Swarm Optimization and Iterative Privacy Generates a Sensitive Rule with the Constraints Data

R. Sasikala
Assistant Professor
Department of Computer Science & Engineering
V.S.B College of Engineering, Coimbatore, India

Abstract— Recently, motivating the demand for the privacy and secure data mining research is the expansion of techniques that include the privacy and security along with the effective data publishing. Most of the research work is developed for the data distribution with the privacy. However, the protocols used in the homomorphic encryption which increased the computational costs and communication. In order to overcome the limitations, a Swarm optimization and Iterative Privacy Rule Preservation (SIPRP) method is designed in the paper to improve the efficiency of the privacy preserving association rule mining with the constraint minimization. Initially, SIPRP method generates the association rules for the privacy preserving distribution database based on the support and confidence threshold. Finally, the SIPRP method obtains the sensitive sets of items for generating the specific sensitive. Experimental evaluation of the SIPRP method is done with the performance metrics such as the number of sensitive rule.

Key words: Computational Cost, Association Rule Mining, Sensitive Item Sets, Sensitive Rules, Principle Component Analysis

I. INTRODUCTION

Most of the research work is developed in the Privacy Preserving Data Mining (PPDM) for hiding the private, confidential, or secure information. Protocol to the secure mining of the association rule was developed in [1] for providing the secured mining association rules using two secure multi-party algorithms. However, the method increased the computational costs. Corporate privacy preserving framework was designed in [2] that introduced an Encrypt/Decrypt (E/D) module to change the client data before it delivered to the server and improves the true patterns with their correct support. However, E/D module assumes that the attacker does not possess knowledge on the hiding aspect and relaxation may break the vulnerabilities encryption scheme and bring privacy. It is ambiguous in providing the corporate privacy preserving association rule mining.

Homomorphic matching technique was introduced in [3] the privacy preservation for improving the privacy level. The secrecy views and null based virtual updates was illustrated [4] for achieving data privacy for reducing the computation cost. The Direct and indirect discrimination was performed [5] using the legitimate classification rules while preserving data quality which results in the improved privacy level at the cost of accuracy.

In the paper, Swarm optimization and Iterative Privacy Rule Preservation (SIPRP) method is designed for enhancing the efficiency of the privacy preserving association rule mining with constraint minimization. In SIPRP method, sensitive rules are subjected to the Particle Swarm Optimization (PSO) for hiding and preserving highly confidential privacy rules. The SIPRP method hides the sensitive rules with aiming at the privacy preserving distribution database.

II. DESIGN OF SWARM OPTIMIZATION AND ITERATIVE PRIVACY RULE PRESERVATION (SIPRP) METHOD

The design of Swarm optimization and Iterative Privacy Rule Preservation (SIPRP) method is described in a detail manner this section. The main goal of the SIPRP method is to hide the sensitive rules form the public aiming at improving the privacy rate. Initially, the SIPRP method generates the association rule based on their support and confidence threshold. The sensitive rules associated with the optimal sensitive item is hidden and then they are estimated for hiding sensitive rules.

