regions, make it form a unified conversion code for data conversion. The Bayesian estimation method is used for static fusion to eliminate the uncertain data with noise. The rough membership function is constructed to distinguish the truth value, complete data preprocessing. According to the principle of relationship matching between data, data feature decomposition is realized. The non sequential Monte Carlo simulation sampling method is adopted to build the data loss probability evaluation model and integrate the data association rules. In the background of cloud computing, permission item sets are generated, and the rationality of association rules is judged by the minimum support. The dynamic programming principle is used to build the mining model, and the improved DTW algorithm is used to read out and analyze the structured, semi-structured and unstructured data to obtain the weak association mining results of mixed data transmission. The experimental results show that the algorithm can completely mine data sets, and the mining error is less than 0.10, with good mining results.

Keywords: Cloud Computing · Long Distance Wireless · Mixed Transmission · Weak Association Mining

1 Introduction

The continuous progress of information technology and science and technology has promoted the development of new information technologies in different fields. Data mining technology is a new technology in the information age. In the Internet, data mining refers to collecting high-value models and rules, using data analysis tools to analyze the collected data models and data information, obtaining the differences and similarities between the data models and data information according to the analysis results,

B. Wang et al. (Eds.): ICMTEL 2023, LNICST 534, pp. 62–78, 2024.
https://doi.org/10.1007/978-3-031-50577-5_6

and using the above results to make predictions on the Internet. With the development of communication network and information age, great changes have taken place in the network mode and the application field of communication network. Mining the valuable content and data in communication network can improve the utilization value of communication network. According to the analysis results, the data rules and data system of the communication network are optimized to establish a safe and stable communication environment and provide people with high-quality and efficient communication services. Therefore, it is of great significance to mine the weak association of mixed data in communication networks. Long distance wireless hybrid transmission under cloud computing brings a new development direction to the network industry, and also brings unprecedented challenges. With the rapid development of information technology, a large number of information is generated under the complex requirements of resource management. Promote information integration and safe operation through the use of advanced information and communication technologies [1]. However, due to the requirements of various types of network services, different functions are used in different periods, which makes the data model and information inconsistent. Massive wireless mixed transmission data is difficult to flexibly realize information sharing and provide data needed for the development of intelligent Internet, which has become a bottleneck restricting the rapid improvement of Internet automation.

At present, there are two main data association mining methods. Reference [2] proposed a fuzzy association rule mining method based on GSO optimized MF in deterministic data. Firstly, the uncertain data is represented by the ternary language representation model. Then, given an initial MF, and taking maximizing the support of fuzzy itemsets and semantic interpretability as the fitness function, the optimal MF is obtained through the optimization learning of GSO algorithm. Finally, according to the best MF obtained, an improved FFP growth algorithm is used to mine fuzzy association rules from uncertain data. This method can adaptively optimize MF according to data sets, so as to effectively mine association rules from uncertain data. However, the efficiency and operation speed of this method are slow when dealing with large amounts of data. Reference [3] proposed an algorithm based on Apriori association rules. The specific method is to analyze each band first, and then use these bands to generate stronger correlation. To apply the Apriori association rule algorithm, first scan multiple databases, and then generate a large amount of commonly used candidate objects, which makes the Apriori algorithm time and space complexity. When mining large amounts of data, the performance of Apriori algorithm is poor. Therefore, a weak association mining algorithm for long-distance wireless hybrid transmission data under cloud computing is proposed.

2 Data Preprocessing of Long-Distance Wireless Hybrid Transmission Under Cloud Computing

The remote wireless hybrid transmission data under the original cloud computing has serious quality problems, such as data loss, data redundancy, noise data, etc., which will reduce the data mining efficiency of remote wireless hybrid transmission. Under the condition of ensuring data integrity, reasonable and effective data preprocessing is the basis of improving data mining efficiency.

2.1 Noise Data Elimination

Mixed transmission data cleaning is to remove and repair the incomplete and noisy data in the data. In the original database, the average value is usually used to fill in the incomplete data. This process needs to be implemented using the moving average method. Moving average method is a key method to take the average value of data at a certain stage as the prediction value of a certain period in the future, and use this data as the later mining data. The formula for calculating the moving average is as follows:

$$a=\sum \frac{l}{n} \tag{1}$$

In formula (1): l represents the moving length; n is the total number used by the moving average. In order to facilitate that all data in the database have the same attributes, its transformation rules need to be defined and unified before mining. Because noise data is illogical deviation data, which often affects the accuracy of data mining, data smoothing technology is used to eliminate noise data.

