Privacy Preserving Association Rule Hiding Techniques: Current Research Challenges

Mohamed Refaat Abdellah
Military Technical College

H. Aboelseoud M.
Military Technical College

Khalid Shafee Badran
Military Technical College

M. Badr Senousy
Sadat Academy

ABSTRACT

Association rule mining is one of the most used techniques of data mining that are utilized to extract the association rules from large databases. Association rules are one of the most significant assets of any organization that can be used for business growth and profitability increase. It contains sensitive information that threatens the privacy of its publication and it should be hidden before publishing the database. Privacy preserving data mining (PPDM) techniques is used to preserve such confidential information or restrictive patterns from unauthorized access. The pattern can be represented in the form of a frequent itemset or association rule. Also, a rule or pattern is marked as sensitive if its disclosure risk is above a given threshold. This paper discusses the current techniques and challenges of privacy preserving in association rule mining. Also, presentation of metrics used to evaluate the performance of those approaches is also given. Finally, Interesting future trends in this research body are specified.

General Terms

Data mining security, Database security

Keywords

Privacy preserving data mining, Association Rule, Hiding Approaches.

1. INTRODUCTION

Now days, the privacy preserving data mining has become an essential concern due to the rapid growth of electronic data in governments, corporations and different organizations. Such data may implicitly contain sensitive information and can lead to privacy or security threats if they are misused. As the data mining technology has rapidly progressed, getting user's sensitive information through data mining technology has become very easy. This led to increasing concerns about the privacy of the underlying data.

Association rule hiding is a subarea of privacy preserving data mining that studies the side effects of data mining methods that generated from the disclose the sensitive information belong to individuals or organizations. The presented of many extended set of application scenarios in which collected data or knowledge patterns extracted from the data have to be shared with others (possibly not trusted) entities to serve owner or organization particular purposes. The sharing of data or knowledge might done at a cost to privacy, primarily due to two main reasons: (i) if the data refer to individuals, then its disclosure can violate the privacy of the individuals who are recorded in the data. If their identity is discovered to not trust third parties or if sensitive knowledge about them can be mined from the data, and (ii) if the data regard business (or organizations) information, then disclosure of this data or any knowledge mined from the data may potentially reveal sensitive trade secrets, whose knowledge can offer a important advantage to business competitors and thus can cause the data owner to lose business over his or her peers.

The problem of hiding association rule can be considered as a type of database inference control, but its main goal is to protect the sensitive rules not the sensitive data [1] (the violation of privacy is coming from sensitive association rules rather than the data itself). In association rule hiding there will be a set of sensitive association rules, which are specified by the security administrator or data owner, the task of the association rule hiding algorithms is to sanitize the data so that the association rule mining algorithms applied to this data, (i) will be unable to extract the sensitive rules and (ii) can mine all the non-sensitive rules. Several techniques have been used to hide sensitive association rules by doing some changes in the original data set. Due to these changes, some non-sensitive patterns may be lost, called lost rules, and new patterns are also generated, known as ghost rules as explained later.

This paper is organized as follows; background and association rule hiding process are discussed in Section 2 and Section 3, respectively. The association rule hiding approaches and metrics and performance analysis are explained in Section 4 and Section 5, respectively. The challenges and future trends and conclusions are discussed in Section 6 and Section 7, respectively.

2. BACKGROUND

In 1999, Atallah et al. [2], in this first work the authors studied the problem of hiding sensitive frequent patterns and presented a heuristic greedy approach that pass through the frequent itemset pattern for choose the transactions and the items that they had to change, so that the support of a sensitive frequent pattern decreases to fall below the support threshold.

In 2000, Dasseni et al. [3], described the hiding of the sensitive frequent pattern problem as the problem of hiding both sensitive frequent itemsets and sensitive rules.

In 2001, Saygm et al. [4], worked in two directions to hide sensitive association rules first based on reducing the minimum confidence of the rules and the second based on reducing the minimum support of the itemsets that generated these rules.

In 2002, Oliveiraet et al. [5], proposed four algorithms based on the conflict degree of selecting the sensitive transactions to sanitize. They divided into Pattern Restriction Based algorithm (Naïve) and Item Restriction Based algorithms (MinFIA, MaxFIA and IGA). The Naïve algorithm kept only the highest frequency items for the selected transaction in the database. The MinFIA (Minimum Frequency Item Algorithm) removes the smallest support items for the sensitive transactions. The MaxFIA (Maximum Frequency Item Algorithm) removes the maximum support items for the sensitive transactions. The IGA (Item Grouping algorithm) groups same itemsets related to different restricted patterns, so the group of all the sensitive patterns can be hidden in one step.

