

Research Article

Equipment Quality Information Mining Method Based on Improved Apriori Algorithm

Fulin Li, Chen Meng , Cheng Wang, and Shuyi Fan

Shijiazhuang Campus, Army Engineering University, Shijiazhuang 050003, China

Correspondence should be addressed to Chen Meng; mengchenlgd@163.com

Received 18 August 2022; Revised 21 March 2023; Accepted 18 April 2023; Published 3 May 2023

Academic Editor: Davide Palumbo

Copyright © 2023 Fulin Li et al. 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.

Equipment quality-related data contains valuable information. Data mining technology seems to be an efficient method for extracting knowledge from large amounts of data. In this paper, a general method for equipment quality information mining based on association rule is proposed for complex equipment. Due to the shortcomings of classical association rule mining algorithms such as long running time and high memory consumption, the candidate itemset generation process is optimized, and an improved Apriori algorithm is proposed. Taking five experimental data sets as the object, the performance of the algorithms is tested using time complexity and spatial complexity as evaluation criteria. Comparative experiments show that the improved algorithm had advantages. To further implement data processing and information representation, a matrix-based strong association rule extraction algorithm was proposed. Taking a certain type of equipment as an example, a simulation experiment was conducted using the method proposed in this article in reliability test data sets, and some interesting knowledge was obtained through mining, verifying the effectiveness of the method. The research in this article seems promising with respect to improving the scientific level of equipment support.

1. Introduction

Equipment quality information is the information related to the quality of equipment in the stages of development, use, and maintenance. The application of big data and the development of information technology have greatly increased the quality-related data throughout the entire life cycle of equipment. It is of great significance to mine massive data; extract information that reflects equipment quality requirements, status, changes, and related factors and their relationships; and serve the use, maintenance, and research of equipment.

Association rule was originally used to solve the problem of shopping basket analysis, with the purpose of discovering customers' buying habits. Over the years, many experts and scholars have carried out extensive research, including the theoretical exploration, application and promotion of association rule mining, and the optimization and improvement of mining algorithm. Association rule has played an important role in scientific research, communication and security, finance, retail, medical care, and other industries [1–5]. In

this paper, association rule will be used to mine equipment quality information.

The classical algorithms of association rule mining include Apriori algorithm [6] and FP-growth algorithm [7]. After decades of development, various solution strategies for association rule mining problems emerge in endlessly. For example, meta-heuristic-based approaches include ant colony optimization, artificial neural network, differential evolution, genetic algorithm, and particle swarm optimization [8–12]. Taking differential evolution as an example, Altay and Alatas proposed hybrid optimization and global search methods based on differential evolution and sine cosine algorithm, which can automatically adjust the appropriate interval of numerical valued attributes and mine without finding frequent itemsets, thus realizing the mining of numerical association rules. This method has strong adaptability and high automation level. For the evolutionary optimization methods, a new representation scheme of evolutionary computation based on chaos number is also proposed and applied to quantitative association rule mining [13].

Because of the simple idea and low implementation difficulty, classical Apriori algorithm is still an enduring association rule mining algorithm, which is widely used in various fields. However, after long-term and in-depth research, experts and scholars have found the shortcomings of Apriori algorithm: (1) there are many redundant steps, and the candidate itemset generated during the operation may be too large. (2) It has a greater number of scans on the database and larger I/O load. When the amount of data is huge, the problems above will lead to a significant decrease in the efficiency of the algorithm. Many efforts have been made to improve the mining algorithm, mainly including the following ways:

- (1) Improvement based on the Apriori algorithm itself: Arcos and Hernandez denoted the constraint of the classical Apriori algorithm in terms of computational cost and accomplished the improvement by minimizing the algorithm generation cost through enhancing transaction reduction approach [14]. Wang and Zheng analyzed the working principle of the traditional Apriori algorithm and pointed out the existing problems. They proposed an improved Apriori algorithm for time series of frequent itemsets and applied this improved time series Apriori algorithm for frequent itemsets to mining association rules based on time constraints [15]. Chiclana et al. proposed a new mining algorithm based on animal migration optimization, greatly reduced the computational time for frequent itemset generation, memory for association rule generation, and the number of rules generated [16]. Such improvements are intended to reduce the number of scans and thus improve efficiency
- (2) Improvement of storage mode: Mar and Oo used linked list and hash table, which makes the processing time and memory space that Apriori algorithm needs reduce [17]. In order to optimize the connection step, pruning step, support counting step, and transaction storage mode of the Apriori algorithm, Sun and Li applied the prefixed-itemset and the compression matrix [18]. Xu improved the mining performance of the Apriori algorithm by introducing dynamic storage space [19]
- (3) Improvement of mining algorithm for specific industry data: some industry-specific data, such as medical big data, have obvious industry data characteristics and attributes. Using these characteristics can improve the mining algorithm of medical data. Sornalakshmi et al. presented an improved Apriori algorithm names enhanced parallel and distributed Apriori (EPDA) for the healthcare industry, which works in terms of the time and number of rules generated with a database of healthcare and different minimum support [20]. In the research of medical big data, He et al. optimized the Apriori algorithm by Amazon Web Services and graphics

processing unit, and the data mining speed was improved [21]

- (4) Improvement based on platform and programming model: due to the heavy I/O load of the traditional Apriori algorithm and the problem of limited storage resources, some scholars have studied the use of Hadoop platform and MapReduce model to realize parallel calculation, so as to improve the original method [22, 23]. Spark is an extension of MapReduce programming framework. When processing large-scale data sets and data iterations, its performance is better than MapReduce. Singh et al. proposed MapReduce-based frequent itemset mining algorithms on Hadoop cluster and investigated the efficiency of various data structures for the Spark-based Apriori algorithm [24]
- (5) Comprehensive use of various improvement methods: to improve Apriori algorithm, Ye proposed a method based on compressed combination itemset technology and Hash technology. The improved algorithm was applied in the library personalized service [25]. Yuan not only optimized the Apriori algorithm itself but also changed the data mapping mode, effectively improving the operation efficiency [26]

Aiming at the problems of classical Apriori algorithm, this paper will improve the algorithm itself. Apriori algorithm with optimized candidate itemset generation strategy is proposed. The improved mining method is used to search frequent itemsets, and then, strong association rules required by users are extracted.