2.2 Data Conversion

The data conversion mainly includes data normalization and continuous data discretization. The continuous data is discretized by compressing the original data to reduce data input and output. The attribute values of continuous data are divided into several discrete regions, and values are taken from them in order to reduce the number of attribute values and facilitate data mining. The data conversion process is to unify different data formats to form a unified conversion code [4]. There is usually detailed data in massive data, and the data in the data warehouse is used for analysis. The data can be aggregated according to the database granularity without detailed data. Different data stores have different storage rules. These rules can not be implemented simply by adding, subtracting, or subtracting. These data need to be calculated and stored in the database for subsequent mining analysis and use.

2.3 Data Truth Screening

Mixed transmission data is composed of network data from different periods, systems or departments, including structured data (such as department data, XML, JSON, etc.), unstructured data (such as remote sensing images, research plans, drawings, etc.), and semi-structured data (such as policies and regulations listed on this page). In order to facilitate data analysis and integration, data attributes, storage methods, read/write methods, etc. need to be preprocessed [5]. The data integration mode of mixed transmission needs to be further studied to realize deeper recognition of truth value. Data conflict refers to the problem of large margin provided by multiple sensors at the same time and differences in redundant data of the same attribute. The usual way to solve the conflict is to minimize the redundancy of data and find the true value from multiple conflicting data. The true value includes the unity of attributes. Because there are replication, dependency and digital similarity between data, Bayesian estimation is used for static fusion, The information of the algorithm conforms to the probability distribution, and can well handle uncertain data with noise [6]. Bayesian estimation method optimizes the data by setting conditions in advance, integrates the information of each sensor with probability principle, and expresses it with probability density function. The basic process is to assume that there is a non empty object set, a non empty condition attribute set and a non empty decision attribute set in a decision system. Construct an approximation of existing knowledge and indiscernibility relation x, which can be expressed as:

$$x_S = (S|[S]_x \cap S \neq \emptyset) \tag{2}$$

In formula (2), $[S]_x$ is the equivalent of x. In the rough set theory, according to the knowledge and the calculation results of indiscernibility relations, the x attribute of the decision set is analyzed and expressed by the rough membership function:

$$\gamma_S^x(S) = \frac{\left|[S]_x \cap^S\right|}{|[S]_x|} \tag{3}$$

The true value is discriminated according to the rough membership function determined above. Assume that the decision set is $S_1, S_2, \cdots S_m$, and the observation result obtained through the sensor observation data is G, and then calculate the posterior probability. The formula is:

$$P(S_m|G) = \frac{P(S_m G)}{P(G)} = \frac{P(G|S_m)P(S_m)}{\sum_{m=1}^{i} P(G|S_m)P(S_m)} \tag{4}$$

In formula (4), $P(S_m)$ represents a prior probability; $P(G|S_m)$ stands for conditional probability; i represents the number of sensors, and the true value is identified according to the decision posterior probability.

3 Weak Association Mining of Mixed Transmission Data in Cloud Computing

In the process of mining massive data in the cloud computing environment, it is easy to have a large number of redundant data, which reduces the correlation between data and can not effectively complete the mining of weak associations [7]. Therefore, a weak association mining method based on weak clustering algorithm for massive data in cloud computing environment is proposed, which decomposes data features through data description features, and fuses all data according to data features. Based on the association decision probability, the massive data in the cloud computing environment is effectively divided, and the calculation of the association probability of all data features is completed. Attribute elements are classified by the weak clustering method, and quantitative elements are converted into category types. The data after clustering is mined by weakening the association rule method [8].