In 2005, E. Bertino [6], worked on examination and improvement of the algorithm proposed in [4], the performance is examined with different itemsets sizes and different sets of sensitive rules. The results show very good performance and efficiency.

In 2005, S. Wang et al. [7], proposed two hiding algorithms ISL (Increase Support of LHS) and DSR (Decrease Support of RHS) algorithm. ISL through rising support of rules' LHS confidence will be reduced under the threshold; as a result the sensitive association rules will be unseen. DSR decreases the whole rule's support and confidence below the threshold to hide sensitive association rules. DSR has no hiding failure; notwithstanding, whereas ISL will fail if there is no suitable transaction to add.

In 2005, X. Sun [8], proposed a border based approach to efficiently estimate the impact of any modification to the database during the hiding process. The quality of the database can be well preserved by carefully selecting the modifications with minimal side effect.

In 2007, X. Sun [9], improved effectiveness of the previous work in [8] and provided an evaluation of the impact of any modification to the database during the hiding process.

In 2007, S. Wang et al. [10], proposed two algorithms, first is Decrease Confidence by Increase Support (DCIS) and second is Decrease Confidence by Decrease Support (DCDS) to automatically hide collaborative recommendation association rules with no pre mining and choice of hidden rules. In DCDS, the support of the recommended item decreased and the support of non-recommended item remained constant. In DCIS, the support of non-recommended item increased and support of the recommended item remained constant.

In 2008, Kaya [11], introduced a new algorithm to hide critical fuzzy association rules from quantitative data. To do this, the support value of LHS of the rule to be hidden is increased. The algorithm provides reliable rule hiding results.

In 2008, Weng et al. [12], proposed FHSAR (Fast Hiding Sensitive Association Rules) algorithm. The main goal is to reduce the execution time by hiding all sensitive association rules during one database scan. In this algorithm all relationships between the sensitive association rules and each transaction are evaluated to efficiently select the appropriate items to modify.

In 2010, Modi et al. [13], proposed DSRRC (Decrease Support of RHS Items of Rule Cluster) algorithm. The sensitive rules are grouped by similarity on RHS and then start the hiding process. This operation will reduce both amounts of changes in the database and the side effects. This algorithm has three side effects: (i) it does not maintain data quality (ii) large number of transactions on the database (iii) execution time is increased due to sort of the database after each change.

In 2011, Yogendra Kumar et al. [14], introduced an algorithm that increases and decreases the support of the LHS and RHS item of the rule at the same time to hide the rule. The advantage of this algorithm is the minimization of the data modification to hide a set of rules, so it need lower CPU time than the previous work.

In 2012, Komal Shah et al. [15], proposed improved algorithms called ADSRRC (Advanced Decrease Support of RHS items of Rule Cluster) and RRLR (Remove and Reinsert LHS of Rule) to overcome the limitations of DSRRC. ADSRRC, the same as DSSRC, tries to cluster sensitive rules based on similar RHS. This algorithm started with sensitivity calculation of the transactions, and then sorted them in descending order. For this, order of transactions has no impact on algorithm result. RRLR can handle multiple RHS of different association rules. Using of two sorting processes, enhance the speed of these algorithms over the DSRRC.

In 2012, D. Jain et al. [16], introduced an algorithm to hide sensitive association rules without changing the support of frequent itemsets. This algorithm used a new concept named Representative Rule RR, where all sensitive rules can be inferred without any access to the main database. The basic idea of this algorithm to hide association rules by changing the position of items, rather than deleting them from the transactions. The advantage of that no modification is done in frequent itemsets support, size of database, and less modifications in database. There is a need to find out a method, which can avoid the computation of the confidence of the rules that has confidence below the minimum confidence.

In 2013, Domadiya et al. [17], proposed Modified Decrease Support of RHS item of Rule Clusters (MDSRRC) to hide association rules. MDSRRC can hide rules with several LHS and RHS. It starts with the calculation of sensitive rule items according to the RHS and delete items with the highest values. The advantage of MDSRRC over DSRRC that it has less side effects and better data quality due to decrease database modification.

In 2013, Dhutraj et al. [18], proposed a hybrid algorithm for hiding sensitive association rules, it combined both DSR and ISL methods. But the proposed algorithm has bad memory utilization.

