

Retraction

Retracted: Application and Analysis of Hypergraph Association Rule Redundancy Algorithm in Data Mining

Mobile Information Systems

Received 26 September 2023; Accepted 26 September 2023; Published 27 September 2023

Copyright © 2023 Mobile Information Systems. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] H. Pang and L. Zhou, "Application and Analysis of Hypergraph Association Rule Redundancy Algorithm in Data Mining," *Mobile Information Systems*, vol. 2022, Article ID 1193586, 11 pages, 2022.

Research Article

Application and Analysis of Hypergraph Association Rule Redundancy Algorithm in Data Mining

Huanli Pang¹ and Lianzhe Zhou ²

¹School of International Education, Changchun University of Technology, Changchun 130012, Jilin, China

²School of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, Jilin, China

Correspondence should be addressed to Lianzhe Zhou; zhoulianzhe@ccut.edu.cn

Received 28 July 2022; Revised 17 September 2022; Accepted 24 September 2022; Published 11 October 2022

Academic Editor: Xuefeng Zhang

Copyright © 2022 Huanli Pang and Lianzhe Zhou. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In order to realize “from individual data research to data system research” and “from passive data verification to active discovery,” this study proposes a hypergraph-based association rule redundancy processing algorithm in data mining. This study introduces the concepts of hypergraph and system, explores the establishment of hypergraph on a three-dimensional matrix model, and adopts a new hyperedge definition method according to the characteristics of big data and the concept of the system, which improves the ability to deal with problems; the association rules are transformed into a directed hypergraph, and the adjacency matrix is redefined. The detection of redundancy and loops is transformed into the processing of connected blocks and cycles in the hypergraph. The experimental results show that two UCI datasets were selected, namely, the balloons dataset and the shuttle landing control dataset, in which the minimum support and minimum confidence of the balloons dataset are both 5%. The dataset has 4 attributes, and 18 association rules are obtained through the Aprior algorithm. Although the running time of the coevolution algorithm is slightly longer than that of the other two global optimization algorithms, the running time is completely within the acceptable range. Moreover, due to the effective introduction of the idea of coevolution, compared with the use of the other two algorithms for association rule mining, it not only has a better mining quality but also has a significant advantage in the ability to jump out of the local optimal solution, realizing the search of high-quality association rules in high-dimensional datasets. *Conclusion.* This model provides a new idea and method for the redundant processing of association rules.

1. Introduction

The Internet of things, cloud computing, and other information technologies are updating day by day and constantly integrated with the human world, economy, politics, military, scientific research, life, and other fields. The speed of data generation is rapid, and the amount of data is increasing day by day, giving birth to a huge amount of data [1]. Data visualization aims to express data clearly and effectively through a graphical representation and uses visualization to find data connections that are not easy to observe in the original data. Information visualization is conducive to enhancing users' understanding of high-dimensional and large-scale data and plays an important role in association rule mining, recognition, and understanding. As an

important method of knowledge discovery and pattern recognition, association rule mining aims to find valuable relationships in the form of if then. Association rule visualization is an indispensable subset of association rule research. Its main goal is to display data and help users with insight into the results of association rule mining.

Data mining can explore the hidden rules in data and give full play to the value of data. Association rule mining can extract potential and valuable frequent patterns or relationships between attributes from data [2]. Text can clearly and intuitively show frequent patterns and related relationships, but due to the limited cognitive ability of users, the value of association rule mining cannot be fully reflected. Hypergraphs are widely used in many fields of information science. Previous information visualization and visual

analysis techniques mainly focused on the simple binary information between data objects. However, the research found that multiple relationships can more natural express the internal relationships and patterns hidden in information. A hypergraph is a generalization of ordinary graphs in a topological structure, and it can intuitively show multivariate relationships [3]. This also provides strong conditions and theoretical support for the visualization of association rules. The hypergraph model combines the advantages of hypergraph and directed graph and can be used for a visual representation of association rules. Nodes in the graph represent data items, and edges represent association relationships. The support and confidence of rules can be expressed by different values and colors.

2. Literature Review

Khan et al. proposed a big data entity recognition algorithm based on a graph, which maps high-dimensional data relationships in the graph, where the edge represents some data relationships, and the weight on the edge represents the near degree of association between items. This method avoids the explicit calculation of the degree of association between high-dimensional data and has made corresponding progress in the reduction of high-dimensional data [4]. Skuratovskii et al. proposed the concept of neighborhood knowledge granularity from the perspective of granular computing to evaluate the granulation ability for high-dimensional data features and combined it with neighborhood dependency as a heuristic function for data attribute reduction [5]. Baert et al. analyzed the concept of information granularity and granularity division for the granularity selection of characteristic attributes of high-dimensional data systems, accurately reflected the roughness of data dimensions in decision-making systems, and made up for the defect of reduction based on domain attributes only when high-dimensional data are granulated in a big data environment [6]. Shekhawat et al. proposed to define neighborhood complementary entropy and neighborhood complementary conditional entropy through analytical simulation and replacement of information particles, so as to obtain nonmonotonic high-dimensional data attribute granulation and nonmonotonic high-dimensional data attribute reduction. In the research process on high-dimensional data features, the above three algorithms ensure data value, accurately capture data features, and reduce data complexity. Therefore, how to preprocess high-dimensional data through granulation and accurately capture their data characteristics is a hot issue in high-dimensional data mining [7]. Elmanakhly et al. proposed the load balancing strategy of high-low frequency division and grouped the nodes evenly by estimating the number of tasks, to avoid the problems of data skew and overload [8]. Liu et al. proposed a TBLB algorithm, which combines node energy and node degree to form a load balance tree for path selection according to the path performance evaluation factor. The formation of the balance tree effectively balances the node load and greatly improves the node energy consumption [9]. Xiao et al. proposed the mrpropost algorithm, which gets the f-list of frequent 1-itemsets after the first MapReduce task is

executed and constructs the PPC tree to mine the frequent itemsets of multiple computing nodes distributed on it. This process does not need to save the PPC tree in memory, which can not only quickly calculate the itemset support but also reduce the time and space consumption of the algorithm [10].

