ANALYSIS OF PRIVACY PRESERVATION IN POPULATION CENSUS DATA PUBLISHING

Dr. P. Tamil Selvan[1]

ABSTRACT

The Privacy Preserving Data Mining (PPDM) is used to protect the sensitive information affecting the privacy of individuals. Hiding the sensitive rules of a transactional database in data mining technique provides the confidentiality of both organizations and individuals. The PPDP requires guarantee for hiding sensitive items in an efficient manner. The property of hiding the sensitive item process is employed to minimize the side effects and provide higher data utility. Privacy information includes personal information in business-like social security numbers, home address, credit card numbers, credit ratings, purchasing behavior, medical records and best-selling services. Some of the privacy preserving techniques do not ensure the sensitive information-hiding process. Hence PPDM is implemented for providing several applications such as population census, homeland security, medical database mining and customer transaction analysis.

Keywords - Privacy preserving data mining, OSA-SIH, SIPRP, RSA-DSO, privacy preservation accuracy.

I. INTRODUCTION

Initially, Optimized Social Ant Based Sensitive Item Hiding (OSA-SIH) technique is introduced with the aim of ensuring higher quality privacy preservation with optimal side effects on the original dataset. Efficient population census process is developed using proposed OSA-SIH technique with the minimum side effects for data-publishing. At first, social ant based relative item set distribution is employed to identify the sensitive items in the given distributed dataset. By using identified dataset, optimal hiding of sensitive items is obtained, based on the social ant based relative item set distribution. Next, sensitive item hiding is executed during the multiplicative and transformational data perturbation process in an effective manner. Thus, the data perturbation mainly depends upon socially cohesive relational rate between sensitive and non-sensitive item sets for providing higher privacy preservation accuracy. Lastly, the side effects on the modified dataset are verified for the multiple users' requested item set distribution, which helps in achieving the accuracy on population census process.

Next, Swarm optimization and Iterative Privacy Rule Preservation (SIPRP) method is proposed with the objective of improving the efficiency of privacy preserving association rule mining in population census process. In the proposed SIPRP method, the sensitive rules associated with the optimal sensitive items being hidden are evaluated. Then, the sensitive rules are given as input to Particle Swarm Optimization

[1] Assistant Professor, Department of S, CA & IT, Karpagam Academy of Higher Education, tmselvanin@gmail.com

(PSO) for hiding and protecting the confidential privacy rules.

Finally, an integrated Reinforced Social Ant and Discrete Swarm Optimization (RSA-DSO) model is proposed to solve multi-objective factor of data hiding and rule hiding, which helps in achieving the accuracy on population census process. In the proposed RSA-DSO model, both Reinforced Social Ant and Discrete Swarm Optimization perform with the same particles. Initially, Reinforced Social Ant (RSA) model is designed to ensure that the sensitive data item hiding is executed through the proposed RSA-DSO framework. Next, Discrete Swarm Optimization (DSO) model is introduced to identify the sensitive rules and further hiding for enhancing the accuracy of Privacy preserving data publishing. Figure 1 shows the process of data-publishing during the population census process using the proposed methods.

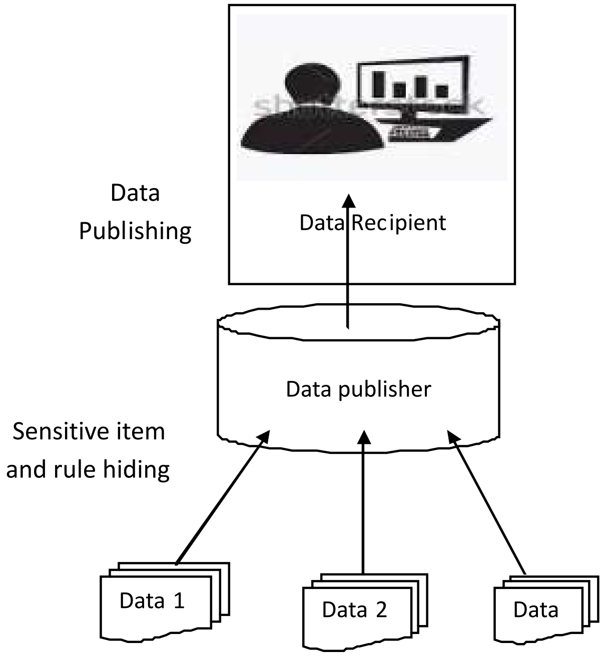

Figure :1 Processing diagram of the data hiding and publishing

As shown in the figure 1, a population census process includes hiding, organizing, analyzing and publishing at a specified time. During the census, the privacy about the individual's personal information is essential for a country to effectively plan growth and deliver services. The information taken for the individuals comprises the adult dataset such as age, gender, country, personal income, marital status, education, work class, occupation, etc. In the figure, the data includes sensitive items and rules for hiding purpose. These kinds of data are needed to be secured while distributing the data to recipient.