3.1 Principle of Relationship Matching Between Data

Massive data in the cloud computing environment have a certain relationship with each other. It is the focus of research to mine the parts associated with other data by describing the information expressed by the data. The whole principle is as follows: First, get the description characteristics of the data, complete the data feature decomposition, and then obtain the following feature matrix:

$$R=\begin{bmatrix} R_{11} & R_{12} & \cdots & R_{1j} \\ R_{21} & R_{22} & \cdots & R_{2j} \\ \cdots & \cdots & \cdots & \cdots \\ R_{l1} & R_{l2} & \cdots & R_{lj} \end{bmatrix} \tag{5}$$

In Formula (5), l represents the amount of data, and j represents the number of data types. Transform the data characteristic matrix, and the average value of the characteristics obtained is:

$$A = l^2 \cdot \sum_{j=2}^{j} R_{lj} \tag{6}$$

All data are fused according to data characteristics [9]. The massive data in the cloud computing environment is effectively divided based on the association decision probability, and the association probability of all data features is calculated by the following formula:

$$f\left(g_l \cdot d_j\right) = \left(f\left(g_l\right) + f\left(g_l \cdot \delta_j\right)\right) / f\left(d_j\right) \tag{7}$$

In Formula (7), $f\left(d_j\right)$ represents the prior probability of data; $f\left(g_l \cdot \delta_j\right)$ represents the corresponding conditional probability.

Because the information objects in the remote wireless hybrid transmission database under cloud computing are described by the same remote wireless hybrid transmission data specification, in different cases, the internal nodes of the remote wireless hybrid transmission data standard tree are part of the data specification. The difference is that the element value at the node [10]. The approximate matching process of long-distance wireless mixed transmission data is shown in Fig. 1.

As shown in Fig. 1, comparing the query tree with the standard tree, it is found that the nodes corresponding to node b in the multimodal data standard tree are $b1$, $b2$ and $b3$, and the nodes corresponding to node b' in the query tree are $b'1$ and $b'2$. According to these, it is unnecessary to match the metadata of object $U2$ with the query tree data when matching the query tree data with the standard tree data. When the query tree matches the metadata tree of $U3$, since no node on the subtree can match the node in the query tree, it is unnecessary to consider the matching of the subtree with node $b3$ as the root node. Before querying, first match the query tree with the multimodal data standard scheme tree of the resource target database, and then record the matching information (i.e. preprocessing information) of the associated nodes. By analyzing the obtained information, we can avoid a large number of non associated nodes matching in the future matching between the query tree and the multimodal data standard tree, and avoid unnecessary duplication.

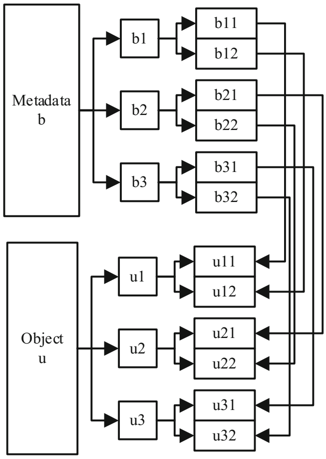


Fig. 1. Approximate matching process of long distance wireless hybrid transmission data

3.2 Mining Model Construction

Before data mining, it is necessary to evaluate the remote wireless hybrid transmission under cloud computing. However, the actual remote wireless hybrid transmission has randomness and dynamic volatility. Therefore, a probability evaluation model for transmission data loss under mixed fluctuation conditions is designed to achieve data loss analysis under combined factor interference.

Data integration is required before analysis. According to the results of data preprocessing and truth screening, the overall plan of data integration based on decision rough set model is designed, as shown in Fig. 2.

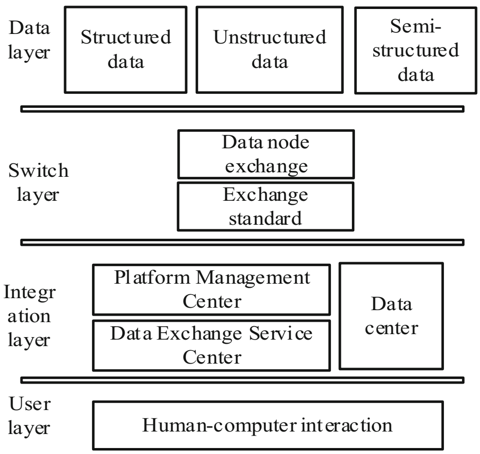


Fig. 2. Data integration scheme

It can be seen from Fig. 2 that the scheme is composed of data layer, exchange layer, integration layer and user layer, and the data layer contains different data types; The main function of the exchange layer is to realize two-way data transmission; The integration layer is the core of the whole system. Its main functions include data conversion, data call, process control and data routing; The user layer is responsible for user login, authority management and maintenance.