In this study, a retarget-based hypergraph analysis based on a three-dimensional matrix model is used for project data analysis. The dataset measures the performance and feasibility of the model and the data mining algorithm according to it.

3. Research Methods

3.1. Hypergraph. The Knowledgebase is a knowledge base that uses semantic research to gather data from multiple sources to improve research efficiency. Knowing Atlas is an art form that contains many objects and elements in the real world and their relationships and is used to represent all objects and their relationships in the real world [11]. As shown in Table 1, the knowledge map can be divided into layer structure and data layer of the logical architecture phase.

Although the knowledge map is widely used, the representation methods based on triples often oversimplify the complexity of the data stored in the knowledge map; especially, for hyper-relational data connecting two or more entities, the loss of high-order structure information will lead to the limitation of knowledge hypergraph representation and reasoning ability. Relevant work has proved that, in Freebase, more than 33.3% of entities and 61% of relationships cannot be represented by binary relationships. A knowledge hypergraph is a special kind of heterogeneous graph. In order to understand the characteristics of a knowledge hypergraph more clearly, we first study the representation of a heterogeneous hypergraph. According to its relevance to knowledge hypergraph, the representation method of knowledge hypergraph is further studied. Finally, a three-tier architecture of the knowledge hypergraph is proposed, which can effectively improve the reasoning ability and efficiency of the knowledge hypergraph. The definition, characteristics, and main tasks of hypergraph and correlation graph are shown in Table 2. Where, V refers to the number of node types and E refers to the number of relationship types.

3.2. Redundancy Rule Detection. Let the association rules $R_1: X \rightarrow Y$ and $R_2: A \rightarrow B$. If $(A \cup B) \subseteq (X \cup Y)$ and $X \subseteq A$ are satisfied, the total number of redundant rules is $(3^{|Y|} - 2^{|Y|} - 1)$, where $|Y|$ is the number of items contained in the itemset Y .

Theorem 1. *The theorem proves that, under the existing evaluation criteria, there will be a large number of redundant rules that can be deleted in the mining association rules, and it theoretically analyzes the total number of redundant rules [12, 13].*

TABLE 1: Logical structure of knowledge map.

Logical structure level	Primary coverage	Example
Mode layer	Data model of knowledge class	Concept and relationship
Data layer	Specific data information	Fact triplet

TABLE 2: Definition of hypergraph and related graph.

Graph classification	Definition	Characteristic	Task
Isomorphic graph	$H = (V, E)$, where V represents the node set and E represents the relationship set	$ TV = 1, TE = 1$, its node and relationship type are single, which can only represent binary relationship data	Link prediction, node classification, and clustering
Heterogeneous graph	$H = (V, E)$, where V represents the node set and E represents the relationship set	$ TV + TE > 2$, which has more than one node and relationship, and can only represent binary relationship data	Link prediction, node classification, and clustering
Heterogeneous hypergraph	$H = (V, E)$, where V represents the node set and E represents the relationship set; $e \in E, e(v_1, v_2, \dots, v_n) \subseteq V$ for any superedge	$ TV > 1$, its node types are diverse, which can represent hyper-relational data, and the relationship usually has no semantics	Link prediction, node classification, and clustering
Knowledge hypergraph	$H = (V, E)$, where V represents the entity set and E represents the hyper-relational set; the basic unit is the hyper-relational fact $\text{facta} = (e, 1, 2, \dots, n)$	$ TV > 1, TE > 1$, with various types of entities and relationships; it can represent hyper-relational data and rich relational semantics	Link prediction and hyper-relational fact judgment

Definition 1. (association rule redundancy). Redundant rules can generally be divided into two forms: one is dependent rules; that is, if the conclusion of rule X_i is the same as that of rule X_j and while the premise of X_i is the sufficient condition of premise X_j , then X_i is redundant, and repeated rules can be regarded as a special case of dependent rules [14]. The second is the repeated path rule. If there are selectors X_i and X_j in the rule base and there are at least two paths between X_i and X_j , it can be determined that there are redundant rules.

Dependent rules can be represented by rules (1) and (2):

$$X_2 \longrightarrow X_4, \quad (1)$$

$$X_2 X_3 \longrightarrow X_4. \quad (2)$$

It can be seen from rules (1) and (2) that the subsequent items of the two rules are the same, and the previous item has an intersection, so we think that rule (2) is a redundant rule; then, delete rule (2) and retain rule (1); that is, retain the party with fewer children in the previous item, in which rules (1) and (2) become dependent rules.

Repeated path rules can be represented by rules (3) and (4):

$$X_1 \longrightarrow X_2 X_3 \longrightarrow X_4, \quad (3)$$

$$X_1 \longrightarrow X_5 \longrightarrow X_4. \quad (4)$$

According to rules (3) and (4), there are two paths from X_1 to X_4 . We think that the path is repeated, and delete one of them.

3.3. Directed Hypergraph Representation of Association Rules.

In a directed hypergraph, the directed hyperedge $e \in E$ is defined as an ordered pair composed of head node $H(e)$ and tail node $T(e)$, and $H(e)$ and $T(e)$ are subsets of vertex set V ; that is, it can be composed of a set of multiple vertices. This feature is conducive to the representation of association rules as a directed hypergraph [15]. According to the correspondence between the head node $H(e)$ and the subsequent term of the association rule and the tail node $T(e)$ and the previous term of the association rule, each association rule can be uniquely represented as a super edge in the directed hypergraph.

