HIDING THE SENSITIVE RULES WITH MINIMUM CONSTRAINTS OF DATA PUBLISHING IN SIPRP METHOD

Dr. P. Tamil Selvan [1], *Dr. S. Veni* [2]

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 effective data publishing. Most of the research work is developed for data distribution with privacy. However, the protocols used in the homomorphic encryption 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. Then, the sensitive rules associated with the optimal sensitive items which are hidden are evaluated. After that, the sensitive rules are subjected to the Particle Swarm Optimization (PSO) for hiding and preserving highly confidential privacy rules. Finally, the SIPRP method obtains the sensitive sets of items for generating the specific sensitive rules. It is hidden with less effect on the privacy being exposed during the data distribution across multiple users. Experimental analysis shows that the SIPRP method is able to improve the privacy rate by 10.5% and also increases the number of hidden rule generated by 26.5 %, when compared to the state-of-the-art works.

Keyword : Association rule mining, sensitive data items, sensitive rules.

A. INTRODUCTION

A majority of the examination work is created in the Privacy Preserving Data Mining (PPDM) for concealing the private, classified, or secure data. In order to improve protected mining, the association rule was developed in [1] for giving secured mining association rules utilizing two secure multi-party calculations. Like that, the strategy expanded the computational expenses. Corporate protection saving structure was planned in [2] that presented an Encode/Unscramble (E/D) module to change the customer information before it was conveyed to the server.

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.

[1]Assistant Professor, Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India e-mail: tmselvanin@gmail.com

[2]Associate professor & Head, Dept. of CS, CA & IT, KAHE, Coimbatore , India, e-mail:venikarthik04@gmail.com

A privacy preserving mining scheme was presented[6] for achieving privacy and scalability on a large scale. Efficient clustering was designed [7] with the aim of improving the computational performance and reducing the computational cost through the Fractional Calculus. Hierarchical K-Means Clustering [8] was applied on the horizontally partitioned data with the objective to improve the communication cost.

In this 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.

B. Design of SIPRP method

The design of Swarm Optimization and Iterative Privacy Rule Preservation (SIPRP) methodology is represented in an exceedingly detailed manner in this section. The major goal of the SIPRP methodology is to cover the sensitive rule type and the general public aiming at the privacy rate. Initially, the SIPRP methodology generatesd the association rule and supported their support and confidence thresholds. The sensitive rules related to the optimum sensitive item are hidden and they are calculable for concealment sensitive rules. SIPRP methodology hides the sensitive rules and mistreatment of the Particle Swarm Optimization (PSO)mechanism with the target of conserving extremely confidential privacy rules.

The SIPRP methodology reduces the constraints which occur, whereas conserving the highly confidential privacy rules through the unvarying generation rules for numerous sensitive sets of things. Finally, the SIPRP methodology ensures the sensitive rules to be hidden with less impact on the privacy that is exposed throughout the data distribution across multiple users. The design diagram of the SIPRP methodology for concealment sensitive rule is shown in the below Figure 1.