On this basis, the non sequential Monte Carlo simulation sampling method is used to build the data loss probability evaluation model. The specific process is as follows:

First, the probability space of wireless network data is simulated by non sequential Monte Carlo, where the statistical value is calculated as follows:

$$x_{\mathrm{i}}' = \frac{1}{g}\sum_{i=1}^{n} z_i \tag{8}$$

In formula (8), g represents a restricted constraint function, and z represents a series of random data. After the data loss of wireless network is constrained, the fluctuation probability of each time point is taken as. According to the daily transmission data volume, loss data volume and loss rate of wireless network statistics, when the loss data fluctuates, the loss data is calculated using power flow calculation coefficient, wireless network loss data, transmission data and loss rate. After evaluating the degree of wireless network data loss, define big data association rules, integrate data association rules, and build a mining model based on big data analysis. The detailed implementation process is as follows:

In the background of cloud computing, permission item sets are generated, association rules are generated according to the minimum support and minimum confidence, and the rationality of association rules is judged by the minimum support. In the context of cloud computing, the association rules are as follows:

$$D \Rightarrow \frac{\sup(x)}{\sup(x \Rightarrow y)} \frac{\sup t_{\min}}{p} \tag{9}$$

In Formula (9): $\sup t_{\min}$ represents the minimum support threshold; p indicates association rule item; $\sup(x)$ stands for support; $\sup(x \Rightarrow y)$ stands for confidence.

According to the above formula, the mining model needs to be built using the principle of dynamic programming, as shown below:

$$M_{xy} = \max D \begin{cases} E_{x-1,y-1} - J_{xy} \\ E_{x,y-1} + p \\ E_{x-1,y-1} + p \end{cases} \tag{10}$$

In Formula (10): E_{ab} is the missing parameter of wireless network association rules; J_{ab} indicates useless weak association rules. According to the above model, multiple data sources are integrated to extract data related to mining. Through data transformation, it is unified into a form suitable for mining, and visualization technology is used to provide users with mining data.

3.3 Mining Process of Transmission Data Weak Association

In the process of weak association mining of transmitted data, frequent itemsets can be generated through the relationship between time and space, and frequent itemsets can be generated through the minimum collection cycle. On this basis, improve the DTW algorithm to improve the accuracy of data mining.

Combined with the improved DTW algorithm, the speed of mining weak associations of transmitted data is greatly improved. The detailed steps are as follows:

Step 1: Because the improved DTW algorithm needs a lot of calculation steps in the process of association mining, it takes up a lot of storage space. Therefore, in order to solve this problem, a path of data weak association mining is designed. Set the number of sampling points as y_1, y_2, y_3, and divide the data into three dimensions, namely, one-dimensional $\left[1, y_1\right]$, two-dimensional $\left[y_1 + 1, y_2\right]$, and three-dimensional $\left[y_2 + 1, y_3\right]$. For the calculation of y_1 and y_3 values, it can be expressed as:

$$\begin{cases} y_1 = \frac{\varepsilon c + \gamma - 1}{\gamma - \varepsilon} \\ y_3 = \frac{\gamma c - \varepsilon + 1}{\gamma - \varepsilon} \end{cases} \quad (11)$$

In formula (11), c represents the number of sampling points; ε, γ indicates the slope of two adjacent sides of the parallelogram. When the mining data is not inside the parallelogram, it indicates that these data do not have relevance and need not be mined; On the contrary, it is associated and can be mined. According to the mining results, data sets are collected.

Step 2: Scan all data sets and record the number of data occurrences each time. Determine whether the time and spatial data are in the same dimension according to the requirement definition, and record them in the header table if they exist;

Step 3: Recycle the dataset, delete the data not in the item header table, and arrange the data in the order of increasing the item header table. Loop the data set again to generate a frequent pattern tree. In the frequent pattern tree, all nodes represent spatial and temporal data, while the tree branches represent the number of data occurrences;

Step 4: In the circular header table, search the entries in the regular pattern tree and the leaf nodes of the entries in descending order, and remove the duplicate node data to obtain a separate tree structure dataset. At this time, the dataset is an associated set.