The form of association rules obtained in practice is $x_1 x_2 \dots x_m \Rightarrow y_1 y_2 \dots y_n$; that is, the first item $\text{Ante}(R)$ is a set composed of multiple items, and the second item $\text{Cons}(R)$ also contains multiple items. We define the rule that the latter item contains only one item as a simple rule and the rule that the latter item contains multiple items as a composite rule. This project defines a directed super edge to represent an association rule. The front term of each association rule corresponds to the head node of a directed hypergraph, and the rear term of the association rule corresponds to the tail node of the same directed hypergraph. There are multiple head nodes and tail nodes for each directed super edge, so the composite rule is successfully represented.

This study adopts a spanning tree-based classification method to remove association rule redundancy. This is a new redundancy check method for association rules, which can effectively check the redundant rules, subordinate rules, and duplicate path rules. Since the adjacency matrix of the directed hypergraph is mainly used in simple graphs and the directed hypergraph we want to use here has composite points, which

makes the composite rules only represented by the directed hypergraph, the adjacency matrix must be redefined. A spatial database is integrated with spatial relational data and object-relational data to realize a database of spatial data. The generation process of a spatial database includes the logical structure design of the database and the integrated storage of spatial data. Among them, the logical structure design of the database uses the classic E-R (entity connection) diagram to describe the real geographical world, and the number of paths between layers is proportional to the number of data attribute features. The specific design is shown in Figure 1.

3.4. Graphic Representation and Processing of Redundancy Rules. The adjacency matrix of a hypergraph expression completely defines the relationships of the vertices in figures. The adjacency matrix of an expressed hypergraph based on organizational rules describes the interrelationships of the objects of organizational policies. Retrieval can be accomplished according to the information hypergraphs according to the definition of redundancy rules in Definition 1 and related items of the diagram.

Definition 2. (road). Let graph $G = (V(G), E(G))$; its path is a finite nonempty sequence $W = v_0 e_1 v_1 e_2 \dots e_k v_k$, that is, the alternating sequence of vertices and edges, where $e_i \in E(G)$ and $v_j \in V(G)$, where e_i is associated with v_{i-1} and v_i , respectively, $1 \leq i \leq k, 0 \leq j \leq k$, which is recorded as (v_0, v_k) path, called vertex v_0, v_k are the starting point and end point of path W , respectively, v_1, v_2, \dots, v_{k-1} is the inner vertex of path W , and k is the length of W . If e_1, e_2, \dots, e_k in path W is different from each other, it is called trace. If $v_0, v_1, v_2, \dots, v_{k-1}, v_k$ in trace W is different from each other, it is called a path. If the starting point and ending points of a path (trace and road) are the same, it is called a closed path (closed trace and closed circuit). Closed trace is also called circle [16].

From the definition in graph theory, we found that, to realize the processing of redundant rules in a directed hypergraph based on association rules, it can be transformed into discovering connected blocks in the hypergraph and transforming it into a spanning tree [17]. Because each edge in the directed hypergraph represents an association rule and when the connected graph becomes a spanning tree, the edge needs to be deleted, this edge is the redundant rule in the association rule. Reduction of redundant rules:

- (1) If $E_i \cap E_j \neq \emptyset$, called the associated super edge, then there is the following formula:

$$|E_i \cap E_j| = 1. \quad (5)$$

- (2) If condition (1) is true and $|E_i| \neq 2$, there must be the following formula:

$$|E_j| \neq 2. \quad (6)$$

Then, hypergraph $H = (X, E)$ has a spanning hyper tree T .

Lemma 1. Let the hypergraph H have n vertices, m hyperedges, and c connected branches. If and only if the following equation $\sum_{i=1}^m (|e_i| - 1) = n - c$ exists, H does not contain a superloop [18].

The bipartite graph corresponding to hypergraph $H = (V, E)$ refers to the vertex set V and hyperedge set E of h as vertex sets. When $v_i \in e_i$ in H , vertex v_i and vertex e_i in the bipartite graph are edge connected. The bipartite graph corresponding to H is represented by $G \langle H \rangle$. Figures 2(a) and 2(b) show a hypergraph and its corresponding bipartite graph.

We get the number of connected blocks contained in the directed hypergraph and the location of the connected blocks where each point is located. On this basis, we must perform spanning tree processing on each connected block. Delete redundant rules by obtaining the spanning tree.

3.4.1. Algorithm to Get Spanning Tree. At present, there are generally two methods to find the spanning tree of a connected graph: the ring-breaking method and the ring-avoiding method. The so-called loop breaking method is to break all loops in a connected graph, and the remaining connected graph without loops is a spanning tree of the original graph. This algorithm is called the "loop breaking method." Take an arbitrary edge e_1 in graph G , find an edge e_2 that does not form a loop with e_1 , and then find an edge e_3 that does not form a loop with $\{e_2, e_3\}$. This continues until the process cannot be carried out. At this time, the obtained graph G is a spanning tree. This algorithm is called the "circle avoiding method." According to the meaning of the hypergraph we generated, each hyperedge represents an association rule. So, obviously, we should use the broken circle method.

Input the adjacency matrix of the connected block and get an adjacency matrix of the spanning tree. By restoring the adjacency matrix of the spanning tree, we can eliminate the redundancy of the association rules. In practice, we find that there is often more than one spanning tree of a connected graph. At the same time, this will present a problem; that is, in the obtained rules, the rules that people are interested in and think are important may be deleted. In order to solve this problem, we give a certain weight to the more important association rules. Reflected in the graph is to give weight to each edge. Combined with the current algorithm, we give a smaller weight to the edge corresponding to the more important association rules and a larger weight to the unimportant and uninterested association rules and then use the prim algorithm to calculate the minimum spanning tree.

Definition 3. (minimum spanning tree). In figure $H = (V, E)$, (u, v) represents the edge-connecting vertex u and vertex v , that is, $(u, v) \in E$, and $W(u, v)$ represents the weight of this edge. If there is a subset of T as E , that is, $T \subseteq E$ and acyclic graph, making $W(T)$ of $W(T) = \sum_{(u,v) \in T} W(u, v)$ minimum, this T is called the

minimum spanning tree of H . The minimum spanning tree is actually an abbreviation of the minimum weight spanning tree.