2. Methods

In association rule theory, the indivisible minimum unit of information in the database is called item, which is represented by the symbol i . The set of items is called itemset, which is expressed as $I = \{i_1, i_2, \dots, i_k\}$. I is called k -itemset, and k is equal to the number of items contained in the itemset. Transaction refers to the set of items contained in one process, which is represented by the symbol t . $T = \{t_1, t_2, \dots, t_n\}$ represents the set of transactions. If I is a collection of all items, the itemset that each transaction in T contains is all subsets of I . The number of transactions that contain a certain itemset in all transactions is called the support count of that itemset, that is, the frequency of occurrence of that itemset.

2.1. Frequent Itemset Mining Algorithm. The meaning of association rule is that when some items appear in the same transaction, other items also appear, which is usually expressed in the form of $X \implies Y$. Both X and Y are itemsets, and both of them are true subsets of I , $X \cap Y = \emptyset$. The support of association rule is the ratio of the number of transactions containing both X and Y to the total number of transactions, expressed in $support(X \implies Y)$, $support(X \implies Y) = support(X \cup Y) = P(XY)$. The confidence of association rule is the ratio of the number of transactions containing both X and Y

to the number of transactions containing X , which is expressed by $confidence(X \implies Y)$. The calculation method is as follows:

$$confidence(X \implies Y) = \frac{support(X \cup Y)}{support(X)} = P(Y|X). \quad (1)$$

In fact, the support is equivalent to the occurrence probability of the items contained in X and Y in all transactions, and the confidence is equivalent to the probability of Y occurring under the condition that X occurs.

Minimum support specifies the minimum importance that association rules must meet, and minimum confidence specifies the minimum reliability that association rules must meet. They are two thresholds set by the user as required. In this paper, they are expressed by min_sup and min_conf . One itemset will be considered frequent itemset if its support is not less than min_sup . One association rule will be considered as strong association rule if its support is not less than min_sup and the confidence is not less than min_conf .

It is undeniable that association rules are deceptive, which is reflected in the fact that the meaning of some strong association rules may not be consistent with objective facts. In order to prevent this kind of situation, this paper introduces the lifting degree as the correlation measure between itemsets, which is expressed by $lift(X, Y)$. The calculation method is as follows:

$$lift(X, Y) = \frac{support(X \cup Y)}{support(X)support(Y)} = \frac{P(X \cup Y)}{P(X)P(Y)}. \quad (2)$$

$lift(X, Y) > 1$ means that X and Y are positively correlated. At this time, rule $X \implies Y$ should be interpreted normally. $lift(X, Y) < 1$ means that X and Y are negatively correlated. At this time, rule $X \implies Y$ cannot be interpreted normally, because the fact is that when the items in X appears, the items in Y will not appear. $lift(X, Y) = 1$ means that X and Y are independent of each other, and there is no correlation, so this rule is meaningless.

There are two steps in association rule mining: the first step is to find all the frequent itemsets, and the second step is to generate strong association rules based on the frequent itemsets. In the first step, the commonly used algorithm is Apriori algorithm, which generates a candidate itemset with a larger number of items from a frequent itemset with a smaller number of items, then filters out the itemsets that meet the minimum support, and obtains the frequent itemset. Apriori algorithm needs a lot of connection and pruning steps in the process of running. Connection is to make all itemsets contained in frequent k -itemsets combine in pairs to produce a union set to obtain a candidate $k + 1$ -itemset. Pruning is to delete infrequent itemsets in candidate itemsets. In order to prevent the amount of computation from greatly increasing due to the large scale of the candidate set, the pruning method of the classical Apriori algorithm is to compress the candidate set according to the Apriori property and then filter the remaining itemsets according to the minimum support.

Classical Apriori algorithm still has some defects: (1) the connection operation adopts the combination method. Although it can ensure that each itemset will not be repeatedly connected with the same itemset, the final candidate itemset may still have redundant items, so necessary comparison and deduplication operations are required, and these steps will waste a lot of time. (2) Whether the itemset generated by the connection meets the conditions or not, it will be added to the candidate set first and then selected according to the Apriori property. The additional running time during this period cannot be ignored. The shortcomings of classical algorithms will become more fatal with the increase of the size of the data set, resulting in a significant reduction in the efficiency of the algorithm.

In order to solve the problems above, the classical Apriori algorithm is improved. The algorithm proposed in this paper continues to use some ideas of the classical algorithm and adopts a new strategy to generate candidate itemsets. Before using frequent k -itemsets to generate candidate $k + 1$ -itemsets, a screening step is added to ensure that the itemsets participating in the connection step meet the following conditions: (1) the items contained in the two itemsets are identical except for the last one. (2) Take any $k - 2$ items from the same items contained in the two itemsets, and combine them with the last item of the two itemsets to engender a k -itemset, and the k -itemset should be included in the frequent k -itemset. If the itemsets that do not meet the above conditions are connected, at least one of the following situations will occur: (1) the generated new itemset is exactly the same as the previously generated itemset; that is, redundant items are generated. (2) The number of items in the generated new itemset is greater than $k + 1$, which is beyond the scope of the current candidate itemset. (3) The generated new itemset does not satisfy Apriori property. No matter which of the situations occurs, additional operations need to be added to eliminate its influence. The improved algorithm proposed in this paper can ensure that the itemsets generated by connection are candidate itemsets and reduce unnecessary connection, comparison, deduplication, and pruning steps.