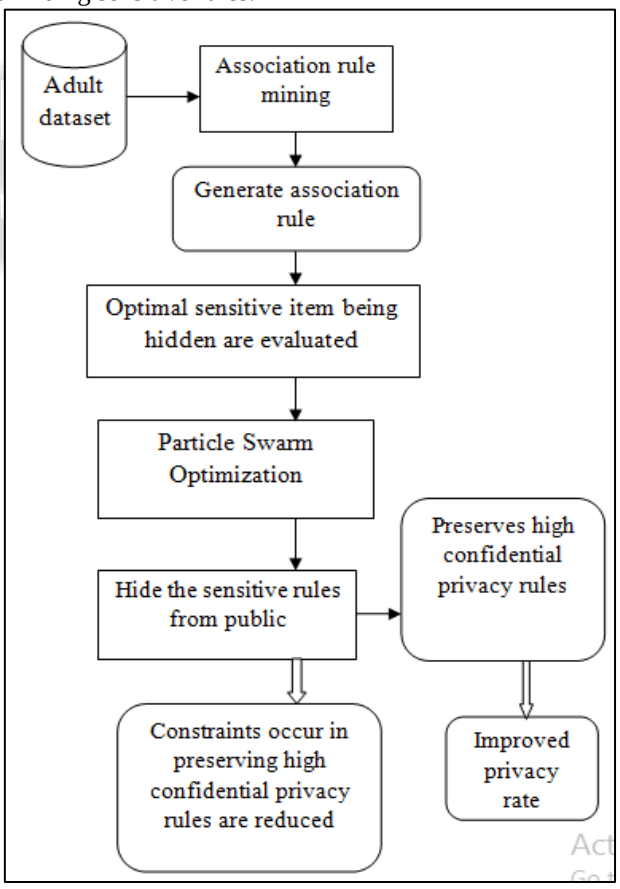

Fig. 1: Architecture diagram of SIPRP technique for sensitive rule hiding

SIPRP method hides the sensitive rules using the Particle Swarm Optimization (PSO) mechanism with the objective of preserving highly confidential privacy rules. The SIPRP method reduces the constraints occur while preserving the

high confidential privacy rules through the iterative generations rules for diverse sensitive sets of items. Finally, the SIPRP method ensures the sensitive rules to be hided with less effect on the privacy which is exposed during the data distribution across multiple users. The architecture diagram of the SIPRP method for hiding sensitive rule is shown in the below Figure 1.

As shown in Figure 1, SIPRP method initially takes the adult data set as an input, and then applies the association rule mining for generating the association rule based on the support and confidence value. After generating the association rule, sensitive rule related with the optimal sensitive item is concealed and evaluated with the objective for improving the privacy rate. Next, the SIPRP method hides the sensitive rules form the public with the help of particle swarm optimization. The PSO mechanism preserves the high confidential privacy rules for reducing the constraints occur which in turn improves the privacy rate of sensitive rules.

A. Association rule mining for generating sensitive rules

SIPRP method generates the sensitive rules with the association rule mining technique. The Association rule mining technique in the SIPRP method protects the sensitive data items by hiding the sensitive rules from the data miners and discloses all the non-sensitive rules to the public. The Association rule mining technique generates the association rule based on the support and confidence threshold value and then evaluation is made. The sensitive rules associated with the optimal sensitive items are hidden to preserve the privacy rate. The task of Association rule mining technique in the SIPRP method is illustrated in the below Figure 2.

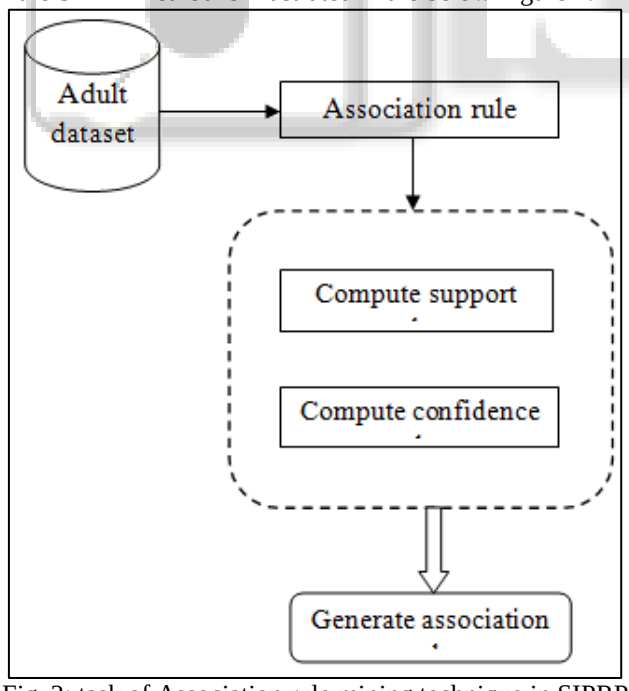

Fig. 2: task of Association rule mining technique in SIPRP method

After performing the association rule, the SIPRP method calculates the sensitive rules associated with the optimal sensitive items. It is hidden for providing high confidential privacy rules.