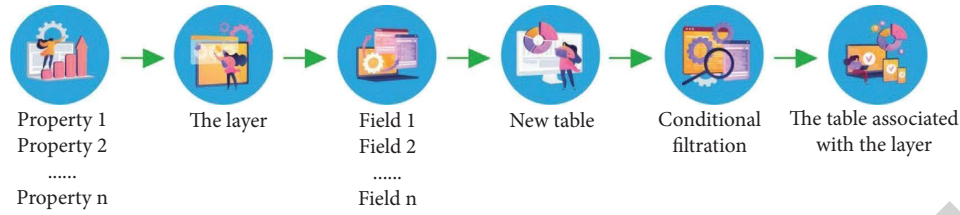


FIGURE 1: E-R model between layers.

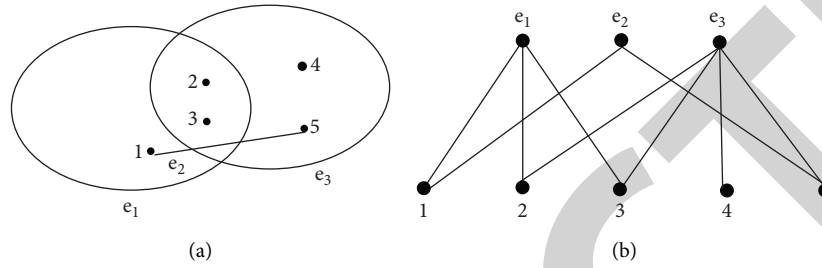


FIGURE 2: Hypergraph and its corresponding bipartite graph. (a) Hypergraph. (b) Divided into two pictures.

3.4.2. *Basic Idea of Prim Algorithm.* Starting from a vertex of the connected graph $H = (V, E)$, select the edge (u_0, v) with the smallest weight associated with it and add its vertices to the vertex set U of the spanning tree. In each subsequent step, select the edge (u, v) with the smallest weight from the edges where one vertex is in U and the other vertex is not in U and add its vertices to the set U . In this way, until all vertices in the graph are added to the vertex set u of the spanning tree, a minimum spanning tree is obtained.

Through practice, it has been found that the edge set of the minimum spanning tree is sometimes different. We introduce weight to deal with the minimum spanning tree, which plays a corresponding protective role in the preservation of important and interesting rules.

3.5. *Algorithm Flow.* Figure 3 shows the outline of the tree spanning plan in this study to eliminate the recurrence of the organizational policy. The special steps are as follows.

- (1) Analyze test data, create aggregation rules, use hypergraphs to represent participatory rules, and revise and obtain its integers
- (2) The preprocessed adjacency matrix is obtained by subtracting the algorithm [16, 19]
- (3) Unspanned and linked spanning trees are obtained by the spanning tree algorithm
- (4) The adjacency matrix of the surrounding tree is reconstructed by the organizational law, and its final completion is possible

This study proposes an algorithm to remove subordinate rules by redefining the adjacency matrix. Each association rule is defined as an edge of a directed hypergraph. According to the previous section, the redefining adjacency matrix is obtained. The columns of the matrix represent the subsequent terms of the association rule. The flowchart of

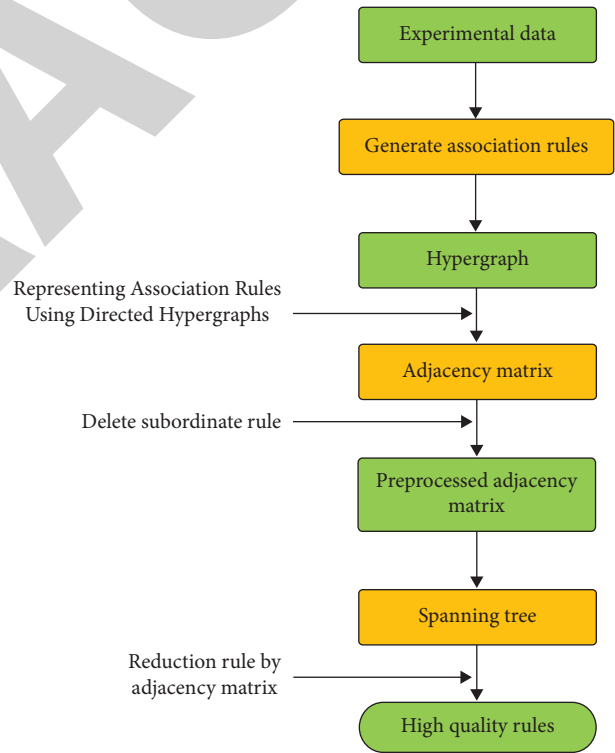


FIGURE 3: Flowchart of association rule redundancy detection.

the algorithm is shown in Figure 4. After this algorithm, all the subordinate rules in the redundant rules can be deleted, and the preprocessed adjacency matrix can be obtained.

4. Result Analysis

4.1. *Verification Results.* In this study, the spanning tree-based classification method to eliminate the redundancy of connection policies consists of three modules: the redefine adjacency matrix module, the delete dependency rule