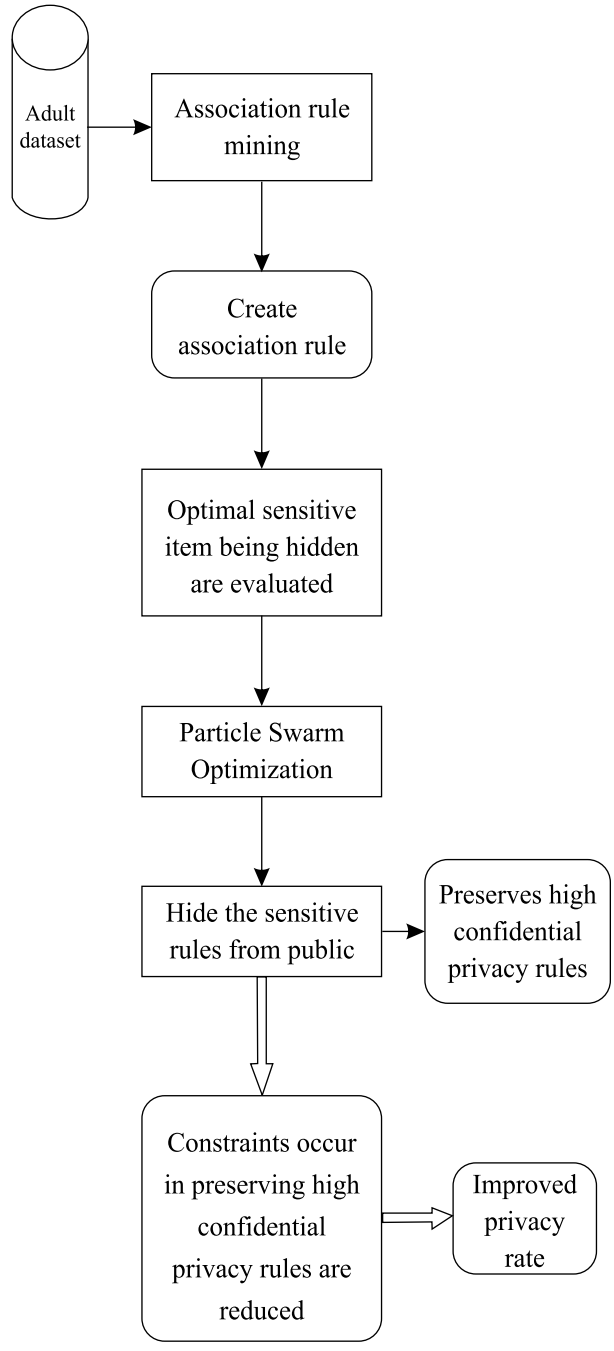

Figure 1 Design of SIPRP technique for sensitive rule thrashing

As shown in the Figure 1, SIPRP methodology, at the start, takes the Adult Data Set as the associate input, then applies the association rule mining for generating the association rule and supports the support and confidence value. Once the association rule is generated, sensitive rule connected with the optimum sensitive item is hidden and evaluated with the target for the privacy rate. Next, the SIPRP methodology hides the sensitive rule type with the help of the general public with the assistance of Particle Swarm Optimization. The PSO mechanism preserves the high confidential privacy rules for reducing the constraints,which occur successively and improves the privacy rate of sensitive rules.

Association rule mining for generating sensitive rules

SIPRP methodology generates the sensitive rules with the association rule mining technique. The Association rule mining technique within the SIPRP methodprotects 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 related to the optimum sensitive things are hidden to preserve the privacy rate. The task of Association rule mining technique within the SIPRP methodology is illustrated within Figure 2 given below.

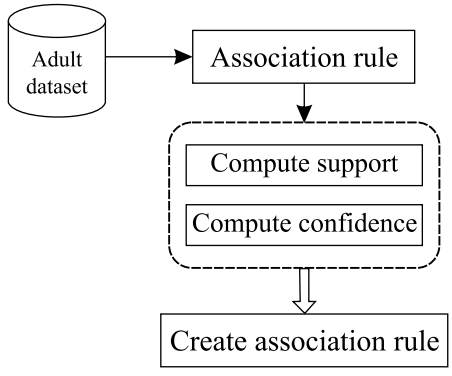

Figure 2 Task of Association rule mining technique in SIPRP method

After performing the association rule, the SIPRP methodology calculates the sensitive rules related to the optimum sensitive things, and it is hidden for providing high confidential privacy rules.

'D' may be information that consists a group of transactions D={T1,T2,…,Tn}and every dealing contains a group of things I={I1,I2,…,Im}. The Association Rule Mining technique recognizes all the association rules X ⇒ Y with a minimum support and confidence value. The support value of an item X ∈ I in the database D is the count of transactions contains X and represented as Sup count(X). Support price of X is denoted as Sup(X) that is mathematically developed as,

$$Sup(X)=(Sup\ count(X))/n*100 \ldots\ldots (1)$$

From (1), n is that the range if dealing is D. Item set X is termed as a frequent item set, once it satisfies the subsequent condition

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

WhereverSUPmin indicates the Minimum Support Threshold (i.e. predefined threshold),the Confidence measure for rule X → Y in dataset D is mathematically formulated as below,

$$Confidence\ (X \rightarrow Y) = (Sup\ (XY*100))/\ (Sup(X)) \ldots\ldots\ldots (3)$$

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

Input: Database '', set of items '', Support Value: , Confidence Threshold Value: X Y
Output: generate sensitive rules
Step 1 : **Begin** Step 2 : **For** each Database '' Step 3 : **For** each Items '

Step 4: Measure the support value using (1)

Step 5: Measure the 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**

The above step showsthat the rule generation algorithm initially measures the support and confidence value in each item and the database. And then it 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.

C. Experimental Performance

The Swarm Optimization and Iterative Privacy Rule Preservation (SIPRP) methodology is developed to enhance the potency of privacy conservationand the association rule mining with the constraint diminution SIPRP methodology is enforced in the java language. The SIPRP methodology uses the Adult DataSet from the University of California Irvine data repository that contains the data informationabout the individuals such as age, level of education and the type of current employment.