Let us Consider, 'D' is a Database that consists a set of transactions $D = \{T1, T2, \dots, Tn\}$ and each transaction contains a set of items $I = \{I1, I2, \dots, Im\}$. The Association Rule Mining technique recognizes all association rules $X \Rightarrow Y$ with a minimum support and confidence value. The support value of an item $X \in I$ in the database D is the count of transactions contains X and represented as $Sup\ count(X)$. Support value of X is denoted as $Sup(X)$ which is mathematically formulated as,

$$Sup(X) = \frac{Sup\ count(X)}{n} * 100 \quad (1)$$

From (1), n is the number if transaction is D. Item set X is termed as a frequent item set when it satisfies the following condition

$$Sup\ (X) > SUPmin \quad (2)$$

Where $SUPmin$ indicates the Minimum Support Threshold (i.e. predefined threshold). The Confidence measure for rule $X \rightarrow Y$ in dataset D is mathematically formulated as below,

$$Confidence(X \rightarrow Y) = \frac{Sup(XY*100)}{Sup(X)} \quad (3)$$

SIPRP method using the rule generation algorithm for generating the association rule is the algorithmic process is described as follows

Input: Database 'D', set of items '$I = \{I1, I2, \dots, Im\}$', Support Value: $Sup(X)$, Confidence Threshold Value: $Confidence(X \rightarrow Y)$

Output: generate sensitive rules

Step 1: **Begin**
Step 2: **For** each Database 'D'
Step 3: **For** each Items 'X
Step 4: measure the support value using (1)
Step 5: measure confidence value using (3)
Step 6:generate the association rule based on support and confidence threshold value
Step 7 **End for**
Step 9: **End for**
Step 10: **End**

Fig. 3: Rule Generation algorithm for generating sensitive rule

As shown in the Figure 3, the rule generation algorithm initially measures the support and confidence value in each item and the database. And then generates the sensitive rules based on the support and confidence values evaluated. After that, SIPRP method evaluates the sensitive rule associated with the optimal sensitive item being concealed to hide the sensitive rules.

III. EXPERIMENTAL SETTING

The Swarm optimization and Iterative Privacy Rule Preservation (SIPRP) method is developed to improve the efficiency of privacy preserving the association rule mining with the constraint minimization. SIPRP method is implemented using the java language. The SIPRP method uses the Adult data set from the University of California Irvine data repository which contains the information about the individuals such as age, level of education and current employment type.

The adult dataset consists of forty nine thousand records and also binomial label that represents the salary of less or greater than fifty thousand US dollars, referred to as

<50K or >50K in SIPRP method. The adult dataset has been divided into a training dataset and test dataset for conducting the experimental work. Training dataset comprises of thirty two thousand records and a test dataset comprises of sixteen thousand records. There are fourteen attributes consisting of seven polynomials, one binomial and six continuous attributes and are used in the SIPRP method to preserve the privacy of certain attributes including the salary, relationship and marital status. The employment class attribute denotes the employer type (i.e. self-employed or federal) and occupation refers to the employment type (i.e. farming or managerial). The education attribute include of high school graduate or doctorate. The relationship attribute includes the information related to unmarried or married.

IV. DISCUSSION

In this section, the result analysis of SIPRP method is evaluated. The performance of SIPRP method is compared with the exiting two methods namely, protocol for secure mining of association rule [1], corporate privacy-preserving framework [2]. The performance of TFVODT framework is evaluated along with the following metrics.

A. Impact of number of sensitive rules

In SIPRP method, the number of sensitive rule describes the ratio in the number of association rules generated to the given set of items which measured in terms of percentage (%) and mathematically formulated as,

$$\text{Number of sensitive rule} = \frac{\text{number of association rule generated}}{\text{set of items}} * 100 \quad (4)$$

When higher the number of association rule generated, the method is said to be more efficient.