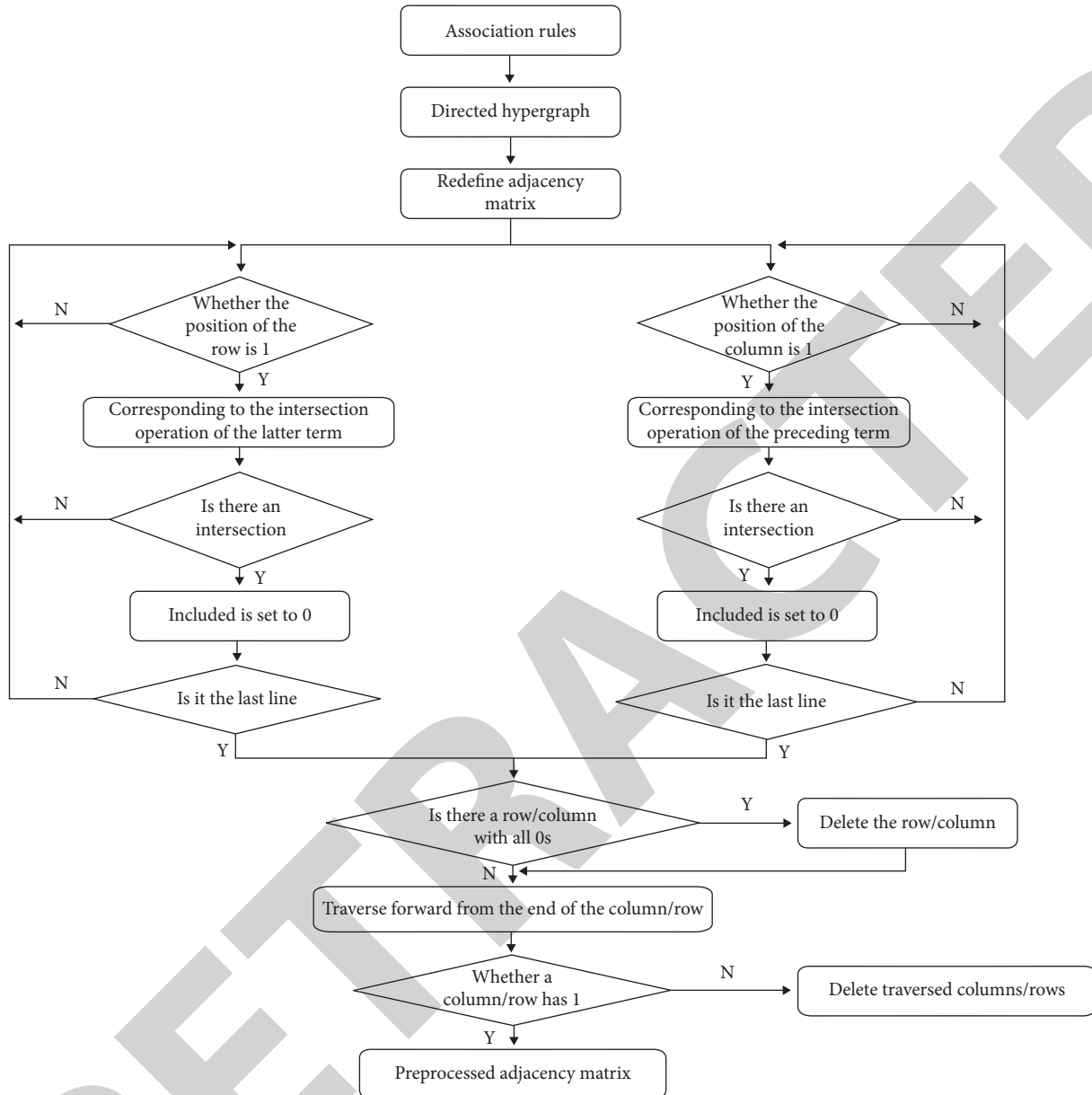


FIGURE 4: Flowchart of removing dependent rules.

module, and the spanning tree module. The first two modules are used in VB programming, while the spanning tree module is used in MATLAB. The special points are shown in Figure 5.

After removing the coding algorithm and the spanning tree algorithm, we obtain a tree spanning without connection. The special procedure of the spanning tree algorithm and hypergraph instructions is shown in Figure 6. It can be seen from Figure 6 that the preprocessed adjacency matrix (the result of the dependency removal algorithm) yields the total deleted adjacency matrix. The redundant, on the right-hand side, shows the variation of the hypergraph indicator before and after the spanning tree algorithm [17, 20].

Through the method introduced in this study, the redundant rules are removed accurately and quickly in both datasets. The specific results are shown in Table 3.

4.2. Experiment. The experimental data studied in this study came from the data obtained from the special task project of Humanities and Social Sciences Research of the Ministry of Education, “Research on building a scientific and complete network culture construction and management system of colleges and universities,” the special project of moral education innovation and development of a city, “analysis of the influence factors and validity of social environment on young students,” and the project of moral education innovation and development

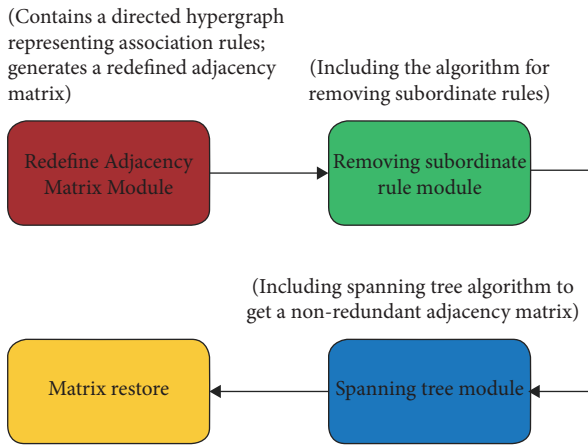


FIGURE 5: Module composition of classification redundancy removal method.