The adult dataset consists of 49 thousand records and additionally binomial label that represents the regular payment of more or less than fifty thousand USdollars, brought up as <50K or &>50K in SIPRP methodology. The adult dataset has been divided into a Training dataset and test dataset for conducting the experimental work. Training dataset consists of thirty two thousand records and a test dataset sixteen thousand records. There are fourteen attributes consisting of seven polynomials, one binomial and 6 continuous attributes and area used in the SIPRP method to preserve the privacy of certain attributes including the salary, relationship and marital status.

D. Discussion

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

Impact of a number of sensitive rules

In SIPRP methodology, the quantity of sensitive rule describes the magnitude relation within the number of association rules generated to the given set of things that measured in terms of share (%) and mathematically developed as,

Number of sensitive rule= (Number of association rule generated)/ (set of items)*100….. (4)

The higher the number of association rule generated, more efficientthe method is said to be.

Table 1 shows that the proportion in the quantity of delicate administer created as for the distinctive number of things and the examination is made with the two existing techniques, namely protocol for secure mining of association rule [1] and corporate privacy-preserving framework [2]. From the table value, it is clear that more than the other state- of-art methods, the proposed SIPRP method increases the number of sensitive rule generated

Table 1 Tabulation for the Number of sensitive rule

Number of items	Number of sensitive rules (%) SIPRP method	protocol for securing the mining of association rule	corporate privacy-preserving framework
1	65	50	43
2	68	53	46
3	71	56	49
4	74	59	52
5	77	62	56
6	80	65	59
7	82	68	62

Figure 3 demonstrates the Number of hidden rules versus different number of sensitive rule input using the three different methods. As illustrated in the Figure, the proposed SIPRP method performs well, when compared to the two other methods protocol to secure the mining of association rule [1] and corporate privacy-preserving framework [2]. This is because of the application of the Particle Swarm Optimization mechanism in the SIPRP method. With the help of PSO mechanism, SIPRP method computes the position and velocity of each particle with the aim of hiding the sensitive rules from the public. Based on the computed value such as position and velocity, SIPRP method significantly hides the sensitive rules from the public. As a result, the Number of hidden rules using SIPRP method is improved by 18%, as compared to the protocol for secure mining of association rule [1] and 35% as compared to corporate privacy-preserving

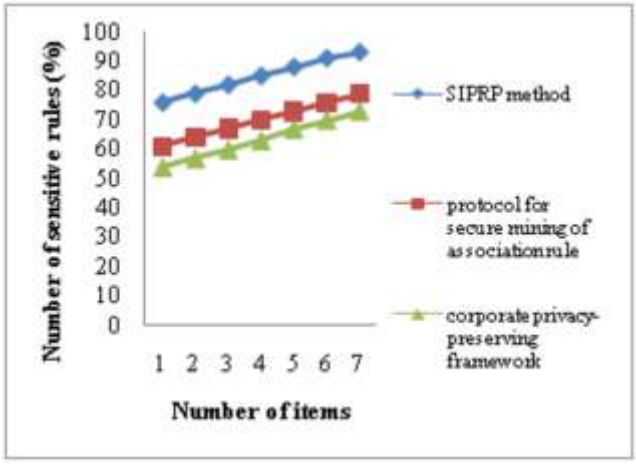

Figure 3 Measure of the Number of sensitive rules

E. Conclusion

In this paper, effectiveand novel framework is composed. It is called Swarm Optimization and Iterative Privacy Rule Preservation (SIPRP) technique. SIPRP method is developed to improve the efficiency of the privacy preserving association rule mining with constraint minimization.The proposed SIPRP method reduces the constraints which arise in preserving the highly confidential privacy rules through the iterative generations of the rules for different sensitive sets of items. The results show that the SIPRP method provides better performance (by 26.5%) with the improvement of a number of hidden rules generated when compared to the state of the art works.

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] DimitriosKarapiperis 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.

[6] Fosca Giannotti, Laks V. S. Lakshmanan, Anna Monreale, DiIssuePedreschi and Hui (Wendy) Wang, "Privacy-Preserving Mining of Association Rules From Outsourced Transaction Databases", IEEE Systems Journal, Volume 7, Issue 3, September 2013, Pages 385-395.

[7] Pawan R. Bhaladhare and Devesh C. Jinwala, "A Clustering Approach for the ??-Diversity Model in Privacy Preserving Data Mining Using Fractional Calculus-Bacterial Foraging Optimization Algorithm", Advances in Computer Engineering, September 2014 , Pages 1-13.

[8] AnrongXue, Dongjie Jiang, Shiguang Ju, Weihe Chen, and Handa Ma, "Privacy-Preserving Hierarchical-k-means Clustering on Horizontally Partitioned Data", International Journal of Distributed Sensor Networks, Volume 5, Issue 1, 2009, Pages 81 - 82.