Number of items	Number of sensitive rules (%)		
	SIPRP method	protocol for securing the mining of association rule	corporate privacy-preserving framework
1	86	71	64
2	89	74	67
3	92	77	70
4	95	80	73
5	98	83	77
6	81	66	60
7	83	69	63

Table 1: Tabulation for the Number of sensitive rule

Table 1 represents the ratio in the number of sensitive rule generated with respect to the different number of items and the comparison is made with the two existing methods, namely protocol for secure mining of association rule [1], corporate privacy-preserving framework [2]. From the table value, it is clear that the proposed SIPRP method increases the number of sensitive rule generated than the other state-of-art methods.

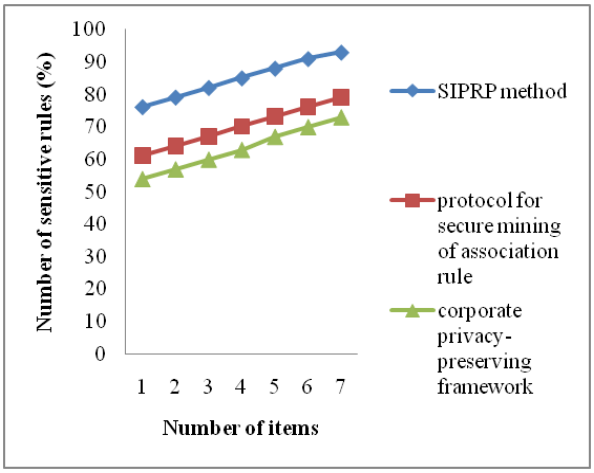

Fig. 4: Measure of the Number of sensitive rules

Figure 4 shows the impact of the number of sensitive rule generated with respect to varying number of items in the range of 1 to 7 using the SIPRP method, protocol for secure mining of the association rule [1], corporate privacy-preserving framework [2]. As illustrated in the Figure, the proposed SIPRP method performs relatively well when compared to the two other existing methods. This is because of the application of the Association rule mining technique in SIPRP method that generates the association rule based on their support and confidence threshold value. Then, SIPRP method evaluates the sensitive rules associated with the optimal sensitive items being hidden for preserving the privacy rate. Therefore, the number of sensitive rule generated using the SIPRP method is improved by 18% as compared to the protocol for securing the mining of association rule [1] and 25% as compared to the corporate privacy-preserving framework [2] respectively

V. CONCLUSION

In the paper, an effective novel framework is designed. It is called as Swarm optimization and Iterative Privacy Rule Preservation (SIPRP) method. SIPRP method is developed to improve the efficiency of the privacy preserving association rule mining with the constraint minimization.

REFERENCES

[1] Tamir Tassa, "Secure Mining of Association Rules in Horizontally Distributed Databases", IEEE Transactions on Knowledge and Data Engineering, Vol. 26, No. 4, APRIL 2014

[2] Fosca Giannotti, Laks V. S. Lakshmanan, Anna Monreale, Dino Pedreschi, and Hui (Wendy) Wang, "Privacy-Preserving Mining of Association Rules From Outsourced Transaction Databases" IEEE Systems Journal, Vol. 7, No. 3, September 2013

[3] Dimitrios Karapiperis and Vassilios S. Verykios, "An LSH-Based Blocking Approach with a Homomorphic Matching Technique for Privacy-Preserving Record Linkage", IEEE Transactions on Knowledge and Data Engineering, Volume 27, Issue 4, April 2015, Pages 909-921.

[4] Leopoldo Bertossi and Lechen Li, "Achieving Data Privacy through Secrecy Views and Null-Based Virtual

Updates", IEEE Transactions on Knowledge and Data Engineering, Volume 25, Issue 5, May 2013, Pages 987-1000.

[5] Sara Hajian and Josep Domingo-Ferrer, "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining", IEEE Transactions on Knowledge and Data Engineering, Volume 25, Issue 7, July 2013, Pages 1445-1459.