2. PERFORMANCE ANALYSIS FOR THE PROPOSED OSA-SIH, SIPRP AND RSA-DSO METHODS

Experiments are conducted using Java language for the proposed OSA-SIH, SIPRP, RSA-DSO methods and existing methods namely Multilevel Trust in Privacy Preserving Data Mining (MLT-PPDM) developed by Yaping Li etal.(2012) and Protocol for secure mining of association rule developed by TamirTassa (2014). The improvement of privacy preservation is obtained using adult data set. The adult data set is taken from the University of California Irvine data repository. It is also called as "Census Income" dataset. The income dataset includes 14 attributes and the number of instances is 48842. The attributes are age, work class, education, marital status, occupation, native country, level, current employment type and so on.

Impact of privacy-preservation accuracy

Privacy-preservation accuracy helps in measuring the number of privacy preserved perturbed copies in PPDM with respect to the total number of perturbed copies considered for data-publishing. The privacy preservation accuracy is mathematically formulated as given below:

$$PPA = \left(\frac{\text{No. of privacy preserved pertubed copies}}{\text{Total Number of pertubed copies}} \right) 100 \quad (1)$$

From the equation (1), privacy preservation accuracy" is measured by percentage (%). If the privacy preservation accuracy is high, then the method is said to be more efficient.

Table1-Tabulation of privacy preservation accuracy

Privacy Preservation Accuracy (%)					
Age (Number of perturbed copies)	**Existing MLT - PPDM**	**Existing protocol for secure mining association rule of**	**Proposed OSA-SIH**	**Proposed SIPRP**	**Proposed RSA-DSO**
10	69.48	60.36	75.35	79.02	82.28
20	75.48	65.45	79.41	83.14	86.52
30	77.54	67.51	80.39	84.32	88.58
40	76.26	66.98	78.28	82.41	87.3
50	75.96	70.93	81.97	85.34	89.32
60	82.43	73.43	85.32	89.26	93.35
70	85.44	75.54	89.36	92.12	95.48
80	88.17	78.37	92.14	95.48	98.32
90	91.11	81.43	94.53	97.22	99.28
100	92.84	83.94	96.26	98.52	97.12

Table 1 illustrates the privacy preservation accuracy based on the number of perturbed copies for data publishing using the proposed OSA-SIH, SIPRP, RSA-DSO methods and the existing MLT-PPDM Protocol for secure mining of association rule methods. The number of perturbed copies is taken from the range of 10 to 100 for conducting the experiment. When increasing the number of perturbed copies, privacy preservation accuracy is also increased in all the methods. However, table 1 shows that the proposed RSA-DSO model effectively achieves a higher privacy preservation accuracy, when compared to the proposed and other existing methods.

Figure 2 shows the measure of privacy preservation accuracy using the proposed OSA-SIH, SIPRP, RSA-DSO methods and existing methods such as MLT-PPDM developed by Yaping Li et al. (2012) and Protocol for secure mining of association rules developed by TamirTassa (2014).

