An integrated method for hiding sensitive association rules of the supply chains

Hui Cheng 1 | Wenjie Zhang 2 | Zhaoyang Wang 2 | Fengjuan Zuo 2 | Zaifang Zhang 2

1 Shanghai Aerospace Equipments Manufacturer Co., Ltd, Shanghai, China
2 School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China

Correspondence
Zaifang Zhang, School of Mechatronic Engineering and Automation, Shanghai University, Shanghai, China.
Email: zaifangzhang@shu.edu.cn

Abstract
Sensitive association rule hiding is an important issue of data sharing for supply chains, which can ensure mutual benefits and avoid information leakages among different enterprises. An integrated method is proposed by using Apriori and the discrete binary particle swarm optimization (BPSO) algorithm, aiming to improve the rule hiding efficiency and effectiveness. The Apriori algorithm is used to extract the association rules from sharing data. The selected sensitive association rules can be hidden using BPSO based on constructing discrete binary space and multi-objective fitness functions. The proposed method is verified through a case study. The results show that the proposed method can effectively hide sensitive information and protect enterprises’ business benefits.

1 | INTRODUCTION

The supply chain effectively integrates the relevant superior resources of different enterprises. Supply chain management plays a key role in many engineering fields, such as industry production, service oriented manufacturing, product service system operating, commercial sales, and so on [1, 2]. Its collaboration can achieve complementary advantages among different enterprises, improve strain capacity and quick response ability to market demands, reduce production and logistics costs, drive market sharing and increase sales [3]. Data sharing, such as transaction records, is a prerequisite for successful supply chain collaboration. Interesting association relationships among these shared data records can offer assistance for production and operation decision. These include product design, procurement planning, cross marketing, inventory control, and so on [4, 5].

However, secret information may be leaked due to association rules’ mining from shared data [6, 7]. The leakages bring some risks to information providers, including market shrinking, revenue reducing, competitiveness decreasing etc.[8]. Privacy-preserving technologies have received widespread attentions in order to avoid the risks of important information leakage during the data sharing process [9–11]. The sensitive association rules that may be used for important business/industry decisions are hidden by using cleaning operations on the original database. The cleaning operations ensure and prevent all negative impacts as small as possible on the non-sensitive association rules. The cleaned database can still be used for information sharing among supply chains.

Association rule hiding was first analysed by Atallah et al. [12]. Two fundamental approaches were presented in order to protect the sensitive rules from disclosures, which include modifying transaction records of supporting sensitive rules and hiding frequent item-sets of generating rules. Verykios et al. [13] proposed association rule hiding methods through deleting or inserting some items of specific records with minimising dataset modification. Wang et al. [14] proposed two methods of increasing the supports of rule antecedents and decreasing the supports of consequences for hiding sensitive rules. Belwal et al. [15] developed a method of increasing support of rule antecedents by changing the definitions of support and confidence, as well as introducing calculating counters. Domadiya & Rao [16] used the algorithm of decreasing the support of rule consequences, where frequent items were handled with sensitivity rankings. A sliding window algorithm and a fast hiding sensitive association rule algorithm were proposed in [17, 18], respectively. The two algorithms can reduce the...
number of scans to the database and improve calculating efficiencies. In order to improve the intelligent hiding levels, Khan et al. [19] and Lin [20] used genetic algorithm (GA) to hide the sensitive association rules with different database operation modes and optimization functions, respectively. Lin et al. [21] used a particle swarm optimization (PSO) algorithm to hide sensitive rules with minimising negative effects. Afshari et al. [22] introduced the Hamming distance into the migration operation to avoid falling into local optimum during the application of the cuckoo optimization algorithm. In [23], the potential risks in retail supply chain collaboration were analysed and an algorithm was developed based on frequent itemsets intersections. Akbar & Asadollah [24] presented a structured analysis for the existing challenges and directions of state-of-the-art sanitisation algorithms.