The steps of the improved algorithm proposed in this paper are as follows.

2.2. Matrix-Based Strong Association Rule Extracting Algorithm. Mining all frequent itemsets from the database according to the minimum support set by users is the main work of association rule mining, and it is also the key research content in this field. The generation of strong association rules based on frequent itemsets is the process of mining the hidden information behind the data and expressing it in a form which is easier to be understood after a lot of data analysis and processing. It is the key link of applying association rule theory to obtain information from data. At present, the extraction of strong association rules has received relatively less attention. Most researchers focus more on the mining of frequent itemsets, and there is little introduction on how to implement the subsequent extraction of strong association rules.

Input: import data set to generate numerical matrix D

Step1 set the minimum support count min_sup

Step2 scan all the data to generate $L1$: the set of frequent 1-itemsets. The main steps are as follows:

Step2.1 create identity matrix E with n rows and n columns, where n is the number of columns of matrix D

Step2.2 calculate the sum of each column of matrix D and store it in matrix B in turn

Step2.3 compare all elements of matrix B with min_sup , if the i -th element of B is less than min_sup , then delete that element and the i -th row of matrix E

Step2.4 connect matrix E and matrix B' in the vertical direction to obtain matrix L . At this time, E is a frequent 1-itemset, and L contains both frequent 1-itemsets and corresponding support of the itemsets

Step3 combines all item sets contained in frequent 1-itemset $L1$ in pairs, connects the newly generated 2-itemsets in the horizontal direction to obtain matrix C , that is, candidate 2-itemset $C2$

Step4 if C is not an empty set, prune C according to min_sup to get the frequent k - itemset Lk . The main steps are as follows:

Step4.1 let m be the number of rows of C , let t be the number of rows of matrix D , and let $i = 1$

Step4.2 create zero matrix N with m rows and 1 column

Step4.3 create an all-one matrix H with t rows and 1 column

Step4.4 find the elements equal to 1 in the i -th row of C , store the subscripts of these elements with matrix ind , and count the number of elements as n

Step4.5 take the value stored in ind as the column subscript, read the corresponding n columns data in matrix D , and perform an and operation with matrix H

Step4.6 find the sum of all elements of matrix H and store it in x , if $x < min_sup$, then delete the i -th row of C and N , $m = m - 1$; Otherwise, let $N(i, 1) = x$, $i = i + 1$

Step4.7 if $i \leq m$, turn to step4.3; Otherwise, connect matrix C and matrix N in the vertical direction to obtain the frequent k -itemset Lk

Step5 if Lk is not an empty set, connect matrix L and matrix Lk in the horizontal direction to expand matrix L

Step6 make $k = k + 1$, and generate candidate $k+1$ -itemset $Ck+1$ from frequent k -itemset Lk . The main steps are as follows:

Step6.1 let m be the number of rows of Lk and n be the number of columns of matrix Lk

Step6.2 create a zero matrix C with 0 rows and n columns; let $i = 1$

Step6.3 make $a = 1$

Step6.4 make $num = 0$, $product = 1$, and $j = 1$

Step6.5 if $Lk(i, j) = 1$, let $num = num + 1$, $product = product * Lk(i, j) * Lk(i + a, j)$

Step6.6 if $num \geq k - 1$ or $product = 0$, turn to step6.8

Step6.7 $j = j + 1$, if $j \leq n$, turn to step6.5; Otherwise, turn to step6.8

Step6.8 if $product = 0$, turn to step6.26

Step6.9 make $p = 0$, $q = 0$, $j = 1$

Step6.10 if $Lk(i, n - j + 1) = 1$, let $p = n - j + 1$

Step6.11 if $Lk(i + a, n - j + 1) = 1$, let $q = n - j + 1$

Step6.12 if $p * q > 0$, turn to step6.14

Step6.13 $j = j + 1$, if $j \leq n$, turn to step6.10; Otherwise turn to step6.14

Step6.14 make $tail = Lk(:, :p) \& Lk(:, :q)$

Step6.15 make $flag = tail$

Step6.16 find the elements equal to 1 in the i -th row of Lk , and store the subscripts of these elements with matrix $ind1$

Step6.17 make $x = 1$

Step6.18 make $ind2 = ind1$ and delete the x -th element of $ind2$

Step6.19 make $y = 1$

Step6.20 make $flag = tail \& Lk(:, :ind2(y))$

Step6.21 $j = j + 1$, if $y \leq k - 2$, turn to step6.20; Otherwise, turn to step6.22

Step6.22 if the sum of all the elements of $flag$ is 0, turn to step6.24

Step6.23 $x = x + 1$, if $x \leq k - 1$, turn to step6.18; Otherwise turn to step6.24

Step6.24 if the sum of all the elements of $flag$ is not 0, let $c = Lk(i, :) | Lk(i + a, :)$, connect matrix C and c in the horizontal direction

Step6.25 $a = a + 1$, if $i + a \leq m$, turn to step6.4; Otherwise turn to step6.26

Step6.26 $i = i + 1$, if $i < m$, turn to step6.3; Otherwise, turn to step 7. At this time, matrix C is candidate $k+1$ -itemset $Ck+1$

Step7 if C is not an empty set, turn to Step4; Otherwise end

Output: numerical matrix L , representing all frequent itemsets and corresponding support

ALGORITHM 1: The process of the improved Apriori algorithm.

A matrix-based strong association rule extracting algorithm is proposed in this paper. The main contents include (1) block the matrix to deal with frequent itemsets with different item numbers, respectively; (2) search the itemset of postrules layer by layer. The idea is incremental construction method, and the implementation

method is recursive calls of function; and (3) calculate the parameters and screen the rules, and then, output the information.

The algorithm steps are as follows.

The main steps of the function `conf_calculate` called in the algorithm above are as follows.