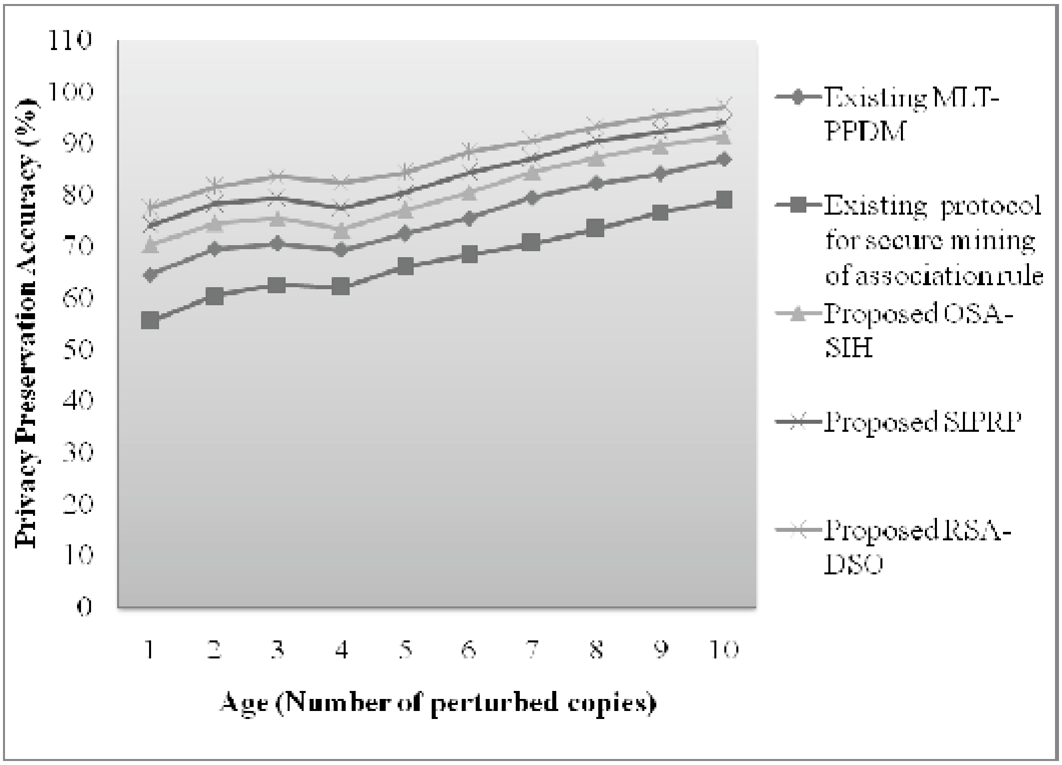

Figure 2 Measure of the privacy preservation accuracy

From figure 2, it is clear that the proposed RSA-DSO model efficiently increases the privacy preservation accuracy, when compared to the proposed and other existing methods. This privacy-preservation enhancement is achieved in the proposed RSA-DSO model with the help of discrete swarm optimization algorithm in the population census data-publishing process. In addition, the discrete swarm optimization algorithm utilizes the fitness function for extracting the sensitive rule in a significant manner, which in turn increases the privacy-preservation accuracy. Therefore, the proposed RSA-DSO model increases the privacy-preservation accuracy by 19%, when compared to the existing MLT-PPDM and Protocol for the secure mining of association rule methods. The proposed SIPRP method increases the privacy preservation accuracy by 15% and the proposed OSA-SIH technique increases the privacy preservation accuracy by 11%, when compared to the existing MLT-PPDM and Protocol for secure mining of association rule methods.

3. CONCLUSION

The proposed OSA-SIH, SIPRP, RSA-DSO methods are successfully compared with the existing methods such as Multilevel Trust in the Privacy Preserving Data Mining (MLT-PPDM) developed by Yaping Li etal. (2012) and the Protocol for secure mining of association rule developed by TamirTassa (2014) and provide secure population census data publishing process. Experimental results clearly show that the rate of side effects using the proposed OSA-SIH technique is reduced by applying the side effects on the modified dataset. Then, the number of hidden rules is improved in the proposed SIPRP method with the help of particle swarm optimization mechanism. Privacy preservation accuracy and the number of sensitive rules are improved in the proposed RSA-DSO model by using the discrete swarm optimization algorithm.

4. REFERENCES

1. TamirTassa., "Secure Mining of Association Rules in Horizontally Distributed Databases", IEEE Transactions on Knowledge and Data Engineering, Volume 26, Issue 4, April 2014, Pages 970 - 983.

2. Yaping Li, Minghua Chen, Qiwei Li, and Wei Zhang, "Enabling Multilevel Trust in Privacy Preserving Data Mining", IEEE Transactions on Knowledge and Data Engineering, Volume 24, Issue 9, September 2012, Pages 1598 - 1612.

3. ZhuoHao., Sheng Zhong., and Nenghai Yu., "A Privacy-Preserving Remote Data Integrity Checking Protocol with Data Dynamics and Public Verifiability" IEEE Transactions on Knowledge and Data Engineering, Volume 23, Issue 9, September 2011, Pages 1432 - 1437

4. R.J. Kuoa,C.M. Chao and Y.T. Chiu, "Application of particle swarm optimization to association rule mining", ELSEVIER: Applied Soft Computing, Volume 11, Issue 1, January 2011, Pages 326-336.

5. R. Mahesh and Meyyappan, "New Method for Preserving Privacy in Data Publishing Against Attribute and Identity Disclosure Risk", International Journal on Cryptography and Information Security (IJCIS), Volume 3, Issue 2, June 2013, Pages 261-266

6. Giannotti F, Lakshmanan VS, Monreale A, Pedreschi D, Wang HW. Privacy-preserving mining of association rules from outsourced transaction databases. IEEE Systems Journal. 2013 Sep; 7(3):385-95.

7. Tassa T. Secure mining of association rules in horizontally distributed databases. IEEE Transactions on Knowledge and Data Engineering. 2014 Apr; 26(4):970-83.

8. Squicciarini AC, Lin D, Sundareswaran S, Wede J. Privacy policy inference of user-uploaded images on content sharing sites. IEEE Transactions on Knowledge and Data Engineering. 2015 Jan 1; 27(1):193-206.

9. Paulet R, Md Kaosar G, Yi X, Bertino E. Privacy-preserving and content-protecting location based queries. IEEE Transactions on Knowledge and Data Engineering. 2014 May; 26(5):1200-10.

10. Pervaiz Z, Walid G, Ghafoor A, Prabhu N. Accuracyconstrained privacy-preserving access control mechanism for relational data. IEEE Transactions on Knowledge and Data Engineering. 2014 Apr; 26(4):795-807.