The above algorithms play important roles in hiding sensitive rules for industry and business. However, these association rule hiding algorithms are mostly based on both decreasing the support and confidence of sensitive rules as well as hiding frequent itemsets. Also, association rule mining process is rarely discussed. Two issues should be further improved as follows: (1) mine or identify the sensitive association rules; (2) increase hiding efficiency. An integrated method is proposed using the Apriori and the discrete binary particle swarm optimization (BPSO) algorithm for hiding sensitive association rules. The Apriori is used to extract association rules from sharing information. The sensitive association rules are selected by using support and confidence rankings or related specific items. These selected sensitive rules can be hidden by using BPSO algorithm within limited victim itemsets.

This article is organised as follows: Section 2 presents sensitive association rule hiding in supply chain management. Section 3 proposes the integrating algorithm for association rule hiding using Apriori and BPSO. A case study is taken as an illustrative example in Section 4. Finally, contribution is summarised and the future work is discussed in Section 5.

2 SENSITIVE ASSOCIATION RULE HIDING PROBLEM

Relationship risks, caused by the lack of necessary communications among enterprises, are the main risk forms for supply chain from cooperation perspective [25]. In order to decrease the relationship risks and obtain an accurate market demand information, an agreement is generally reached on sharing the transaction data among suppliers, manufacturers and sellers. This can be advantageous to develop reasonable production plans, reduce inventories and increase sales. This way of data sharing can reduce the costs and expand the benefits, while may also bring potential risks to data providers. The interests or competitiveness of different enterprises will be changed due to the unintentional disclosure of sensitive knowledge in the released or shared data. Protecting the sensitive knowledge becomes a critical step of sharing data for guaranteeing sustainable enterprise cooperation and core business interests.

A data sharing model is given in Figure 1 with taking a Seller and a Supplier as an example. Protecting sensitive knowledge is generally achieved using sensitive association rule hiding. The general principle of hiding sensitive association rules can be summarised as follows: mine the association rules from the original transaction dataset of the Seller, and get a series of association rules set, R. The subset Rs of R is identified as sensitive rules. These sensitive rules can be hidden using the proposed algorithm for building a new cleaned database. Supplier will mine association rules for their production or business decision using the cleaned database. The mined result is the subset $R' = R - Rs$. Necessary data is shared and sensitive knowledge leakage is avoided.

The following described scenario demonstrates how business groups use data mining technique to obtain confidential information and gain superiority in market competition [26]. The significance of hiding sensitive association rules will be more clearly recognized for supply chain management.

Suppose Seller A has an agreement with Supplier (b) If A is willing to let B read sales transaction database, B will supply products to A at a more favourable price in return. A accepts the agreement, and B can use association rule mining tool to analyse the database. A sensitive rule is found where customers who usually buy Product Z also purchase their competitor C’s paper products. As a result, B launches a promotional strategy - customer will get a discount on the purchase of Supplier B’s paper products when buying Product Z. The sales promotion will seriously affect C’s paper sales. B acquires the market initiative. However, C increases the supply price for A due to the paper sales reduction. The next time, B may refuse to provide A with preferential supply prices because of competitor reduction. This scenario makes A passive in the relationships with suppliers.

Conversely, Seller A released a cleaned database by hiding the sensitive association rules that can influence key decision results in advance, which will make a difference. For example, A removes some transactions of company C’s paper products from the shared database in order to hide the sensitive rule ‘Product Z → C’ paper’. So B cannot extract the sales relationship between these two products and will not promote their products sale with Product Z. As a result, A will not only

![Figure 1](image-url)
maintain the sales of C’s papers, but also gets B’s products and C’s papers at discounted prices. In this way, A will maintain a stable and sustainable cooperation with both B and C.

3 SENSITIVE RULE HIDING USING APRIORI AND BPSO

Sensitive association rule hiding is shown in Figure 2. The first step uses the Apriori algorithm to mine the association rule set R from the transaction database D. The second step is to find out sensitive association rules Rs that affect important business decisions. Victim items are then identified according to these sensitive rules. The final step applies the BPSO algorithm to hide these sensitive rules based on victim items and acquires the cleaned database D’ for data sharing.