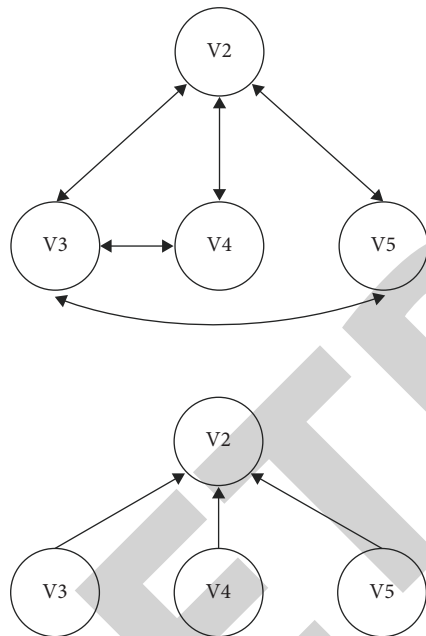


FIGURE 6: Matlab simulation and its directed hypergraph.

of a city, “large-scale special research on contemporary universities in the Internet environment.” The purpose of this project is to understand whether the current Internet environment impacts contemporary college students’ life, learning, ideology, especially their outlooks on life, world outlook, and values, and strive to determine the major influencing factors that affect young students, to provide decision support for constructing network culture in colleges and universities [21]. The data consisted of 63 questions and 30143 valid records. The specific questionnaires include the following: A1~A7 are basic personal information, T1~T23 involve college students’ habits of surfing the Internet (online time, online purpose, habits of going to social networking sites, views on hot online events, etc.), and T24~T29 are college students’ political attitude and learning attitude. We have preprocessed this data, established a three-dimensional matrix model, and done a lot of data analysis. Some examples of data analysis are listed below.

4.2.1. *Basic Statistical Analysis (Sample)*. T1: how long do you spend online every day?

As shown in Figure 7, there are 29933 cases of effective data in this part, and 0.7% of this item is missing. For the problem of young students’ online time, 3608 students (11.97%) spend less than 1 hour online every day, 8986 students (29.81%) spend between 1-2 hours online every day, 11735 students (38.93%) spend between 2-4 hours online every day, 4370 students (14.5%) spend between 4-8 hours online every day, and 1234 students (4.09%) spend more than 8 hours online every day.

According to the test statistics, $\chi^2 = 12177.003$, $p = 0 < 0.05$, reaching a significant level, indicating that there is a significant difference in the number of times the five options in “daily online time” are checked by the sample.

4.2.2. *Crosstab Analysis (Example)*. This case studies whether there is a significant difference in the percentage of choices the five options in “daily online time” among young students of different genders. The statistical results are shown in Figure 8. According to the Chi-square test statistics, the Pearson Chi-square value is 2.368, the degree of freedom is 3, and the significance probability value $p = 0 < 0.05$, reached a significant level, indicating that there is a significant difference between the percentages of at least one choice time of young students of different genders in the five T1 options.

4.2.3. *Logistic Regression Analysis (Sample)*. In this case, t26.1 “contemporary young students should take realizing the great rejuvenation of the Chinese nation as their own responsibility” as the dependent variable, A1, A4, A5, A6, A7, and T4 “browsing the content of interest online” as the independent variables for logistic regression analysis.

First, we test the likelihood ratio of each independent variable. If $p = 0 < 0.05$, it means that the independent variable has statistical significance for the corresponding variable. From Table 4, we can see that A7 is “whether it is the only child,” and the p value of D option “science and technology trends” and I option “job hunting and employment” in T4 is greater than 0.05, indicating that A7, T4.D science and technology trends, and T4.I job hunting and employment have no influence on whether young students agree to “realize the great rejuvenation of the Chinese nation as their own responsibility.”

4.2.4. *Association Rule Analysis (Example)*. In this example, the basic information (A1~A7) and T1~T23 are used as the antecedents of association rules, and T24~T29 are used as the antecedents of association rules for association rule analysis. We set the support of association rules to 70% and the confidence to 90%, expecting to get high-quality rules and reduce the number of rules. Some rules we obtained are shown in Table 5.

We take the basic information in A1~A7 and T1~T29 in the questionnaire as the antecedents of association rules and T1~T29 as the antecedents of association rules for

TABLE 3: Results of removing redundant rules.

	Balloons	Shuttle landing control
Total association rules	18	15
Subordinate rule	8	9
Duplicate path rule	7	2
Number of redundant rules removed	15	11
Number of remaining association rules	3	4

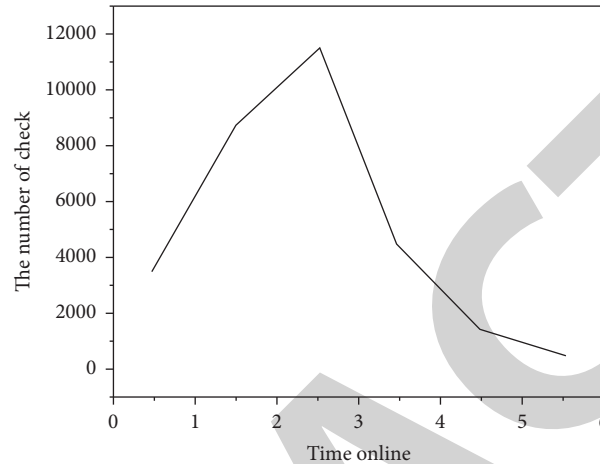


FIGURE 7: Sample of basic statistical analysis.

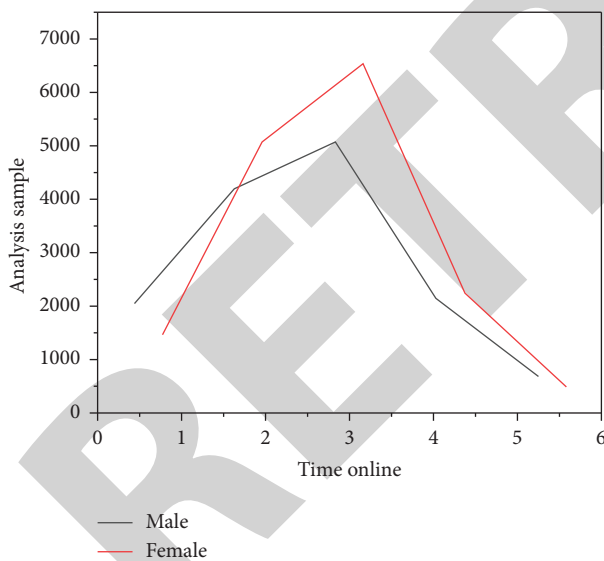


FIGURE 8: Example of Crosstab analysis.

association rule analysis. In order to get higher quality rules and reduce the number of redundant and loop rules as much as possible, because there are many rules, we set the support and confidence as high as possible. We set the support at 75% and the confidence at 85%. Taking some association rules in the results as examples, the specific explanations are as follows:

(1) Young students who agree that “China must develop a low-carbon, green economy and take the path of sustainable development” and agree that “honesty,

trustworthiness and doing what one says is the bottom line that everyone should abide by” the main topic of chatting with people online does not love “(support 78.436% and confidence 89.599%). (2) Young students who agree that “filial piety to parents and respect for teachers are natural” and agree that “honesty and trustworthiness and doing what one says are the bottom lines that everyone should abide by” will also agree with the view that “tolerance is a virtue” (82.258% support and 97.1% confidence).