In 2014 P. Cheng et al. [19], proposed a Multi-Objective Optimization (EMO) algorithm (a hybrid algorithm uses a genetic algorithm with data distortion algorithm). The proposed approach can effectively hide all sensitive rules while generate fewer side effects. But it suffers from existence of non-sensitive lost rules and the selection of deleting items need more effort.

3. ASSOCIATION RULE HIDING PROCESS

3.1 Problem Description

Association rule hiding problems can be defined as: create a sanitized database from the original database to prevent that data mining techniques from mine sensitive rules from the database and keep the visibility of all non-sensitive rules. A general definition of the problem can be given as:

Given transnational database D, Minimum confidence, Minimum support, and generated set of association rules **R** from D, a subset $\mathbf{R}_{\mathbf{sen}}$ of **R** as sensitive rules, which database owner wants to hide. The Problem is to find the sanitized database D' such that when a mining technique is applied to the D', all sensitive rules in set $\mathbf{R}_{\mathbf{sen}}$ will be hidden while all non-sensitive rules $\mathbf{R}_{\mathbf{non\text{-}sen}}$ can be mined. After applying mining technique on D', the $\mathbf{R}_{\mathbf{non\text{-}sen}}$ will be divided into true association rules and lost rules, $\mathbf{R}_{\mathbf{sen}}$ is also will be divided into a set of sensitive rules that not be hidden ($\mathbf{R}_{\mathbf{non\text{-}Hide}}$) and a set of sensitive rules that will be hidden as shown in Figure 1 [17].

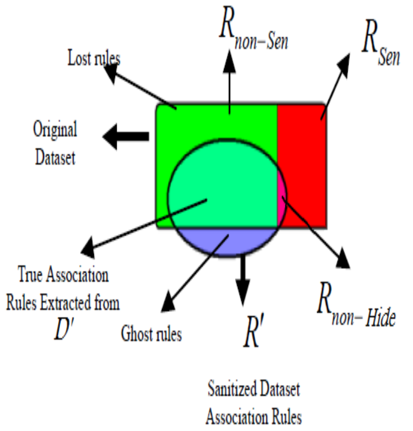
$R_{non-Sen}$
R_{Sen}
Lost rules
Original Dataset
True Association Rules Extracted from D'
Ghost rules
R'
$R_{non-Hide}$
Sanitized Dataset Association Rules

Fig 1: Association rule hiding process

3.2 Side Effects of Privacy Preserving In Association Rule Mining

1. Sensitive rules that not be hidden (**R**$\mathbf{R_{non\text{-}Hide}}$): No sensitive association rules mined from original database can be mined from sanitized database with predefined support and confidence. The sensitive rules may be of a specific form or from the owner's perspective.

2. Non-sensitive rules lost (Lost rules): All the non-sensitive rules that can be mined from the original database with predefined support and confidence, should also be mined from the sanitized database at the same support and confidence level.

3. New rules generated after sanitization (Ghost rules): No rule that was not derived from the original database with predefined support and confidence can be derived from a sanitized database at the same support and confidence level.

The problem is to find an optimized sanitized database, which that minimize or eliminate all these side effects.

4. ASSOCIATION RULE HIDING APPROACHES

In this section, the different associations rule hiding approaches are shown in Figure 2.

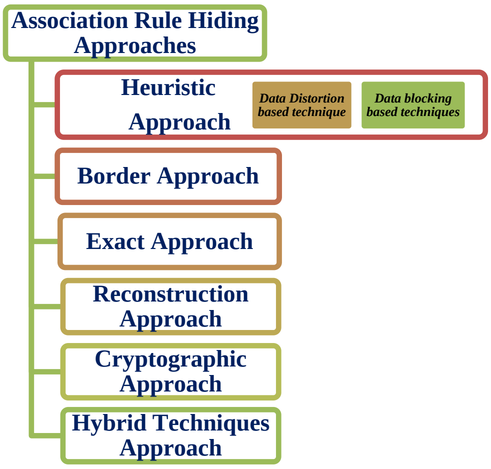
Association Rule Hiding Approaches
Heuristic Approach
Data Distortion based technique
Data blocking based techniques
Border Approach
Exact Approach
Reconstruction Approach
Cryptographic Approach
Hybrid Techniques Approach

Fig 2: Association rule hiding approaches

4.1 Heuristic Approach

4.1.1 *Data Distortion technique*

Data Distortion technique is a technique for modifying data using a random process. This technique apparently distorts sensitive data values by adding noise, data transpose matrix, or adding unknown values etc. [20]. This technique can handle different data types: character, Boolean, classification and integer. Discrete data need original data set to be processed. The processing of data is classified into attribute coding and obtaining sets coded data set [21].