3.1 Association rules mining

Assume that I = \{i1, i2, i3, ..., im\} is the complete item set and D is the transaction database. Each transaction record T of D is a subset of I, which is called TID. R is the extracted association rule set from D and Rs is sensitive association rule set, where Rs ⊂ R. Each association rule is expressed as ‘X → Y’, where X is called rule antecedent (RA) and Y is named as the rule consequence (RC). \(X, Y \subseteq I\) and \(X \cap Y = \emptyset\).

The Apriori algorithm is used to mine association rules in the transaction database D. Two indicators of support and confidence are used during the mining process, which are the measures of rule importance and accuracy. Support and confidence are abbreviated as Sup and Con, respectively. Their calculation is given according to the formula (1) and (2), respectively.

\[
\text{Sup}(X \Rightarrow Y) = \frac{|X \cup Y|}{|D|} \quad (1)
\]

\[
\text{Con}(X \Rightarrow Y) = \frac{|X \cup Y|}{|X|} \quad (2)
\]

where \(|X \cup Y|\) expresses the number of transactions containing both X and Y, \(|D|\) is the total number of transactions, and \(|X|\) is the number of X.

In order to ensure the rule usefulness, the association rule is mined by using a pre-set minimum support threshold (MST) and minimum confidence threshold (MCT).

3.2 Sensitive rules and victim items identifying

In order to identify the critical information, sensitive association rules should be selected from the association rule set by using support and confidence rankings or related specific items. Specific items are generally determined by considering competition or confidentiality. Also, a set of sensitive rules Rs is identified from R.

Generally, the sensitive rule is considered to be hidden when its support and confidence levels are lower than the given thresholds. To successfully hide the sensitive association rules, it is only necessary to reduce the values of Sup and Con lower than the thresholds. The reduction can be achieved through revising some transactions in database D. However, it is unnecessary to operate the entire database because sensitive rules exist only in partial transactions. Otherwise, that will increase the computation complexity and generate a lot of useless solutions. There are two critical steps before hiding rules, which involves selecting key transactions and determining victim items. The key transaction is a record in D that fully supports one or more sensitive rules. The victim is the item that is to be hidden operation. The key transactions are selected to form a new data set Dc according to the sensitive rule set Rs.

It is necessary to reasonably select the victim items in order to successfully hide the sensitive rules and minimise the influences on the non-sensitive rules. The rules are more easily to be hidden successfully by reducing the number of RA than RC according to the formula (1) and (2). The victim items are then selected from RC. Only one victim item is selected for each sensitive rule in order to decrease the impacts on database. If there is only one RC of a sensitive rule, the RC is selected as the victim. For example, a sensitive rule ‘A→B’ has only B as RC, thus B is also the victim. If there are multiple items in RC, the victim is the item with most occurrences in sensitive rules. If the numbers of occurrences are same, and the one with least occurrence times in non-sensitive rules which is selected as the victim item. For example, the sensitive rule ‘A and C → D and E’ has two consequence items, D and E. Compare the numbers of their occurrences in both the sensitive and the non-sensitive rules, and then select the eligible one as the victim.

3.3 Binary particle coding

The PSO algorithm is originally used for solving continuous problem, but many practical engineering problems are described as discrete combinatorial optimization problems.
The proposed BPSO algorithm based on discrete PSO algorithm is given as follows.

Assume \( M \) particle population, where each particle represents a solution in the proposed binary coding method. One or 0 is used to represent the item that appears or disappears in the transactions. The coding diagram is shown in Figure 3. The key transactions \( D_c \) are lined up to form an array represented by 0 and 1. The length of the array is equal to the total number of victim items in key transactions. If there are \( N \) transactions in the key transaction set, each array contains \( N \) parts. The number of victim items contained in each key transaction is represented as \( n_i \), and the sum of the victim items as \( \sum_{i=1}^{N} n_i \) is the array dimension \( N_d \). The remaining \( (M-I) \) particles are generated by randomly quantifying the victim terms to 0 or 1. Finally, a particle population with a size of \( M \times N_d \) is obtained. In the coding mode, the particles are only composed of key transactions supporting sensitive rules, and the array dimension only retains the victim items without considering other items in the transaction dataset. Therefore, this coding method greatly reduces the search space and improves the search efficiency.