Input: numerical matrix L

Step1 import the names of the items in the data set and store them in the matrix $label$ in the form of characters

Step2 set the minimum confidence threshold min_conf

Step3 calls function $fopen$ to open text document

Step4 determine the number of non-zero elements in the last row of matrix L and store it in a , then let n be the number of columns of L

Step5 scan matrix L . Divide the itemsets according to the number of items contained, and count the number of each itemset, then store them in matrix $edge$

Step6 make $chunk = mat2cell(L, edge, [n\ 0])$ to complete the blocking

Step7 make $i = 2$

Step8 make $b = chunk\{i, 1\}$

Step9 let aff be the number of rows of B

Step10 make $j = 1$

Step11 find the element equal to 1 in the j -th row of B , and store the subscript in matrix d . Calculate the ratio of the value of $B(j, n)$ to the total number of transactions to obtain the support sup

Step12 make $num = 1$

Step13 make $k = 1$, $index = 1$, and create an empty matrix $str2$

Step14 make $a = chunk\{i - num, 1\}$, $y = chunk\{num, 1\}$

Step15 assign the value of d to the formal parameter ind , the value of $str2$ to the formal parameter $stry$, the value of $index$ to the formal parameter $index$, then call function $conf_calculate$

Step16 $num = num + 1$, if $num < i$, turn to step12; Otherwise turn to step16

Step17 $j = j + 1$, if $j \leq aff$, turn to step10; Otherwise turn to step17

Step18 $i = i + 1$, if $i \leq a$, turn to step8; Otherwise turn to step19

Step19 call the function $fclose$ to save and close the document

Output: text document recording strong association rules

ALGORITHM 2: The process of the matrix-based strong association rule extracting algorithm.

Step1 make $j = index$

Step2 extract the j -th subscript of the remaining subscripts in ind

Step3 save the extracted subscript, and make its corresponding item be one of the post rules, then connect the string of corresponding item name with $stry$

Step4 delete the extracted subscript in ind

Step5 if the number of subsequent rules is less than num , assign the value of ind to the formal parameter ind , the value of $stry$ to the formal parameter $stry$, the value of j to the formal parameter $index$, then call function $conf_calculate$; Otherwise, turn to step6

Step6 connect the string of item names corresponding to all the remaining subscripts in ind and store them in $strx$

Step7 find the corresponding support count in B and Y according to the divided pre rule itemsets and post rule itemsets, then calculate the ratio to the total number of transactions separately to obtain the support $supx$ and $supy$

Step8 calculate the confidence $conf$ and the correlation measure $lift$, $conf = sup/supx$, $lift = conf/supy$

Step9 if $conf \geq min_conf$, make $str = [strx '='>' stry' Support=' num2str(sup) Confidence=' num2str(conf) Lift=' num2str(lift)]$, call function $fprintf$ to write str into text document; Otherwise turn to step10

Step10 make $j = j + 1$, if $j \leq i - num + 1$, turn to Step2; Otherwise end

ALGORITHM 3: The process of function $conf_calculate$ called in Algorithm 2.

2.3. Parameters of Algorithms. According to the basic principle, the parameters of the classical algorithm and the improved algorithm include support, minimum support, confidence, minimum confidence, and lifting degree. In practical applications, the scanning results of support are values of quantity. In order to avoid repeated conversion, the minimum support count is used to replace the minimum support in parameter setting (support can be considered as the ratio of the support count to the total number of transactions). Therefore, the parameters that need to be set by the users are the minimum support count and the minimum confidence; support, confidence, and lifting degree are output parameters. In addition, if the users only need to extract partial association rules, the setting of the lifting degree

threshold can also be added, which has a certain filtering effect.

2.4. Analysis of Algorithm Complexity. The complexity of the two algorithms is further analyzed theoretically. The difference is in the process of generating candidate k -itemsets using frequent $k - 1$ -itemsets. In terms of time complexity, the duplicates generated by the classical algorithm in the connection process need to be identified by comparison one by one, and the number of calculations required for comparison is

$$(k - 1)|L_{k-1}|^2, \quad (3)$$

where $|L_{k-1}|$ represents the length of frequent $k-1$ -itemsets. The redundant items that do not meet the conditions need to be pruned, and the number of calculations is

$$|C_k| \cdot k \cdot |L_{k-1}| \cdot (k-1), \quad (4)$$

where $|C_k|$ represents the length of the candidate k -itemsets before pruning. Add formula (3) and formula (4) to get the number of calculations required by the classical algorithm when generating candidate k -itemsets:

$$(k-1) \cdot |L_{k-1}| \cdot (|L_{k-1}| + |C_k| \cdot k). \quad (5)$$

For the improved algorithm proposed in this paper, assume that the total number of items included in the data set is n , and the maximum number of calculations required for the filtering process is

$$(k-1) \cdot n. \quad (6)$$

The maximum number of calculations required for the validation process is

$$(k-1) \cdot 2 \cdot (3n + (k-1) \cdot (k-2)). \quad (7)$$

Add formula (6) and formula (7) to get the number of calculations required by the improved algorithm when generating candidate k -itemsets:

$$(k-1) \cdot (7n + (k-1) \cdot (k-2)). \quad (8)$$

Obviously, the computing times of the classical algorithm mainly depend on the length of each itemset in the running process, which can usually reach several times of the number of items, while the computing times of the improved algorithm depend on the number of items in the candidate itemsets and the total number of items contained in the data set, so the time complexity of the improved algorithm is lower. With the increase of the size of the data set, the calculation amount of the classical algorithm increases significantly, while the calculation amount of the improved algorithm is in a relatively controllable range, with better time characteristics.