Step 5: Output the tree data set of all single paths to form the final result set.

Step 6: First, read and analyze the filtered structured, semi-structured and unstructured data from the hybrid transmission database through the above model. For structured data, establish a database according to the data type and directly select it into the database; Semi structured data can be divided into two categories according to data types: structured data and unstructured data. In Category 1, create an associated class library and connect it directly to the class library. For the unstructured data in category 2, using the mapping relationship between different data, the corresponding data is embedded into these data as additional parameters to achieve a one to many direct mapping; In Category 2, coordinate transformation is carried out for various types of data, and then the single level segmentation algorithm of artificial neural network is used to partition. On this basis, structured data and unstructured data are combined into multiple single level fusion parameters, which are overlapped with quasi space to obtain the integration results.

Step 7: Regard the integration result as a fuzzy attribute set, and build a fuzzy database through the original database. Let spatial data $T\left(y_{x_1}\right)$ and temporal data $T\left(y_{x_2}\right)$ be the support degree of spatial data x_1 and temporal data x_2 respectively. The support degree of rule $x_1 \Rightarrow x_2$ in database H can be expressed as:

$$T(x_1 \Rightarrow x_2) = \frac{\sum_{i=1}^{m} T_i(x_1 \cup_2^x)}{|H|} \tag{12}$$

It can be seen from Formula (6) that in the fuzzy association relationship, it is necessary to calculate the fuzzy support degree, that is, the implication degree, which can effectively reduce the mining steps and shorten the mining time. The k-th data implication can be expressed as:

$$\nu(x_1 \Rightarrow x_2) = I[T(x_1), T(x_2)] \tag{13}$$

In formula (13), I represents the implication degree operator. By calculating support, frequent itemsets can be determined, and the result is the weak association mining result of mixed data transmission.

4 Experimental Analysis

In order to verify the effectiveness of the weak association mining algorithm for long-distance wireless mixed transmission data under cloud computing, the GSO-based optimization MF method (the method of reference [2]) and Apriori-based association rule algorithm (the algorithm of reference [3]) are used as comparison methods, experiments are conducted on the Matlab platform through Unix operating system.

4.1 Experimental Platform

As this experiment is based on the laboratory environment, the cloud platform is designed as a private cloud platform for LAN testing in the design, but the core modules can still be used under the public cloud platform. A standard private cloud platform should include at least four platforms: the underlying virtualization platform, virtual machine management platform, distributed storage platform and distributed computing service platform. Therefore, the designed architecture diagram of the test private cloud platform is shown in Fig. 3.

It can be seen from Fig. 3 that the experimental platform provides a place for data integration, through which data can be found in time and uploaded to the response system to generate feedback information.

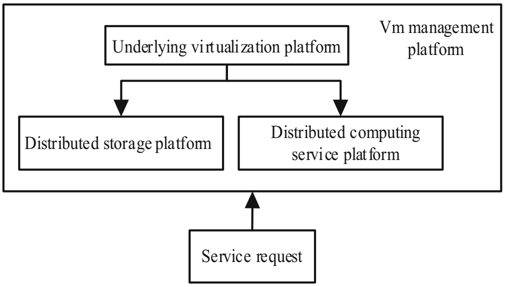


Fig. 3. Experimental platform

4.2 Experimental Data Set

The experiment used a set of public data sets, with a total of 2500 documents, each of which contains a picture with corresponding instructions. Each picture and a group of files correspond to a specific category directory, and all information in the category directory can be divided into 15 categories. The SIFT feature description method is used to describe the image as a 128 bit feature vector. Text of the dataset is presented in the form of 10 topics. During the experimental test, 2/3 of the data are used as training data and 1/3 of the data are used as test data.

According to the difficulty of field data acquisition, a data acquisition and analysis system is used, which mainly includes two parts: the field end and the communication end. The field terminal is used to collect and transmit field data. The communication terminal is responsible for field data processing and background data interaction. Its architecture is shown in Fig. 4.

