International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 10, October-2019

A Survey on Sensitive Association Rule Hiding for Privacy Evaluation of Methods and Metrics

M. Rajasekaran
Assistant Professor
Department of CSE
Kamaraj College of Engg & Tech
Madurai – 625701

Dr. M. S. Thanabal
Professor
Department of CSE
PSNA College of Engg & Tech
Dindigul – 624004

***Abstract*— Privacy preserving or Data and knowledge hiding is a novel research area in distributed collaborative data mining to protect the privacy of confidential or sensitive information of individuals. Many of the researchers have been proposed methods in Privacy Preserving Data mining (PPDM) to hide sensitive information in association rule mining. Association rule hiding is the process to modify the original database for vanishing sensitive association rule while generating rules using rule mining algorithms. The better rule hiding methods are not affecting the quality of the database and nonsensitive rules. In this paper, privacy preserving association rule hiding methods in the literature are studied in detail to find the problem in each method and metrics used for evaluating these methods. The performance of metrics, merit and demerit of every method are thoroughly compared. Finally, the remarkable future direction is suggested in the association rule hiding area based on the problems has been found from the literature.**

Keywords: ***Privacy Preserving Data Mining, Sensitive Itemset Hiding, Association Rule Hiding***

1. INTRODUCTION

In state-of-the-art scenario of internet world, the security and privacy has to be given more attention while information sharing among organizations. The association rule mining discovers the relationship among itemsets which helps to improve the business of enterprises. The sharing of relationship among itemsets between organizations helps them to acquire extraordinary business knowledge. However, information of sensitive itemsets or items to be preserved while sharing between organizations. The methods for hiding of sensitive items from frequent itemset or associations rules were proposed by many authors in the literature. The reducing support value of sensitive item Atallah et al. [1] was the general idea for providing security and privacy of itemsets in association rule mining. The sensitive item hiding was achieved by the modifying support value of items to decrease below the minimum support values. However, it was considered as a very basic solution for privacy preserving, because the data loss and spurious rule generations were not considered. Also the items inappropriately hidden can be simply be recovered through the use of inference channels. Algorithm parameters like computational speed, the number of arithmetic operations were only for privacy preserving Bertino et al. [2]. The expectation was that the computation time for hiding sensitive items proportionally related with strength of hiding. The metric for maintaining data quality and selecting sensitive items for retaining quality were not discussed.

Some articles in the literature focused only on the metrics and their efficiency for measuring the quality of the different privacy preserving techniques Fletcher et al.[3]. The metrics to measure the quality of classification and clustering was available more but lacked the metrics used in association rule mining algorithms. Some other articles in the literatures reviewed combines methods and metrics. Mendes and Vilela [4] mainly conducted the survey on methods and metrics used especially in association rule hiding. However, the thorough comparison between various methods and model were not presented and only listed out the metrics used for measuring the quality of rule hiding and did not elaborate about metrics.

The main objective of this work is to study and understand the sensitive item hiding methodologies of association rules without affecting the quality of original database as fewer as possible. Additionally, the sensitive knowledge is obtained by intended recipients through various intentional attacks will be studied from some article. Even though several privacy preserving association rules hiding methods proposed to protect privacy, privacy preserving methods become a hot directive in security of data mining.

2. SURVEY ON SENSITIVE RULE HIDING TECHNIUQES

The accuracy of frequent itemset mining was reduced in the conventional privacy protection techniques, because the different items were provided the same level of privacy protection. Sun,C et al., [5] proposed a personalized privacy protection method which provided different level of privacy protection to different attributes or items. The randomized response technique was used in this method. The provision of different privacy level to different item offered the accuracy of frequent itemset mining.

Domadiya & Rao [6] proposed a hybrid method for privacy preserving association rule mining. The items from the transaction were selected and modified based on two heuristic such as reduce support and reduce confidence of sensitive rules. The total quantity of modification required in transaction to hide any sensitive rule was calculated based on the relation of support and confidence. The hybridization of two heuristics balanced the quality of database while preserved the privacy.

Chenga et al.,[7] used Evolutionary Multi-objective Optimization (EMO). The multi objective functions were proposed to reduce various side effects occurred when tradeoff accuracy and privacy. The selected items are

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 10, October-2019

removed to hide sensitive rules. The minimization of missing normal rules, spurious or ghost rules and data loss were the multi objective of EMO. The objective functions were used to select appropriate itemsets of transactions for modification with minimal side effects.