In terms of spatial complexity, the classic algorithm uses the combination method to connect directly. The length of the candidate itemset generated is $C_{|L_{k-1}|}^2$, where $|L_{k-1}|$ represents the length of the frequent $k-1$ -itemset. Because the candidate itemset contains a large amount of duplicates and redundant items, more memory space will be occupied. The length of the candidate itemset generated by the improved algorithm in the worst case is C_n^k , where n is the total number of items contained in the data set. The improved algorithm produces a candidate itemset without redundancy, so when the data set is larger, its length will be much less than the candidate itemset generated by the classical algorithm, which can greatly reduce the storage space occupation.

TABLE 1: Experiment data set.

Data set	Number of transactions	Number of items	Average length of frequent itemsets
D1	500	10	5
D2	800	14	7
D3	1800	27	9
D4	6000	100	20
D5	8125	125	25

3. Experimental Results and Analysis

3.1. Comparative Analysis of Algorithms. In order to test the mining ability and operation efficiency of the improved algorithm, the classical Apriori algorithm and the algorithm proposed in this paper are compared and analyzed through experiments in this section. The improved algorithm follows partial ideas of the classical algorithm. Begin with the itemsets with the minimum item number, and find out the objects that meet the requirements, then generate more itemsets with a larger items number based on this, and so on, until the new itemsets cannot be generated. Therefore, the starting search point of the two algorithms is the same, which is all the itemsets with 1 item in the data set and their support count. In order to guarantee a fair comparison as much as possible, the experiments are carried out under the same conditions and the simulations are performed in the same situations, so that the experimental results are only affected by the performance of the algorithms. The experimental platform is PC (Intel Core i5-8265U, CPU 1.60 GHz, RAM 8 GB), and all the algorithms are implemented by MATLAB 2016. The data of several kinds of equipment recorded in the tests is obtained, involving the parameters, status, and environmental conditions related to the equipment. According to the preset number of tests and the number of items included in each test, partial data is extracted from these historical data to form five experimental data sets of different sizes. The basic information of the data sets used in the experiment is shown in Table 1.

In the actual process of equipment quality information mining, there is a comparatively higher requirement for the complexity of mining algorithms and the load of storage resources. Select running time as the evaluation index of algorithm complexity, and use the number of candidate itemsets generated to measure the memory occupation. The comparison experiment is divided into two parts: one is to compare the time consumption of the two algorithms in the same conditions, and the other is to compare the number of generated candidate itemsets when the two algorithms work in the same situation. Run the algorithms in MATLAB environment, repeat the experiment for 5 times in each case, and take the average of the measured values.

3.1.1. Comparison of Time Consumption. The minimum support is set to 10%, and two algorithms are used to mine five data sets, respectively; then, the running time is measured. The results are shown in Figure 1. It can be seen that

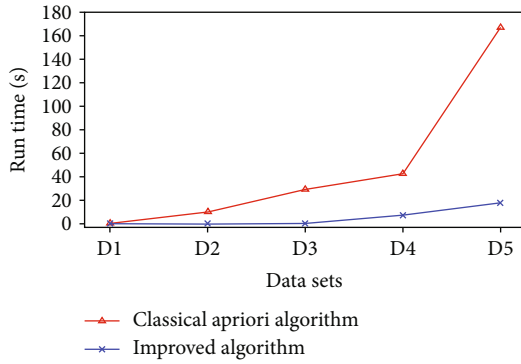


FIGURE 1: Time consumption under different data sizes.

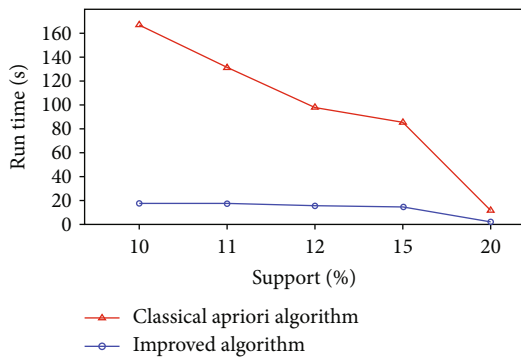


FIGURE 2: Time consumption under different supports.

both of the algorithms can quickly mine all frequent itemsets when the data scale is small, and the running time has little difference. The reason is that the reduction of the number of transactions and items will compress the search space of the algorithms and reduce the size of the candidate sets. As the data scale continues to increase, the running time of the classical Apriori algorithm increases sharply. The reason is that the algorithm contains some unnecessary operations such as connection, comparison, removing duplication, and pruning and needs to scan the database repeatedly. When the data size increases, the running time of the Apriori algorithm with optimized candidate itemset generation strategy increases slowly, indicating that the performance of the algorithm is less affected by the amount of data.

Select data set D5, and set the minimum support to 10%, 11%, 12%, 15%, and 20%, respectively. Measure the running time of the two algorithms on the same data set, and the results are shown in Figure 2. It can be seen from Figure 2 that with the increasing minimum support, the running time of both algorithms tends to decrease. The reason why the classical algorithm shows dramatic change is that the increase of the support threshold will reduce the number of frequent itemsets and candidate itemsets, which will reduce many unnecessary operations during the operation of the classical algorithm. The change of the improved algorithm is relatively gentle, because the algorithm itself does not have a large number of redundant steps and the reduced steps will be fewer.

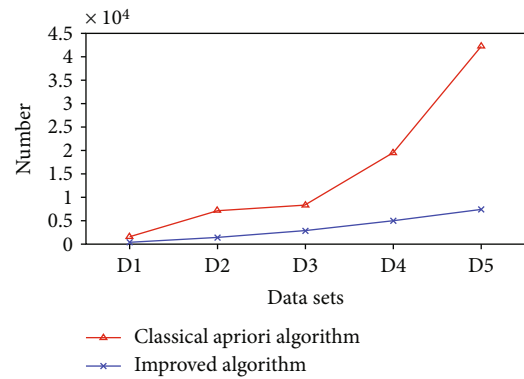


FIGURE 3: Number of candidate itemsets under different data sizes.