Through the interpretation of the mined association rules, we can find the important relationships and laws contained therein to provide a practical basis for enriching the theoretical research of young students’ education and promoting the cultural education of young students.

It can be seen from Figure 9 and Table 6 that although the running time of the coevolutionary algorithm is slightly longer than that of the other two global optimization algorithms, the running time is completely within an acceptable range. Moreover, due to the effective introduction of the idea of coevolution, compared with the use of the other two algorithms for association rule mining, it not only has better mining quality but also has a significant advantage in the ability to jump out of the local optimal solution, realizing the search of high-quality association rules in high-dimensional datasets.

After obtaining the association rules, we use the algorithm introduced above for redundancy processing. The experimental results are shown in Table 7.

As shown in Table 7, it can be seen that the use of hypergraph-based redundancy and loop detection methods can reduce the release of redundancy and loop policy,

TABLE 4: Example of logistic regression analysis.

Influence factor	<i>p</i> value
A1 gender	0
A4 nationality	0
A5 political outlook	0
A6 home location	0
A7 is it an only child	0.628
T4.A current affairs and politics	0
T4.B economic finance	0
T4.C education and training	0
T4.D technology trends	0.146
T4.E game entertainment	0.006
T4.F sports events	0.002
T4.G emotional marriage	0.007
T4.H culture and art	0.016
T4.I job hunting and employment	0.171

TABLE 5: Example of association rule analysis.

Consequent	The aforesaid	Support (%)	Confidence (%)
T28.4 = agree	T20 talk about news or interesting things that happen around you = Y	70.899	91.371
T28.3 = agree	T20 talk about news or interesting things that happen around you = Y	70.899	91.353
T28.3 = agree	T11 classmates and friends around = Y	70.258	90.452
T28.4 = agree	T11 classmates and friends around = Y	70.258	90.339
T28.3 = agree	T15 seek visual stimulation = N and T23 pretend to be someone else and make trouble = N and nationality = han nationality	71.721	90.222
T28.4 = agree	T15 seek visual stimulation = N and T23 pretend to be someone else and make trouble = N and nationality = han nationality	71.721	90.148
T28.3 = agree	T15 seek visual stimulation = N and nationality = han nationality and T12 I have experienced it, but I will never believe it again = N	75.948	90.076
T28.4 = agree	T23 pretend to be someone else and make trouble = N and nationality = han nationality and T12 I have experienced it, but I will never believe it again = N	77.142	90.074
T28.3 = agree	T23 pretend to be someone else and make trouble = N and nationality = han nationality and T12 I have experienced it, but I will never believe it again = N	77.142	90.049
T28.4 = agree	Political outlook = masses and T12 I have experienced it, but I will never believe it again = N and T12 I hope online love can come to me = N	70.235	90.019
T28.4 = agree	T15 seek visual stimulation = N and nationality = han nationality and T12 I have experienced it, but I will never believe it again = N	75.948	90.01

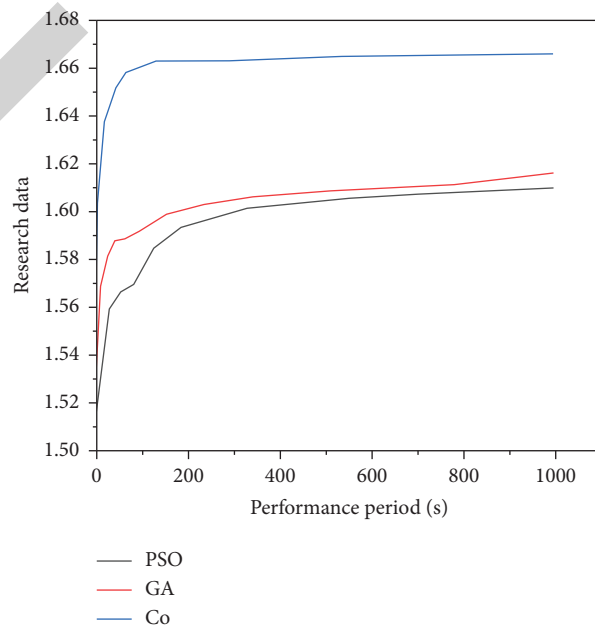


FIGURE 9: Comparison diagram of Co, GA, and PSO applied to research database.

TABLE 6: Running time of Co, GA, and PSO applied to the research database.

Algorithm	Running time (s)
Co	5946.13
GA	4601.43
PSO	3153.56

TABLE 7: Test results of dataset.

Dataset	Number of rules to be detected	Number of redundant rules	Number of rules after processing
Set A	23	6	17
Set B	7	1	6
Set C	98	34	64
Set D	9	2	7

without wasting the capital mining, and facilitate user selection and implementation.