3.4 Fitness function building

Fitness is used to evaluate the pros and cons of the particle individual. Fitness function directly affects the execution efficiency of BPSO and the quality of mining results. The association rule hiding algorithm is evaluated by the negative impacts on the database after the hidden operation. In this article, the negative effects involve hiding failure (HF), lost rules (LR), and average hide deviation (AHD). Fitness function adopts the linear combination of objective functions. The calculation formula is:

\[
\text{Fitness}(R_i) = w_1HF + w_2LR + w_3AHD
\]  

(3)

where \( HF, LR, AHD \) are the three objective variables, and \( w_1, w_2, w_3 \) are related weights, respectively.

The first objective variable is calculated as follows:

\[
HF = \sum_{i=1}^{|R_d|} W_i / |R_d|
\]  

(4)

\[
W_i = \begin{cases} 0, & \text{Sup}(I_i) < \text{MST} \text{ or } \text{Conf}(r_i) < \text{MCT} \\ 1, & \text{otherwise} \end{cases}
\]  

(5)

where \(|R_d|\) is the number of sensitive rules; \( I_i \) is the itemset of the rule \( r_i \); \( \text{MST} \) and \( \text{MCT} \) represent the thresholds of support and confidence, respectively.

The second objective variable is calculated as follows:

\[
LR = \sum_{i=1}^{|\hat{R}_d|} W_i / |\hat{R}_d|
\]  

(6)

\[
W_i = \begin{cases} 0, & \text{Sup}(I_i) \geq \text{MST} \text{ and } \text{Conf}(r_i) \geq \text{MCT} \\ 1, & \text{otherwise} \end{cases}
\]  

(7)

where \(|\hat{R}_d|\) is the number of non-sensitive rules; \( I_i \) is the itemset of the rule \( r_i \).

The third objective variable is calculated as follows:

\[
\text{Fit3} = \frac{\sum U_i}{|\hat{R}_d|} + \frac{\sum V_i}{|\hat{R}_d|}
\]  

(8)

\[
U_i = \begin{cases} 0, & \text{Sup}(I_i) < \text{MST} \\ 1, & \text{Conf}(r_i) < \text{MCT} \end{cases}
\]  

(9)

\[
V_i = \begin{cases} 0, & \text{Sup}(I_i) \geq \text{MST} \text{ and } \text{Conf}(r_i) \geq \text{MCT} \\ \text{MCT} - \text{Conf}(r_i), & \text{Conf}(r_i) < \text{MCT} \end{cases}
\]  

(10)

where \(|R_d|\) and \(|\hat{R}_d|\) are the number of sensitive and non-sensitive rules; \( I_i \) is the itemset of the rule \( r_i \).

The three objectives can be obtained without mining the database during iterative calculation. Therefore, the efficiency of the algorithm is improved and the running time is saved. It is necessary to merge the non-key transactions separated in Section 3.2 into the database during the population iteration process. This can make the particles fitness value represent the entire database state.

3.5 Speed and location update

The BPSO algorithm improves the position update formula, where velocity value is converted into a probability that the bit variable takes a value of 1. Function sigmoid is used to map the velocity value into the interval \([0,1]\). The particle velocity and position update formula are expressed as follows:

\[
\text{FIGURE 3 Binary particle coding}
\]
\[ v_{ij}(t + 1) = w_i v_{ij}(t) + c_1 r_1(x_{pbest} - x_{ij}(t)) + c_2 r_2(x_{gbest} - x_{ij}(t)) \]
\[ x_{ij}(t + 1) = \begin{cases} 1, & \text{if} \quad r() < \text{sig}(v_{ij}) \\ 0, & \text{otherwise} \end{cases} \]
\[ w_i = w_{\text{max}} - \frac{t \times (w_{\text{max}} - w_{\text{min}})}{t_{\text{max}}} \]
\[ \text{sig}(v_{ij}) = \frac{1}{1 + e^{-v_{ij}(t+1)}} \]

where \( w_i \) is inertia weight with the linear decreasing model, and \( w_{\text{max}} \) and \( w_{\text{min}} \) represent the maximum and minimum values of inertia weight. The two weights are set as 0.9 and 0.4, respectively; \( t \) represents the current iteration steps, and \( t_{\text{max}} \) represents the maximum number of iteration steps; the parameter \( c_1, c_2 \) are the learning factors; \( r() \) is the random number between \((0,1)\); and the value range of \( \text{sig}(v_{ij}) \) is \((0,1)\).