In most of the distortion techniques it is very difficult to preserve the original data.

4.1.2 *Data blocking techniques*

Blocking method works by reduction of the degree of support and confidence of sensitive association rules and replacing some attribute values of data items with unknown values (?) or replace '1' by '0' or '0' by '1'. In this technique preserving privacy is done in two steps. First is to recognize transactions of sensitive rule and second is to replace the known values to the unknown, so the support of certain items goes down to a certain level and rule mining algorithm not able to mine the sensitive rules [22]. One problem with block-based privacy preserving association rule mining is that it is too hard to calculate the support and confidence of a sensitive association rule since the some of the original data is replaced with unknown value [23], [24]. This can be solved by using uncertain symbols which then can be restored with actual support and confidence [25].

4.2 Border Approach

The process of border revision is introduced by X. Sun [8], The authors propose a heuristic approach that uses the notion of the border (improve effectiveness of the previous work in [9]) of the non-sensitive frequent itemsets to track the impact of altering transactions in the database. The proposed scheme first computes the positive and the negative borders in the lattice of all itemsets and then focuses on preserving the quality of the computed borders during the hiding process. The quality of database can be well maintained by greedily selecting the changes with minimal side effect.

In the proposed heuristic, a weight is assigned to each element of the calculated positive border to quantitative the effect of deleting an item. During the sanitization process these weights are dynamically computed according to the current support of the equivalent itemsets in the database. To reduce the support of a sensitive itemset from the negative border, the algorithm calculates the effect of the possible item deletions by calculating the sum of the weights of the positive border elements that will be affected. Then, it proceeds to delete the items that will have the minimal impact on the positive border.

In [26], authors improves the hiding solutions of [8], The proposed algorithms follow a similar approach and try to modify this item in such a way that the support of the max-min itemset is minimally affected. In case of multiple itemsets the hiding process starts with lower support itemset at one at a time base.

4.3 Exact Approach

Exact approaches are normally able to offer better quality solutions compared to the heuristic approaches, but with a high complexity cost. This is coming through represent the sanitization process as a constraint satisfaction problem and by solving it using linear or integer programming solver. The sanitization process is done as an atomic operation to prevent the local minima experienced by the heuristic approaches.

It solves the problem as a Constraint Satisfaction Problem (CSP) with a goal to discover the minimum number of transactions that need to be sanitized for the suitable hiding of all the sensitive knowledge. It works with the sensitive itemsets only to reduce the problem size, apply for their support stays lower than the minimum support threshold. The optimization process is determined by a standard measure function that is glorious by the measure of accuracy. Moreover, the constraints obligatory in the CSP formulation catch the number of supporting transactions that need to be sanitized for the hiding of each sensitive itemset. The best solution of the CSP can be identified by using integer programming solver to satisfy the objective [27].

4.4 Reconstruction Approach

Reconstruction approach has two steps, first perform distortion of data and then reconstructing the distributions. There are many algorithms for reconstructing the distributions and data types [28].

For distributed data, Bayesian reconstruction process is used which is based on EM algorithm. EM algorithm is robust and it can estimate the original distribution when a large amount of data is obtained.

Another way of data reconstruction is to keep the original data aside and start from sanitizing knowledge base. The new data are reconstructed from the sanitized knowledge base [29].

4.5 Cryptographic Approach

Cryptography is a technique through which sensitive data can be encrypted. It is a good technique to protect the data.

In [28], the authors introduced cryptographic technique which is very common because it provides security and safety of sensitive attributes. There are different algorithms of cryptography available. But this technique has many disadvantages. It fails to protect the output of computation. It prevents privacy leakage of the computation. This algorithm does not give successful results when it talks about more parties. It is very complex to apply this algorithm for huge databases. Final data mining result may violate the privacy of the individual's record.

4.6 Hybrid Techniques Approach

Hybrid technique is a new approach through which one can combine two or more techniques to preserve the data. In [24],the authors proposed a hybrid technique in which they used randomization and generalization. In this approach first they randomize the data and then generalized the modified or randomized data. This technique protects private data with better accuracy; also it can reconstruct the original data and provide data with no information loss. In [19], a hybrid algorithm was proposed using genetic algorithm with data distortion algorithm to optimize the hiding side effects. But it suffers from existence of non-sensitive lost rules.