5. Conclusion

In the context of big data, the task of exploring the organization's policy redundancy and improving the efficiency and effectiveness of the organization's policy mining is more pressing, and it has been increased to become the gold research and a key tool in the organization of the mining law industry. In this study, a wood spanning partition method is adopted to eliminate the redundancy of organizational policies. By redefining the adjacency matrix and its directed hypergraph, it uses the adjacency matrix to reflect the relationship between the association rule items to be detected and uses the adjacency matrix to find the connected blocks. By running the program, it obtains the number of connected blocks contained in the directed hypergraph and the location of the connected blocks where each point is located. On this basis, the adjacency matrix of the spanning tree is obtained using the ring breaking method of the spanning tree of the connected graph, and a certain weight is given to avoid the possibility of deleting important rules, to obtain the minimum spanning tree. In this way, we can remove redundant rules by checking dependent rules and repeated path rules.

Data Availability

The labeled dataset used to support the findings of this study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by Project of Jilin Provincial Department of Science and Technology (20180201129GX).

References

- [1] Y. Wen, J. Liu, W. Dou, X. Xu, B. Cao, and J. Chen, "Scheduling workflows with privacy protection constraints for big data applications on cloud," *Future Generation Computer Systems*, vol. 108, pp. 1084–1091, 2020.
- [2] H. Xiao, S. Ali, Z. Zhang, M. S. Sarfraz, F. Zhang, and M. Faisal, "Big data, extracting insights, comprehension, and analytics in cardiology: an overview," *Journal of Healthcare Engineering*, vol. 2021, no. 7541, pp. 1–14, 2021.
- [3] X. Lv and M. Li, "Application and research of the intelligent management system based on internet of things technology in the era of big data," *Mobile Information Systems*, vol. 2021, no. 16, pp. 1–6, 2021.
- [4] B. Khan, R. Naseem, M. A. Shah et al., "Software defect prediction for healthcare big data: an empirical evaluation of machine learning techniques," *Journal of Healthcare Engineering*, vol. 2021, no. 2, pp. 1–16, 2021.
- [5] R. Skuratovskii, A. Bazarna, and E. Osadhyy, "Analysis of speech mel scale and its classification as big data by parameterized knn," *Artificial Intelligence*, vol. 26, no. 1, pp. 42–57, 2021.
- [6] L. Baert, E. Chèrière, C. Sainvitu, I. Lepot, A. Nouvellon, and V. Leonardon, "Aerodynamic optimization of the low-pressure turbine module: exploiting surrogate models in a high-dimensional design space," *Journal of Turbomachinery*, vol. 142, no. 3, pp. 031005–031005.11, 2020.
- [7] S. S. Shekhawat, H. Sharma, S. Kumar, A. Nayyar, and B. Qureshi, "Bssa: binary salp swarm algorithm with hybrid data transformation for feature selection," *IEEE Access*, vol. 9, no. 99, pp. 14867–14882, 2021.
- [8] D. A. Elmanakhly, M. M. Saleh, and E. A. Rashed, "An improved equilibrium optimizer algorithm for features selection: methods and analysis," *IEEE Access*, vol. 9, pp. 120309–120327, 2021.
- [9] S. Liu and K. Zhang, "Under-sampling and feature selection algorithms for s2smlp," *IEEE Access*, vol. 8, pp. 191803–191814, 2020.
- [10] Y. Xiao, X. Li, B. Liu et al., "Multi-view support vector ordinal regression with data uncertainty," *Information Sciences*, vol. 589, pp. 516–530, 2022.
- [11] M. Lin, X. Chen, Z. Wang, D. Wang, and J. L. Zhang, "Global profiling and identification of bile acids by multi-dimensional data mining to reveal a way of eliminating abnormal bile acids," *Analytica Chimica Acta*, vol. 1132, pp. 74–82, 2020.

- [12] Y. Wu, Q. Zhang, Y. Hu et al., "Novel binary logistic regression model based on feature transformation of xgboost for type 2 diabetes mellitus prediction in healthcare systems," *Future Generation Computer Systems*, vol. 129, pp. 1–12, 2022.
- [13] G. Zhou, X. Bao, S. Ye, H. Wang, and H. Yan, "Selection of optimal building facade texture images from uav-based multiple oblique image flows," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 2, pp. 1534–1552, 2021.
- [14] A. E. Radwan, "Three-dimensional gas property geological modeling and simulation," *Sustainable Geoscience for Natural Gas Subsurface Systems*, vol. 63, pp. 29–49, 2022.
- [15] Z. Guo and Z. Xiao, "Research on online calibration of lidar and camera for intelligent connected vehicles based on depth-edge matching," *Nonlinear Engineering*, vol. 10, no. 1, pp. 469–476, 2021.
- [16] R. Huang, "Framework for a smart adult education environment," *World Transactions on Engineering and Technology Education*, vol. 13, no. 4, pp. 637–641, 2015.
- [17] A. Jc, B. JI, L. B. Xin, A. Wg, Z. Jing, and C. Fza, "Degradation of toluene in surface dielectric barrier discharge (SDBD) reactor with mesh electrode: synergistic effect of UV and TiO₂ deposited on electrode," *Chemosphere*, vol. 288, 2021.
- [18] S. Shriram, B. Nagaraj, J. Jaya, S. Shankar, and P. Ajay, "Deep learning-based real-time AI virtual mouse system using computer vision to avoid COVID-19 spread," *Journal of Healthcare Engineering*, vol. 2021, Article ID 8133076, 8 pages, 2021.
- [19] R. Kumar and A. Sharma, "Risk-energy aware service level agreement assessment for computing quickest path in computer networks," *International Journal of Reliability and Safety*, vol. 13, no. 1/2, p. 96, 2019.
- [20] J. Jayakumar, B. Nagaraj, S. Chacko, and P. Ajay, "Conceptual implementation of artificial intelligent based E-mobility controller in smart city environment," *Wireless Communications and Mobile Computing*, vol. 20218 pages, 2021.
- [21] A. Sharma and R. Kumar, "A framework for pre-computed multi- constrained quickest QoS path algorithm," *Reliability Modeling and Optimized Planning of Risk-based Resilient Networks*, vol. 45, 2017.