The maximum velocity \( v_{\text{max}} \) determines the maximum moving distance of an iteration. In general, it cannot exceed the maximum width of a particle. If it is too large, the particles may fly over the optimal solution, and lose local search capability partially. If it is too small, the global search ability of the particles will be reduced, and result in local extremum. The value is generally set according to data characteristics.

### 3.6 Algorithm flow

The proposed rule hiding algorithm using Apriori and BPSO is shown in Figure 4. Phase 1 is the association rule mining process, and phase 2 is the sensitive rule hiding process. The specific steps are described as follows:

**Phase 1:**

1. Select sensitive rule \( R_s \), identify the key transactions \( D_c \) and determine the victim items.
2. Set the population size \( M \), the maximum number of iterations \( L \), support and confidence thresholds \( \text{MST}, \text{MCT} \).
3. Initialise the position and velocity of each particle in the particle swarm randomly.

**Phase 2:**

1. Identify the sensitive rules \( R_s \) and victim itemsets.
2. Initialise the position and velocity. Set iterations \( L \) and population size \( M \).
3. Calculate the fitness value of each particle \( f(X) \).
4. Use Apriori algorithm to obtain association rules \( R \).
5. Keep the original \( p_{\text{best}} \).
6. Update the \( p_{\text{best}} \) of each particle.
7. Let \( X_i \) be \( p_{\text{best}} \).
8. Update speed and position of particle.
9. Condition is satisfied?
10. Output results.

**Figure 4** Process of BPSO based algorithm
(3) Calculate objective values of particles, and calculate the fitness values of particles by using the proposed fitness function.
(4) Compare the current position fitness \(f(x_i)\) of each particle with its own optimal position fitness \(f(p_{best})\). If it is smaller than \(f(p_{best})\), it is updated to the current optimal value.
(5) Find the global optimal particle in the particle swarms, and compare the fitness \(f(x_i)\) with \(f(g_{best})\). If it is less than \(f(g_{best})\), \(f(x_i)\) is updated to the current global optimal value.
(6) Update the velocity and position of the particles according to the formula, and then generate a new population.
(7) Judge whether the end condition is satisfied. If it is satisfied, the optimization ends, and the final result is output. Otherwise, steps (3) to (6) are to be repeated until the iteration satisfies the end condition.

## 4 | CASE STUDY

In many supply chain applications, suppliers usually reach an agreement on sharing the transaction data with the sellers by providing some discounts of order prices. It is critical for sellers to avoid the disclosure of sensitive information caused by sharing data. This study takes a Seller A’s case as an illustrative example. Seller A and Supplier B reach an agreement on sharing the transaction records. The proposed association rule hiding method is applied to the Seller’s original transaction database D for deleting some items of specific transactions in order to hide sensitive rules.

There are 14,517 transaction records among the transaction database D of Seller A during a non-continuous 1 year. The transactions include 16,470 different product numbers. The average transaction length is 13, in other words, each customer buys 13 different products once on an average. The specific transaction data is shown in Table 1.

One purchase per customer is a transaction record: the first column indicates the transaction number, and the second column represents the commodity number included in the corresponding transaction. Most items are identified by a unique barcode, while some item numbers in the data set represent a group of products rather than a single one. One hundred and ten association rules are mined from the database D using the Apriori algorithm with parameters set as MST = 2% and MCT = 10%. Some of the mined rules are shown in Table 2.

For example, the association rule ‘39\(\rightarrow\)48’ numbered one in Table 2 means that the customer who buys Product 39 also buys Product 48 under the specific support and confidence. Supplier B can discover this rule when mining the transaction database, which makes Seller A in face of risk. Supplier B launched a targeted promotion strategy in order to increase the sales of H brand Product 48 to be sold. Customers can enjoy a 10% discount on the H-brand Product 48 while purchasing Product 39. This can significantly increase the sale of H-brand Product 48. On the contrast, this strategy causes rapid drop on the sales of another Brand E’ Product 48 originally sold by Seller A. As a result, A failed to reach the agreed sales volume of the Brand E’ Product 39. The Brand may increase their supply price to A, while supplier B does not agree to give A more discounts due to the reduction of competitors. At this time, the relationships between A and suppliers are passive for influencing the stability of the supply chain.