A comparative analysis of different hiding approaches given in Table 1 [30]:

Table1: different hiding approaches comparison

Approach		Advantage	Limitation
Heuristic approach	**Data Distortion** [20],[21]	More efficient, scalable	Difficult to revert the changes made in database
	Data Blocking [22],[23],[24],[25]	It maintains veracity of database, since instead of inserting false value it just block original value.	Suffer from various side effects like ghost rule, lost rule etc.
Border approach [8],[9],[26]		Maintain database quality by selecting the transaction that produces minimal side effect.	Theory of border difficult to understand Based on heuristic approach.
Exact approach [27]		Provides an optimal solution without any side effects	High complexity due to linear integer programming
Reconstruction approach [28],[29]		Lesser side effect than heuristic based approaches	Number of transaction is restricted in new released database
Cryptography approach [28]		Provide security in multi party computation or where data distributed in different locations	Does not provide security for the output of the computation and it is very difficult to apply on huge databases
Hybrid Techniques Approach [19],[24]		Can provide better data private protection or better measures	High complexity due to combining of two or more different techniques

5. METRICS AND PERFORMANCE ANALYSIS

(i) In this section, two categories of measures related to the performance of a hiding algorithm are presented.

(ii) Data sharing measures: measure the effect of side effects considering sensitive association rules that failed to be hidden, non-sensitive rules that were accidentally missed, and ghost association Rules that were created by the sanitization process [5].

(iii) Process performance measures: measures a set of measures that are correlated to the performance of a hiding algorithm as much as outside parameters are concerned [6].

5.1 Data-sharing measures:

Performance of any privacy preserving association rule mining is estimated using the following metrics:

5.1.1 Hiding Failure (HF)

It is the measure of sensitive association rules that appear in the sanitized database. It is the percentage of data that remain uncovered in the sanitized dataset. It measures the percentage of the number of sensitive association rules that cannot be hidden $S_R\,(D')$ over the number of sensitive rules to be hidden $S_R\,(D)$. It is calculated by using the below formula:

$$H_F = \frac{S_R\,(D')}{S_R\,(D)} \qquad (5.1)$$

Where D is the original data set, D' is the sanitized data set, $\mathrm{S_R}$ is the number of sensitive association rules [5].

5.1.2 Misses Cost (MC)

It is the measure of amount of valid association rules that are hidden by accident after sanitization (lost rules). It is the percentage of non-sensitive data hidden during the sanitization process. It is calculated as follows:

$$M_C = \frac{S'_R\,(D) - S'_R\,(D')}{S'_R\,(D)} \qquad (5.2)$$

Where $S'_R\,(D)$ is the size of set of all non-sensitive rules in the original database D**,** $S'_R\,(D')$ is the size of set of all non-sensitive rules in the sanitized database D' [5].

5.1.3 Artifactual Patterns (AF)

It is the measure of artificial association rules (ghost rules) created by adding the noise in the data. It is the measure of discovering ghost rules. It is calculated by:

$$A_F = \frac{|P'| - |P \cap P'|}{|P'|} \qquad (5.3)$$

Where P is the set of discovered association rules in the original database D , P' is the set of association rules in the sanitized database D' and |X| denotes the cardinality of X. The ghost rules information represents the set of new rules that can be extracted from the database after applying the sanitization technique [5].

5.1.4 Difference (Diff)

It is the measure of difference between original database and sanitized database. It is calculated by:

$$Diff(D, D^{'}) = \frac{1}{\sum_{i=1}^{n} f_D(i)} \times \sum_{i=1}^{n} [f_D(i) - f_{D'}(i)] \quad (5.4)$$

Where $f_D(i)$ represents frequency of $\boldsymbol{i^{th}}$ item in the original database, $f_{D'}(i)$ is the frequency of $\boldsymbol{i^{th}}$ item in the sanitized database and n is the number of distinct items in the original database [5].

5.1.5 Side-Effect Factor (SEP)

It is the amount of non-sensitive association rules that are removed during the sanitization process (lost rules). The side effect factor is used to quantify the amount of non-sensitive association rules that are removed as an effect of the sanitization process. It is calculated by:

$$S_{EF} = \frac{|P| - (|P'| + |S_R\,(D)|)}{|P| - |S_R\,(D)|} \qquad (5.5)$$

It is similar to the measure of misses cost [5].