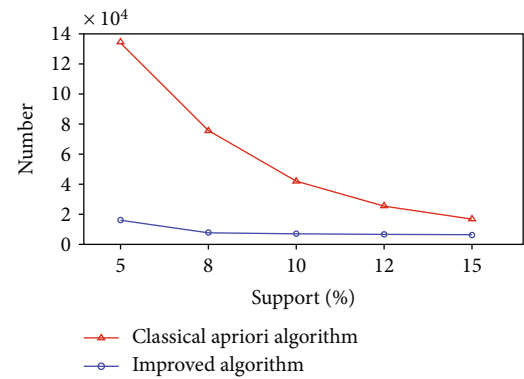


FIGURE 4: Number of candidate itemsets under different supports.

3.1.2. Comparison of Generated Candidate Itemset Number. Set the minimum support to 10%, and the number of candidate itemsets generated by the two algorithms running on five experimental data sets is shown in Figure 3. From the results in Figure 3, for smaller data sets, the numbers of candidate itemsets generated by the algorithms are not much different due to the minor search space. The candidate itemsets generated by the classical algorithm contain duplicate items and redundant items. With the continuous increase of data size, the proportion of invalid itemsets increases significantly. All the candidate itemsets produced by the improved algorithm are valid, and the increment of the total number of candidate itemsets is relatively fewer.

Take data set D5 as an example, and compare the number of candidate itemsets generated by the two algorithms by setting different minimum supports (5%, 8%, 10%, 12%, and 15%), as shown in Figure 4. It can be seen from Figure 4 that with the increase of the minimum support set value, the number of candidate itemsets is gradually decreasing. As a whole, the number of candidate itemsets generated by the improved algorithm is smaller than that of the classical algorithm, especially when the support threshold is set low, mainly because the number of effective itemsets will raise with the decrease of the support threshold; besides, the classical algorithm will also generate duplicate items and redundant items, and the growth of these invalid itemsets is more obvious.

TABLE 2: Parameter setting and experimental results.

Data set	Support threshold	Time consumption/s		Number of candidate itemsets	
		Classical algorithm	Improved algorithm	Classical algorithm	Improved algorithm
D1	10%	0.6	0.1	1584	350
D2	10%	10.3	0.2	7156	1445
D3	10%	29.1	0.4	8321	2893
D4	10%	42.5	7.3	19486	4950
D5	10%	167.2	17.8	42220	7387
D5	10%	167.2	17.8	/	/
D5	11%	131.1	17.3	/	/
D5	12%	97.7	15.5	/	/
D5	15%	85.3	14.9	/	/
D5	20%	11.5	2.3	/	/
D5	5%	/	/	134317	16252
D5	8%	/	/	75624	8028
D5	10%	/	/	42220	7387
D5	12%	/	/	25593	6822
D5	15%	/	/	16961	6430

3.1.3. *Results and Discussion.* The parameter setting and experimental results are summarized in Table 2. From the results of two groups of comparative experiments, compared with the classical Apriori algorithm, performance of the Apriori algorithm with optimized candidate itemset generation strategy has been greatly enhanced. The positive impact of the algorithm proposed in this paper is mainly reflected in the following aspects:

- (1) Better time characteristics: under the same experimental conditions, the improved algorithm consumes less time, especially under more stringent conditions. In the experiment, the time saved by the improved algorithm to complete the same work as the classical algorithm can reach 90%
- (2) Less resource consumption: the improved algorithm significantly reduces the candidate itemsets generated in the mining process, which makes the storage space requirements less
- (3) Better adaptability: with the alteration of experimental conditions, the classical algorithm will change dramatically in both time consumption and memory consumption, while the improved algorithm will be less affected and its performance will be more stable
- (4) A simpler process: the improved algorithm omits the links of deduplication and pruning, so the process of algorithm is generally simplified. These links require a lot of scanning of the database, so the specific steps are also deleted massively

Although the methods proposed in this paper have advantages over the classical methods, the research work still has certain limitations. It is studied in this paper that how to improve the classic Apriori algorithm itself. The idea of the algorithm is changed, but it does not completely deviate

from the basic idea of the original algorithm. Due to the restrictions of research conditions, the optimization method is relatively simple. If other methods are integrated or a better platform is used on this basis, it may achieve better results. The contradictions considered in the improvement process are not comprehensive enough, and there may be room for optimization in other aspects of the algorithm, such as pruning method, itemset frequency calculation, and other processes that seem still need much scanning.

3.2. *Simulation Experiment and Analysis.* Some equipment has the characteristics of complex structure, long storage time, and short working time, so it is very necessary to predict its life. If the effectiveness of the equipment at present and in a certain period of time in the future can be known more accurately, the equipment using unit can carry out replacement maintenance in a timely manner. It is also helpful for the equipment development unit to find the defects of the equipment and improve the design and upgrade.

For complex equipment, it is very difficult to directly establish the degradation model of the whole machine. More feasible way is to adopt appropriate evaluation methods for different levels of equipment. For example, the comprehensive evaluation method based on block diagram is adopted for the system level, the local circuit SPICE is adopted for the component level, and the degradation mechanism model is established for the component level or material level [27–29].

A large number of reliability test data have been collected during the long-term storage and use of a certain type of equipment. These data can reflect the characteristics of this type of equipment in the process of natural degradation. Compared with the data obtained from the accelerated test, the reliability is higher and has greater research value. In this section, the equipment quality information mining method based on the improved Apriori algorithm will be used for experiments.

TABLE 3: Partial test data of the equipment.

Serial number	Infrared device delay too high	Rudder reverse deflection speed too high	5 years	Temperature 22.3-23.5°C	Humidity 54.5%-55.5%
1	1	0	0	0	0
2	0	0	0	1	1
3	0	1	0	0	0
4	1	0	1	0	0

TABLE 4: Partial results after mining.