In order to ensure interests, five sensitive rules (in Table 3) have been identified according to protect specific product sales. The proposed BPSO is used to hide the selected sensitive rules in the set D. The weights of fitness function are set as \(w_1 = 0.7, w_2 = 0.2, w_3 = 0.1\), respectively. \(v_{max}\) is according to the attribute characteristics of the case data. The sensitive association rules that support important business decisions are hid and the cleaned transaction database is released for sharing.

### Table 1: Transaction database of Seller A

| Number | Transaction Records |
|--------|---------------------|
| 1      | 0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29 |
| 2      | 30,31,32           |
| 3      | 33,34,35          |
|        | ⋮                  |
| 2297   | 39,41,48,80,310,426,514,640,777,1184,1393,1756,2011,2437,2524,3245,3521,4843,5153,5154,5155,5156 |
| 2298   | 349,1237,3641,4145,4198,5157 |
|        | ⋮                  |
| 14,517 | 9,38,104,342,371,647,1034,1121,2318,5596,8542 |

### Table 2: Association rules

| Number | Rules | Support(%) | Confidence(%) |
|--------|-------|------------|---------------|
| 1      | 39→48 | 30.02      | 53.8357       |
| 2      | 48→39 | 30.02      | 68.5005       |
| 3      | 39→41 | 20.4519    | 36.677        |
| 4      | 39→54 | 3.7956     | 21.2823       |
| 5      | 41→55 | 3.7956     | 13.5036       |
| 110    | 475→39 | 2.0114    | 66.9725       |

### Table 3: Sensitive rules

| Number | Rule | Support(%) | Confidence(%) |
|--------|------|------------|---------------|
| 1      | 48→41 | 15.4577    | 35.2719       |
| 2      | 38,48→39 | 6.0274 | 75.8232      |
| 3      | 39→32,48 | 6.0481 | 10.8462       |
| 4      | 38→48 | 7.9493     | 46.364        |
| 5      | 41,48→32,39 | 3.024 | 19.5633      |
in order to ensure the healthy cooperation relationship between the two parties. Three indices HF, LR, and AHD are used to evaluate the performance of the proposed algorithm for minimising the negative impacts on transaction data sharing. Furthermore, the results are compared with an improved GA algorithm from four aspects: convergence speed, HF, LR and AHD. The results are given in Figure 5–8.

As shown in Figure 5, the BPSO algorithm can converge before the 25th generation, whereas the improved GA algorithm converges at about the 80th generation. It is obvious that the convergence speed of the proposed algorithm is significantly faster than that of the latter. The improvement of the convergence speed greatly is more important for database with a large capacity. As a result, enterprises can save the time for processing association rules when sharing data and respond to changes of market and supply chain more quickly. A comparison of HF is shown in Figure 6.

**FIGURE 5** Comparison of fitness function

**FIGURE 6** Comparison of rules HF ratio
Both of the two algorithms can achieve the main purpose that the sensitive rules are hidden successfully. But the former is faster, hiding at the fifth generation while the latter at the 25th generation. A comparison of the LR is shown in Figure 7. The LR values obtained are much smaller than the values of the improved GA algorithm because the proposed operation only selects transactions that support sensitive rules for subsequent processing and encoding. Also, as the number of iterations increases, the LR value obtained by the BPSO-based algorithm converges faster, reaching a stable value of 0.55%, while the improved GA algorithm tends to converge at a speed four times slower than the proposed algorithm, reaching a
stable value of 0.65%. Furthermore, due to the unique characteristics of particle swarm, such as self-learning and learning from surrounding particles, particles are easily close to the optimal solution and searched nearby. The weight values of BPSO uses linear weight decreases style, which has strong global search ability in the early stage and local search ability in the later stage. Its is reflected in the AHID comparison chart in Figure 8.