Telikani & Shahbahrami [8] proposed a method to hide sensitive association rules by hybridized border and heuristic approaches. In literature, the border approaches were almost used to hide sensitive items in frequent itemset only, where as heuristic method has been utilized to hide sensitive items in both association rules and in frequent itemsets. The author in this paper utilized the advantage of both border and heuristic methods to hide sensitive items. The border based solution was improved by two heuristic methods. The first heuristic utilized maxmin solutions to select victim items to hide, then the second heuristic removed victim items from transactions.

Wu,J,M et al.,[9] proposed an ant colony system (ACS) to reduce side effects and enhance the performance of the sanitization process. The every ant from the populated solutions randomly deleted the transaction in the original database in each iteration. In each iteration the deleted position of the ant in the database was updated towards global best solution. Some termination conditions were introduced in this paper to stop the sanitization processes. The proposed ACS based heuristic function adjusted the degree of hiding with respect to the side effects obtained.

Bux,N,K et al. [10] proposed a Genetic Algorithm (GA) based method for hiding sensitive items with the objective of minimizing side effects and iteration. The objective functions were applied recursively to achieve better time cost and control the side effects. In this method, the set of rules were hidden in one optimization run instead of hiding one rule in every run. The decrease the confidence rule was chosen to select the victim rules.

Krishnamoorthy,S et al. [11] proposed a Particle Swarm Optimization (PSO) based method to hide a set of sensitive patterns with minimum side effects like lost rules, ghost rules, and number of modifications. In this method, the number of modified entries was reduces by modifying transaction in certain order to keep up the accuracy and security. The PSO based approach is scalable in terms of database size, because PSO clustered the transaction before hiding sensitive patterns.

Peng,M et al., [12] proposed a privacy preserving association rule hiding over the heath data. The degree of items association was reduced by random permutation and random deletion. Because the random permutation, the confidence of rules fell down considerably even though the support of item was unchanged. The data quality was mainly retained by confidence of the rule than support of the item, because confidence was more sensitive then support.

Taleb & Dehkordi [13] proposed an Electromagnetic Field Optimization Algorithm (EFOA) to hide sensitive items. This EFOA used the data distortion method with two objective functions to arrive at the solution with the minimum side effects. This algorithm first obtained local optimal solution for hiding sensitive rule with objectives then moved toward global optimal solutions. The global solution was updated from the local solutions in each iteration reduced considerable amount of the runtime of this algorithm.

Domadiyaa & Rao [14] proposed a privacy preserving collaborative association rule mining on vertically partitioned healthcare datasets. Since the dataset was vertically partitioned the privacy of each stored portion was maintained without knowing the item details of other portions. Also the communication cost of transmitting item details among the stored portioned was reduced by only discovering the items which were correlated to diseases without disclosing patients information .

Baby & Chandra [15] proposed a homomorphic cryptography based scheme for privacy preserving data mining. The main problem in the well known privacy preserving methods like high computational complexity and large communication cost were reduced in this method while providing perfect secrecy and defy various attacks to some extent in association rule mining process.

Mohan & Angamuthu [16] proposed a method to hide sensitive items based on creating and inserting dummy item in the dataset, then hide items using Genetic Algortim(GA). The created dummy items were used to maintain the same cost of original and modified database after removing sensitive items. The victim items for modification was selected optimally based on the of chromosome of GA. The best item selection was achieved by regeneration of chromosome population based on cross over and mutation operations.

Surendra H & Mohan H S [17] proposed a method for hiding sensitive itemsets based on patterns obtained in the Closed Itemsets. The novelty of this work was hiding sensitive items in the Closed Itemsets instead of items in the transaction. This method was best suitable for privacy preserved pattern sharing applications. The multiple sensitive itemsets were hidden recursively irrespective of the size and support of them. The itemsets after hidden of sensitive items satisfied the closeness property and supported for frequent itemset mining and association rule generations. The non-sensitive itemsets were unchanged during the hiding process.

Table 2.1 illustrates an overview of merits, demerits and performance metrics of above discussed load balancing techniques.