Serial number	Association rule	Support	Confidence	Lift
1	Infrared device delay is on the high side => rudder reverse deflection speed too high	40.27%	89.35%	1.2
2	6 years => remote control response test value on the high side	23.36%	82.45%	1.3
3	8 years => remote control response test value too low	26.47%	85.67%	1.3
4	5 years and humidity exceeding 67.5% => infrared device capacitance termination voltage is on the high side	15.35%	80.05%	1.4

3.2.1. Data Set. The data used in this experiment are the data recorded after testing the same type of equipment in different batches during storage. According to the measured values of various indicators of equipment, it can be divided into three states: normal, degradation, and failure. Degradation can be divided into two trends: increasing and decreasing. Failure can be divided into two situations: less than the minimum threshold of the nominal value and greater than the maximum threshold of the nominal value. In addition, some storage conditions are also recorded, such as storage time, temperature, and humidity. The items that occur in each test are represented by "1," and those that do not occur are represented by "0." Some data are shown in Table 3.

3.2.2. Information Mining. Using the mining algorithm proposed in this paper, the minimum support is set to 10%, the minimum confidence is set to 80%, and all frequent itemsets are obtained from the data set. Strong association rules are generated according to frequent itemsets. Table 4 shows some mining results. If users want to further extract key rules from the literature, they can use the postrules or the prerules as constraints to filter out association rules that need to be focused on. For example, taking the quality degradation of a certain component as a postrule can discover the potential factors affecting this component. If it is taken as a prerule, the result will be the impact of its degradation on other parts of the equipment, subsystems, performance indicators, etc.

The results of the simulation experiment indicate that the delay of the infrared device is closely related to the rudder reverse deflection speed, and the degradation of the delay of the infrared device may lead to the failure of the rudder reverse deflection speed. The response test value of the remote control device is prone to degradation or failure with the extension of the storage time, and the remote control device is a weak link. Long-term storage under the condition of humidity exceeding 67.5% will easily lead to the degradation of the performance of the capacitor of the infra-

red device, and its termination voltage will show an increasing trend.

From the experimental results, we can find the degradation law of some indexes of this type of equipment, as well as some weak links. These conclusions can be used in the next equipment reliability analysis. For example, some weak links can be disassembled in the life prediction, so as to improve the work efficiency. The stress shall be properly loaded during the acceleration test to shorten the test cycle.

As quality measurement methods commonly used for complex equipment, the method of establishing degradation mechanism model requires a lot of ready-made reusable component-level models and material-level models, and a relatively complete underlying model library is very necessary. Local circuit SPICE can evaluate the impact of component degradation on the key parameters of the circuit and can also analyze the failure threshold of components from the performance of the circuit. Whether it is transferred up or decomposed down, it is inseparable from prior knowledge. The premise for realizing the comprehensive evaluation method is that there is a lot of basic work to provide support. The method proposed in this paper is independent of prior knowledge and directly obtains the laws hidden behind the data, which can make up for the defects of common methods. The useful information obtained by mining includes those consistent with the results of traditional methods, as well as some knowledge that is difficult to obtain by common methods.

The simulation results show that the equipment quality information mining method proposed in this paper is effective. This is a data-driven method, which is not restricted by specific objects and equipment models. It has better generalizability and can be used as a general method. This method is suitable for mining Boolean data, while equipment quality-related data may contain other types of data, so it has certain limitations. Continuing to improve methods to adapt it to other data types or make it have the ability to process data may be a future research direction.

4. Conclusions

In this paper, an Apriori algorithm with optimized candidate itemset generation strategy that is originated from the notion of association rule is proposed. Association rule mining can effectively obtain useful information hidden behind massive data. A matrix-based strong association rule extracting algorithm is proposed for data processing and information representation. Applying the proposed algorithms to complex equipment, an equipment quality information mining method based on improved Apriori algorithm is developed.

In five experimental data sets, the proposed algorithm and the classical algorithm were tested with respect to time complexity and spatial complexity. According to all the metrics, the proposed improved algorithm is significantly superior. Taking a certain type of equipment as an example, a simulation experiment was conducted on its reliability data set to test the effectiveness of the proposed equipment quality information mining method.

The research in this paper has significantly improved the performance of a mature association rule mining algorithm and further enriched association rule mining methods based on previous achievements. However, the optimization method proposed in this article has limitations such as a single approach and incomplete consideration of contradictions. Integrating other methods or platforms, as well as optimizing other aspects of the algorithm, seems to be worth expecting from the future papers. In addition, a universal method for mining equipment quality information is proposed in this article, which expands the application of association rule mining technology in the field of equipment support. More abundant information support will be provided for the implementation of equipment quality degradation characteristic analysis, equipment remaining using life prediction, and other research works. Due to the limitations of the data types applicable to the proposed method, it may be the direction of further research to try to adapt it to other data types or have the ability to process data.

Data Availability

The data used to support the findings of this study have not been made available because it is not allowed to be disclosed.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (Grant No. 61501493).