The results show the proposed BPSO-based algorithm hides sensitive rules successfully and has less negative impacts on the database. The proposed algorithm has good application potentials on data sharing and protecting for supply chain management.

Contrast to the current existing research, the proposed algorithm uses consequence operation instead of the antecedent and also during the victim identifying process. The particle coding is limited to victim items of key transactions. The proposed fitness function only needs to traversal access the database once. BPSO has close computational complexity with GA and other similar optimization methods, while BPSO convergence speed is faster. Consequently, the proposed method has lower complexity and better processing ability to tackle with the sensitive rule hiding problem.

Data sharing is one of the key factors for sustainable supply chain. However, it may bring the leakage of sensitive information, which will cause great negative impacts on themselves. The proposed method can effectively support data sharing among enterprises through intelligent hiding of sensitive rules. The negative impact on the actual data as low as possible to ensure that the data can reflect the actual truth. Data sharing among enterprises can be persistent and sustainable to create value and economic benefits. Useful association rules can be identified to promote the formulation of the production strategy and reduce production and inventory costs. At the same time, data sharing will not bring leakage risk of trade secrets to themselves. It can benefit for the entire supply chain. A win-win situation can be achieved, which keeps a good cooperative relationship and maintains the long-term health and stability of the supply chain. The application results also have important guiding significance for other companies.

Furthermore, the proposed method can also be applied to service-oriented manufacturing, product service system and other emerging manufacturing modes. Through the protection of sensitive information and the sharing of non-sensitive information, the proposed method can support to improve the efficiency of cooperation and realize the balanced coexistence of data sharing, privacy protection and sustainable cooperation among different stakeholders.

5 | CONCLUSION

The study elaborates the mechanism of risks caused by the leakage of sensitive information during data sharing of the supply chain. An integrated association rule hiding approach is proposed using the Apriori and the BPSO algorithm. The sensitive rules are selected from the association rules that are mined from transaction database using the Apriori algorithm. The key transactions and victim items are identified, which play remarkable roles in improving the efficiency of the algorithm, optimising the results and decreasing the search space. And the BPSO is proposed for sensitive rule hiding. The fitness function in BPSO algorithm is established based on considering the hiding efficiency, negative impacts on the database, and the hiding deviations. The quality of the optimal solution can be obtained by the algorithm iteration. The BPSO-based association rule hiding algorithm was applied to the actual case. The results show that the proposed method has a higher efficiency and effective hiding ability. The proposed algorithm solves the association rule hiding problem more effectively, which is significant for both the supply chain data sharing and their healthy operation management.

In the future, the study will follow an intelligent identification on the sensitive rules under specific constraints. Database dynamic update will be considered with its influence on sensitive rule extraction.

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (No. 51,205,242). The authors express sincere appreciation to the anonymous referees for their helpful comments to improve the quality of the paper.

REFERENCES

1. Li, H., et al.: Bi-level coordinated configuration optimization for product-service system modular design. IEEE Transactions on Systems, Man, and Cybernetics-Systems. 47(3), 537–554 (2017)
2. Burcu, A., Işık Özge, Y.H., Martin, C.: A quantitative performance management framework for assessing omnichannel retail supply chains. J Retailing Consum Serv. 48(5), 257–269 (2019)
3. Corsten, D., Kumar, N.: Do Suppliers Benefit from Collaborative Relationships with Large retailers? An empirical investigation of efficient consumer response adoption. J Market. 69(3), 80–94 (2005)
4. Kumar, N., Gangopadhyay, A., Karabatis, G.: Supporting mobile decision making with association rules and multi-layered caching. Decis Support Syst. 43(1), 16–30 (2007)
5. Chen, Y.J.: Knowledge integration and sharing for collaborative moulding product design and process development. Comput Ind. 61(7), 659–675 (2010)
6. Zhang, D.Y., Zeng, Y., Wang, L.: Modelling and evaluating information leakage caused by inferences in supply chains. Comput Ind. 62(3), 351–363 (2011)
7. Lee, H.L., Whang, S.: Information sharing in a supply chain. Int J Manuf Technol Manag. 1(1), 79–93 (2000)
8. Connelly, C.E., Zweig, D.: How perpetrators and targets construe knowledge hiding in organisations. Eur J Work Organ Psychol. 24(3), 479–489 (2015)
9. De Geoffroy, Z., Evans, M.M.: Are emotionally intelligent employees less likely to hide their knowledge? Knowl Process Manag. 24(2), 81–95 (2017)
10. Kumar Jha, J., Varkkey, B.: Are you a cistern or a channel? Exploring factors triggering knowledge-hiding behaviour at the workplace: Evidence from the Indian R&D professionals. J Knowl Manag. 22(4), 824–849 (2018)
11. Parmar, S., Gupta, M.P., Sharma, M.P.: A comparative study and literature survey on privacy preserving data mining techniques. Int J Comput Sci Mobile Comput. 4(4), 480–486 (2015)
12. Atallah, M., Bertino, E., Elmagarmid, A.: Disclosure limitation of sensitive rules, Workshop on knowledge and data engineering Exchange. In: Proceedings of Workshop on knowledge and data engineering Exchange, pp. 45–52. Chicago November (1999)