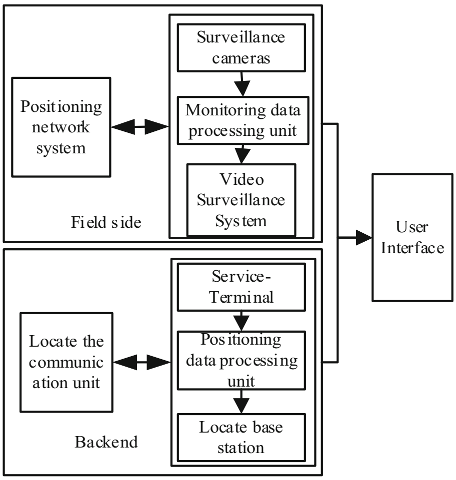


Fig. 4. Experimental data acquisition architecture

On the field side, cameras, positioning base stations, smart helmets, identification tags and other sensors can be used to collect infrastructure in real time, and upload the collected information to the server, providing good data support for infrastructure management. The data collection results are shown in Table 1.

Table 1. Data collection results

Wireless transmission	Document/KB	Picture/KB
Port # 1	75	21
Port # 2	80	22
Port # 3	68	20
Port # 4	63	17
Port # 5	59	18
Port # 6	67	21
Port # 7	66	19
Port # 8	62	18

Take the data in Table 1 as the standard data for experimental verification and analysis.

4.3 Standardized Processing of Experimental Data

According to the data characteristic attribute rules, the data is standardized. The change of trust interval in this process is shown in Fig. 5.

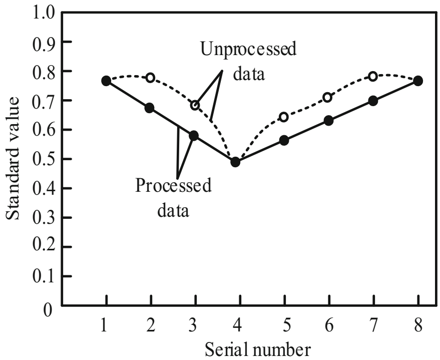


Fig. 5. Change of trust interval in data integration process

It can be seen from Fig. 5 that only the processed value under No. 4 is consistent with the untreated value, and it is taken as the best data integration point to verify the data mining effect on the premise of ensuring the integration stability.

4.4 Determination of Experimental Indicators

Set the experimental indicators as mining integrity rate and mining error, and the mining integrity rate calculation formula is as follows:

$$P = \frac{\partial}{\partial'} \times 100\% \tag{14}$$

In formula (14), ∂ represents the amount of data mined; ∂' represents the total data volume. The larger the calculation result is, the more complete the data mining result is.

The mining error calculation formula is as follows:

$$e = \frac{\left|\delta_1 - \overline{\delta_c}\right| + \left|\delta_2 - \overline{\delta_c}\right| + \cdots \left|\delta_\tau - \overline{\delta_c}\right|}{\tau} \tag{15}$$

In formula (15), τ represents the number of times of excavation; $\overline{\delta_c}$ indicates information whose data has not been searched. The larger the calculation result is, the more accurate the data association mining result is.

4.5 Analysis of Data Integration Results

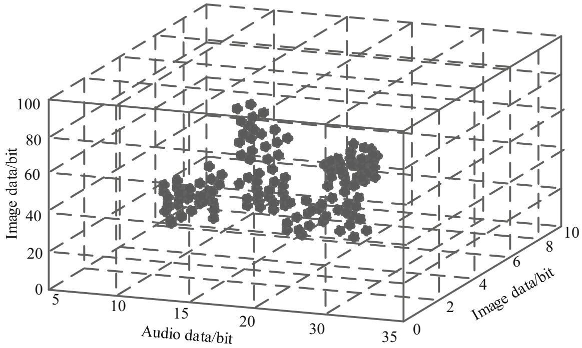
Through the established experimental platform, data are collected in the open data set, standardized processing is carried out, experimental indicators are set, and data integration analysis is carried out. For the verification of the integration effect, the GSO optimized MF method, Apriori based association rule algorithm and cloud computing based weak association mining algorithm are respectively used for integration. The comparison results are shown in Fig. 6.

It can be seen from Fig. 6 that the data integration based on Apriori association rule algorithm is too decentralized, and the data integration process based on GSO optimized MF method does not reach the full integration state. However, the weak association mining algorithm under cloud computing can integrate all data together, so it can be seen that the weak association mining algorithm has a good integration effect.