References

- [1] Z. Y. Zhao, Z. Jian, G. S. Gaba, R. Alroobaea, M. Masud, and S. Rubaiee, "An improved association rule mining algorithm for large data," *Journal of Intelligent Systems*, vol. 30, no. 1, pp. 750–762, 2021.
- [2] H. R. Wang, P. Huang, and X. Chen, "Research and application of a multidimensional association rules mining method based on OLAP," *International Journal of Information Technology and Web Engineering (IJITWE)*, vol. 16, no. 1, pp. 75–94, 2021.
- [3] M. Li, "Investigation on application of association rule algorithm in English teaching logistics information," *Cluster Computing*, vol. 22, Supplement6, pp. 13703–13709, 2019.
- [4] G. F. Tian and H. W. Ma, "Application of multidimensional association mining in shield tunneling optimization," *Geotechnical and Geological Engineering*, vol. 37, no. 3, pp. 1869–1876, 2019.
- [5] R. M. Vazquez, E. Bonilla, E. Sanchez, O. Atriano, and C. Berruecos, "Application of data mining techniques to find relationships between the dishes offered by a restaurant for the elaboration of combos based on the preferences of the diners," *Applied Computer Science*, vol. 15, no. 2, pp. 73–88, 2019.
- [6] R. Agrawal and R. Srikant, "Fast algorithm for mining association rules in large database," in *Proc of the 20th International Conference on Very Large Databases*, pp. 487–499, Santiago, Chile, 1994.
- [7] J. W. Han, J. Pei, and Y. W. Yin, "Mining frequent patterns without candidate generation," in *Proc of the 2000 ACM SIGMOD International Conference on Management of Data*, pp. 1–12, Dallas, Texas, 2000.
- [8] H. Y. Zhou and K. Hirasawa, "Evolving temporal association rules in recommender system," *Neural Computing and Applications*, vol. 31, no. 7, pp. 2605–2619, 2019.
- [9] Y. Dounia, Y. Hayat, and C. Samira, "Extraction of the association rules from artificial neural networks based on the multi-objective optimization," *Network: Computation in Neural Systems*, vol. 33, no. 3-4, pp. 233–252, 2022.
- [10] E. V. Altay and B. Alatas, "Differential evolution and sine cosine algorithm based novel hybrid multi-objective approaches for numerical association rule mining," *Information Sciences*, vol. 554, pp. 198–221, 2021.
- [11] K. Kala, "An effective text mining framework using adaptive principle component analysis," *Multimedia Tools and Applications*, vol. 81, no. 30, pp. 44467–44485, 2022.
- [12] D. Sasikala and K. Premalatha, "A swarm-optimized tree-based association rule approach for classifying semi-structured data using soft computing approach," *Soft Computing*, vol. 25, no. 20, pp. 12745–12758, 2021.
- [13] E. V. Altay and B. Alatas, "Chaos numbers based a new representation scheme for evolutionary computation: applications in evolutionary association rule mining," *Concurrency and Computation: Practice and Experience*, vol. 34, no. 5, article e6744, 2022.
- [14] J. R. D. Arcos and A. A. Hernandez, "Efficient Apriori algorithm using enhanced transaction reduction approach," in *2019 IEEE 13th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, pp. 97–101, Bali, Indonesia, 2019.
- [15] C. X. Wang and X. Zheng, "Application of improved time series Apriori algorithm by frequent itemsets in association rule data mining based on temporal constraint," *Evolutionary Intelligence*, vol. 13, no. 1, pp. 39–49, 2020.
- [16] F. Chiclana, R. Kumar, M. Mittal, M. Khari, J. M. Chatterjee, and S. W. Baik, "ARM-AMO: an efficient association rule mining algorithm based on animal migration optimization," *Knowledge-Based Systems*, vol. 154, pp. 68–80, 2018.

- [17] Z. Mar and K. K. Oo, "An improvement of Apriori mining algorithm using linked list based hash table," in *2020 International Conference on Advanced Information Technologies (ICAIT)*, pp. 165–169, Yangon, Myanmar, 2020.
- [18] R. Sun and Y. Li, "Applying prefixed-itemset and compression matrix to optimize the MapReduce-based Apriori algorithm on Hadoop," in *Proceedings of the 2020 9th International Conference on Software and Computer Applications*, pp. 89–93, Langkawi Malaysia, 2020.
- [19] S. S. Xu, "An Apriori algorithm to improve teaching effectiveness," *International Journal of Performability Engineering*, vol. 16, no. 5, p. 792, 2020.
- [20] M. Sornalakshmi, S. Balamurali, M. Venkatesulu et al., "An efficient Apriori algorithm for frequent pattern mining using MapReduce in healthcare data," *Bulletin of Electrical Engineering and Informatics*, vol. 10, no. 1, pp. 390–403, 2021.
- [21] P. He, B. Zhang, and S. Shen, "Effects of out-of-hospital continuous nursing on postoperative breast cancer patients by medical big data," *Journal of Healthcare Engineering*, vol. 2022, Article ID 9506915, 14 pages, 2022.
- [22] H. L. Yu, J. Wen, H. M. Wang, and L. Jun, "An improved Apriori algorithm based on the Boolean matrix and Hadoop," *Procedia Engineering*, vol. 15, pp. 1827–1831, 2011.
- [23] S. Singh, R. Garg, and P. K. Mishra, "Performance optimization of MapReduce-based Apriori algorithm on Hadoop cluster," *Computers & Electrical Engineering*, vol. 67, pp. 348–364, 2018.
- [24] P. Singh, S. Singh, P. K. Mishra, and R. Garg, "A data structure perspective to the RDD-based Apriori algorithm on Spark," *International Journal of Information Technology*, vol. 14, pp. 1–10, 2019.
- [25] F. Ye, "Research and application of improved Apriori algorithm based on Hash technology," in *2020 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, pp. 64–67, Dalian, China, 2020.
- [26] X. Yuan, "An improved Apriori algorithm for mining association rules," *AIP Conference Proceedings*, vol. 1820, no. 1, article 080005, 2017.
- [27] B. Jakkula, G. R. Mandela, and C. S. N. Murthy, "Reliability block diagram (RBD) and fault tree analysis (FTA) approaches for estimation of system reliability and availability – a case study," *International Journal of Quality & Reliability Management*, vol. 38, no. 3, pp. 682–703, 2021.
- [28] A. Aziz, S. Ghosh, S. Datta, and S. Gupta, "Physics-based circuit-compatible SPICE model for ferroelectric transistors," *IEEE Electron Device Letters*, vol. 37, no. 6, pp. 1–808, 2016.
- [29] X. S. Si, W. Wang, C. H. Hu, M. Y. Chen, and D. H. Zhou, "A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation," *Mechanical Systems and Signal Processing*, vol. 35, no. 1-2, pp. 219–237, 2013.