13. Verykios, V.S., Elmagarmid, A., Bertino, E.: Association rule hiding. IEEE Trans on Knowledge and Data Engineering. 16(4), 434–447 (2014)

14. Wang, S.L., Lee, Y.H., Billis, S.: Hiding sensitive items in privacy preserving association rule mining, IEEE International Conference on Systems, Man and Cybernetics, pp. 3239–3244. IEEE Xplore, USA October (2004)

15. Belwal, R.C., Varshney, J., Khan, S.A.: Hiding sensitive association rules efficiently by introducing new variable hiding counter, IEEE International Conference on Service Operations and Logistics and Informatics, pp. 130–134. IEEE Xplore, Beijing CHINA October (2008)

16. Domadiya, N.H., Rao, U.P.: Hiding sensitive association rules to maintain privacy and data quality in database, IEEE International Advanced Computing Conference, pp. 1306–1310. IEEE Xplore, Ghaziabad, India February (2013)

17. Weng, C.C., Chen, S.T., Lo, H.C.: A Novel algorithm for completely hiding sensitive association rules, Eighth International Conference on intelligent systems design and applications, pp. 202–208. IEEE Computer Society, Kaostung, Taiwan December (2008)

18. Mogtaba, S., Kambal, E.: Association rule hiding for privacy preserving data mining, Industrial Conference on Data Mining: ICDM 2016, pp. 320–333. NY July (2016)

19. Khan, A., Qureshi, M. S., Hussain, A.: Improved Genetic algorithm approach for sensitive association rules hiding. World Appl Sci J. 31(12), 2087–2096 (2014)

20. Lin, C.W., et al.: The GA-based algorithms for optimising hiding sensitive itemsets through transaction deletion. Appl Intell. 42(2), 210–230 (2015)

21. Lin, J.C.W., Liu, Q., Fournier-Viger, P.: A sanitisation approach for hiding sensitive itemsets based on particle swarm optimization. Eng Appl Intif Intell. 53(3), 1–18 (2016)

22. Afshari, M.H., Dehlordi, M.N., Akbari, M.: Association rule hiding using cuckoo optimization algorithm. Expert Syst Appl. 64, 340–351 (2016)

23. Le, H.Q., Arch-Int, S., Nguyen, H.X.: Association rule hiding in risk management for retail supply chain collaboration, vol. 64, pp. 776–784 Comput Ind 7 (2013)

24. Akbar, T., Asadollah, S.: Data sanitisation in association rule mining: an analytical review. Expert Syst Appl. 96, 406–426 (2018)

25. Anand, K.S., Goyal, M.: Strategic information management under leakage in a supply chain. Manag Sci. 55(3), 438–452 (2008)

26. Gkoulalas-Divanis, A., Verykios, V.S.: Association rule hiding for data mining. Springer US (2010)

How to cite this article: Cheng H, Zhang W, Wang Z, Zuo F, Zhang Z. An integrated method for hiding sensitive association rules of the supply chains. IET Collab. Intell. Manuf. 2021;1–10. https://doi.org/10.1049/cim2.12026