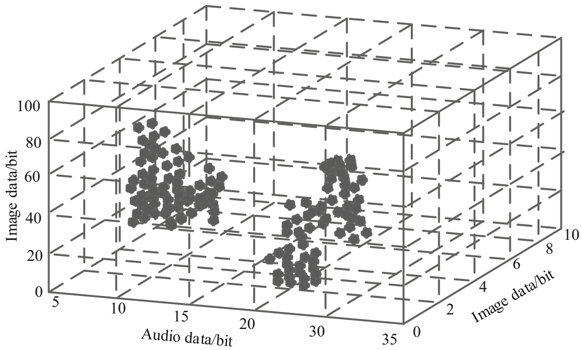
4.6 Analysis of Data Mining Results

The complete mining results of actual data are shown in Table 2.

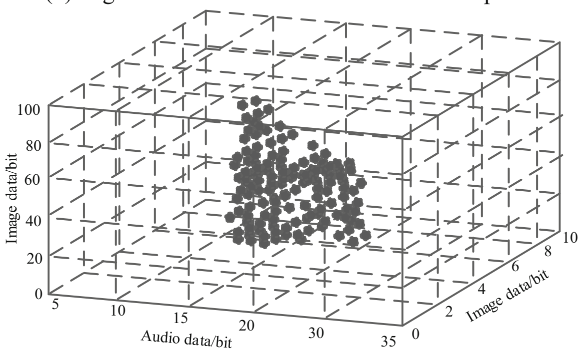
Based on the data in Table 2, use the GSO optimized MF method, the Apriori association rule algorithm and the cloud computing weak association mining algorithm to mine, as shown in Fig. 7.



(a) Optimization of MF method based on GSO



(b) Algorithm of association rules based on Apriori



(c) Weak association mining algorithm in cloud computing

Fig. 6. Comparison and analysis of integration effects of different methods

It can be seen from Fig. 7 that the instance dataset collected using the GSO based optimization MF method is inconsistent with the actual situation, and its dataset is {1, 2, 3, 4, 5, 6, 7, 9}; The instance dataset obtained based on Apriori association rule algorithm

Table 2. Data complete mining analysis

Item	Data
1	{1, 4, 5}
2	{3, 4, 5, 6}
3	{1, 6}
4	{1, 5}
5	{2, 4, 6, 7}
6	{3, 5, 7}
7	{2, 5, 6}
8	{1, 3, 4}

is inconsistent with the actual situation, and its dataset is $\{1, 2, 3, 4, 5, 6, 8\}$; The instance dataset obtained by using the weak association mining algorithm under cloud computing is consistent with the actual situation, and its dataset is $\{1, 2, 3, 4, 5, 6, 7\}$. It can be seen that the data can be completely mined using this method.

For data mining error analysis, the three methods are compared again, and the comparison results are shown in Fig. 8.

It can be seen from Fig. 8 that the mining error of GSO-based optimization MF method and Apriori-based association rule algorithm is about 0.32 and 0.35 respectively, while the mining error of weak association mining algorithm under cloud computing is less than 0.10. It can be seen that the mining error using the method studied is small.

data

{1，2，4} {1，2，5}
{1，4，5} {2，3， 6}
{2，4，5} {2，5，6}
{2，5，7} {2，6，7}
{5，6，7} {5，6，9}
{1，2，3，4，5，6，7，8，9}

item	data
1	{5，6}
2	{1，3，5}
3	{1，3，6}
4	{2，4，7}
5	{3，4，5，6，7}
6	{4，5，9}
7	{1，2}
8	{2，4，5}

(a) Optimization of MF method based on GSO

data

{1，2，4} {1，2，5}
{1，4，5} {2，3， 6}
{2，4，5} {2，5，6}
{2，5，7} {2，6，7}
{5，6，7} {5，6，9}
{1，2，3，4，5，6，7，8，9}

item	data
1	{1，2，4，5，6}
2	{2，3，4，5}
3	{1，3，6}
4	{2，4}
5	{3，4，5，6}
6	{4，8}
7	{1，2，6}
8	{2，4，5}

(b) Algorithm of association rules based on Apriori

data

{1，2，4} {1，2，5}
{1，4，5} {2，3， 6}
{2，4，5} {2，5，6}
{2，5，7} {2，6，7}
{5，6，7} {5，6，9}
{1，2，3，4，5，6，7，8，9}

item	data
1	{1，4，5}
2	{3，4，5，6}
3	{1，6}
4	{1，5}
5	{2，4，6，7}
6	{3，5，7}
7	{2，5，6}
8	{1，3，4}