Table 2.1 Comparison of load balancing techniques

Ref. No.	Methods Used	Merits	Demerits	Performance Metrics
[5]	randomized response technique	different attributes makes higher accuracy	computation and communication cost is medium	Dataset : T3.I4.D500K.N10. minimum support : 0.009

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 8 Issue 10, October-2019

Ref. No.	Methods Used	Merits	Demerits	Performance Metrics
[6]	Decrease Support and Decrease Confidence of sensitive rule for selection and modification of items	maintain the quality of the database while improving the privacy of database.	missing cost (MC) and side effect factor (SEF) are to be considered	Minimum confidence threshold = 63.89% Minimum Support Threshold:20% HF (hiding failure)- MDSRRC: 0
[7]	Evolutionary multi-objective optimization (EMO) was used to find an appropriate subset of transactions for rule hiding	make several hiding solution in a single run	more sensitive rules and higher support or confidence levels lead to a longer running time	Dataset :Mushroom Missing Rule - SIF-IDF: 7.867 % Missing Rule- WSDA: 2.578% Missing Rule- EMO : 2.574%
[8]	MaxMin and Decrease Confidence rule	Accuracy is decreased below the confidence threshold values	The optimization of victim item selection strategy to be utilized for further reducing side effects	Dataset: Mushroom Sensitive Rule:40% DCR - Data utility : 97.1 % ARHIL - Data utility :95.4 %

Ref. No.	Methods Used	Merits	Demerits	Performance Metrics
[9]	Ant colony system (ACS) decrease side effects	pre-large method to observe side effects and computes the degree of hiding items to change the selecting policy	Require to calculate the value of the heuristic function for all candidate transactions.	Dataset :Mushroom Minimum Support :45% Fail to be hidden FTH: PSO2DT -FTH:7
[10]	Simple genetic with objective of estimating the effect of non sensitive rule	provide recursive computation to reduce the time	Hiding one rule in every run.	Dataset:Mushroom Minimum Confidence:60% CPUTime- DCR: 4000s CPUTime –EARH-GA :400s Data utility-DCR :95% Data utility –EARH-GA :97% Accuracy- DCR :92% Accuracy-EARH-GA:94%

3. CONCLUSION

In this paper, many papers regarding privacy preserving has been analyzed and studied their performance and demerits. The demerits and performance of metric provide future direction to provide new solutions to solve demerits and further improve the performance of metrics. The detailed study of the existing algorithms concludes that the selection of appropriate minimum thresholds, rule pruning methods, multiple objectives and stopping criterions are required to provide the privacy preserving methods to handle streams of datas from Internet of things clouds.

REFERENCES

[1] M. Atallah, E. Bertino, A. Elmagarmid, M. Ibrahim, and V. S. Verykios.(1999).Disclosure limitation of sensitive rules. In Proceedings of the 1999 IEEE Knowledge and Data Engineering Exchange Workshop (KDEX'99), pages 45–52.

[2] E. Bertino, I. N. Fovino, and L. P. Povenza. (2005) A framework for evaluating privacy preserving data mining algorithms. Data Mining and Knowledge Discovery, 11(2):121–154.

[3] S. Fletcher and M. Z. Islam, Measuring information quality for privacy preserving data mining, Int. J. Comput. Theor. Eng., vol. 7, no. 1,pp. 21-28, May 2015.

[4] R. Mendes and J. P. Vilela,.(2017),Privacy-preserving data mining: Methods, metrics, and applications, IEEE Access, vol. 5, pp. 10562-10582.

[5] Sun C.,Fu Y., Zhou J., & Gao ,H., (2014). Personalized Privacy-Preserving Frequent Itemset Mining Using Randomized Response. The Scientific World Journal, vol. 2014, Article ID 686151, 10 pages.

[6] Domadiya N.H., Rao U.P., (2016) A Hybrid Technique for Hiding Sensitive Association Rules and Maintaining Database Quality. (eds) Proceedings of First International Conference on Information and Communication Technology for Intelligent Systems: Volume 2. Smart Innovation, Systems and Technologies, vol 51. Springer, Cham

[7] Cheng, P., Lee, I., Lin, C., & Pan, J. (2014). Association rule hiding based on evolutionary multi-objective optimization. Intell. Data Anal., 20, pp. 495-514.

[8] Telikani, A. & Shahbahrami, A. (2017). Optimizing association rule hiding using combination of border and heuristic approaches. Applied Intelligence:Springer. Volume 47, Issue 2, pp 544–557.

[9] Wu, J. M.,Zhan,J., & Lin, J,C., (2017) Ant Colony System Sanitization Approach to Hiding Sensitive Itemsets, IEEE Access, vol. 5, pp. 10024-10039.

[10] Bux,N,K., Lu,M., Wang,J., Hussain,S., & Aljeroudi,Y(2018).Efficient Association Rules Hiding Using Genetic Algorithms. Symmetry 2018, 10(11), pp.576.