(c) Weak association mining algorithm in cloud computing

Fig. 7. Comparison and analysis of complete data mining results of different methods

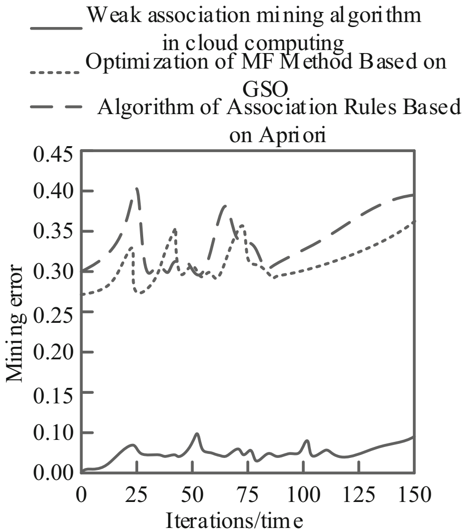


Fig. 8. Comparative analysis of data mining errors of different methods

5 Conclusion

A weak association mining algorithm for long-distance wireless mixed transmission data under cloud computing is proposed. Pre-process the long-distance wireless mixed transmission data, and realize the weak association mining of transmission data through the principle of data relationship matching and the construction of mining model. The data weak association mining under cloud computing studied in this paper can improve the transmission quality of long-distance wireless mixed data, eliminate redundant data in long-distance wireless mixed data, and use the implication degree to mine the weak association of data, which solves the problems existing in current data weak association mining methods and provides a favorable guarantee for the development of data mining technology. By constructing the rough membership function, the truth value is discriminated. By solving conflict problems, efficient data integration can be achieved, which improves the data integration effect to a certain extent. The experimental results show that using the weak association mining algorithm under cloud computing can integrate all data together, and the obtained instance data set is consistent with the actual situation. The mining error is less than 0.10, which can effectively realize the weak association mining of long-distance wireless mixed transmission data.

While obtaining the above research results, due to the limitation of lost data, there are still some areas to be optimized: (1) The coverage of the mining model should be further expanded, and the main transmission data should be analyzed during the verification process; (2) Establish a unified model transmission mechanism to achieve efficient control of the model.

References

1. Yin, Y., Jia, Y., Wang, C., et al.: Research on multi-load remote wireless power transfer system. Power Electron. **55**(09), 139–142 (2021)
2. Zhang, W., Dong, Q.: Fuzzy association rules mining method based on GSO optimization MF in uncertainty data. Appl. Res. Comput. **36**(08), 2284–2288 (2019)
3. Wang, M., Zhu, X.: Analysis of association rules based on improved Apriori algorithm. Comput. Sci. Appl. **11**(06), 1706–1716 (2021)
4. Wang, R., Zhao, L., Hu, S.: Fast estimation method for power loss of three phase unbalanced distribution network based on data correlation mining. Water Resour. Power **39**(05), 202–206 (2021)
5. Wang, P., Meng, Y.: Simulation of mining frequent pattern association rules of multi-segment support data. Comput. Simul. **38**(05), 282–286 (2021)
6. Xin, C., Guo, Y., Lu, X.: Association rule mining algorithm using improving treap with interpolation algorithm in large database. Appl. Res. Comput. **38**(01), 88–92 (2021)
7. Wang, X.: Research on PDE-based improved method for correlation feature data mining. Modern Electron. Tech. **44**(18), 111–113 (2021)
8. Jiang, F., Yuen, K.K.R., Lee, E.W.M.: Analysis of motorcycle accidents using association rule mining-based framework with parameter optimization and GIS technology. J. Safety Res. **75**, 292–309 (2020)
9. Liu, S., Hu, R., Wu, J., et al.: Research on data classification and feature fusion method of cancer nuclei image based on deep learning. Int. J. Imaging Syst. Technol. **32**(3), 969–981 (2022)
10. Liu, S., Liu, D., Muhammad, K., Ding, W.: Effective template update mechanism in visual tracking with background clutter. Neurocomputing **458**, 615–625 (2021)