5.1.6 Recovery Factor (RF)

This measure expresses the possibility of an adversary to use non-sensitive rules to recover a sensitive rule. The recovery factor of a pattern takes into consideration the existence of its subsets. If all the subsets of a sensitive rule can be recovered from the sanitized dataset, then the recovery of the rule itself is possible then RF=1; otherwise RF=0. However, this measure is not certain since, for instance, an adversary may knowing its subsets but cannot learn an itemset [5].

5.2 Process performance measures:

5.2.1 Efficiency

It is the measure of the ability of a privacy preserving algorithm to expeditiously use the existing resources and execute with fine performance. Efficiency is measured in terms of resource requirements (memory usage, required storage, CPU time, and communication requirements) and handling different sizes of the data [6].

5.2.2 Data Quality

The data quality of a privacy preservation algorithm depends on the quality of the sanitized database D' against the original database D and the quality of sanitized database data mining results [6].

Some of the possible measures for the quantification of the data quality are:

(i) Accuracy: measures the closeness of a sanitized value to the original one and is related to the data failure formed by the hiding strategy,

(ii) Completeness: used to measure the amount of lost data in the sanitized database

(iii) Consistency: This is associated with the relationships that have to maintain to hold between the different fields of a data item or between data items in a sanitized database.

5.2.3 Privacy Level

This category contains a set of measures that calculate roughly the quantity of uncertainty, according to the possibility that the sensitive data can be disclosed. It can be measured by the information entropy, the level of privacy and the J-measure [6].

6. CHALLENGES AND FUTURE TRENDS

(i) Very high complexity of the exact hiding approaches, mainly for large databases.

(ii) The blocking algorithms required more researches to improve it is capabilities.

(iii) Distortion techniques fail to offer a mapping between the original values in the dataset and the ones that were distorted in the sanitized dataset.

(iv) The border revision design needs to be able to direct hiding of association rules, instead of their indirect hiding through their generating itemsets.

(v) The need for more advanced measures for the comparison of the different hiding Techniques.

(vi) Good hiding algorithm needs to minimize or eliminate all hiding side effect and minimizing the modifications on database to increase the efficiency.

7. CONCLUSIONS

Privacy preserving data mining is concerned with the privacy of knowledge that is hidden in large databases. More specifically, the research directions which study how sensitive association rules can be hidden, side effects and different metrics and performance measures surveyed for the evaluation of the association rule hiding algorithms. This paper also presented a through analysis and comparison of the several surveyed approaches, as well as literature review, challenges and future trends. Privacy preserving is at the stage of improvement and optimization. Many privacy preserving algorithms of association rule mining are discussed; however, privacy preserving technology needs to be further researched because of the complexity of the privacy problem. In future, the goal is to develop a new distortion technique that use data evaluation function together with other optimization techniques to minimize the data modifications required for hiding sensitive rules and preforming that through one iteration.

8. REFERENCES

[1] Jajodia., C.F.a.S., The inference problem: A survey, in SIGKDD Exploration Newsletter. 2002. p. 6-11.

[2] M. Atallah, E.B., A. Elmagarmid, M. Ibrahim, V.S. Verykios, Disclosure limitation of sensitive rules, in Proceedings of the IEEE Knowledge and Data Engineering Exchange Workshop (KDEX'99). 1999. p. 45-52.

[3] E. Dasseni, V.S.V., A.K. Elmagarmid, E. Bertino. Association rule hiding. in Proceedings of the Fourth International Workshop on Information Hiding. 2001.

[4] Y. Saygm, V.S.V., C. Clifton, and Y. Saygin, Using unknowns to prevent discovery of association rules. ACM SIGMOD Record, 2001. Vol.(30)(No.4): p. 45-54.

[5] S.R.M. Oliveira, O.R.Z., Privacy preserving frequent itemset mining, in IEEE International Conference on Privacy, Security and Data Mining (CRPITS 2002). 2002. p. 43-54.

[6] E. Bertino, I.N.F., L.P. Povenza, A framework for evaluating privacy preserving data mining algorithms. Data Mining and Knowledge Discovery 2005. Vol.11 (No.2): p. 121-154.

[7] Jafari, S.-L.W.a.A., Hiding sensitive predictive association rules, in Systems, Man and Cybernetics, IEEE International Conference 2005. p. 164-169.

[8] Xingzhi Sun, Y., P.S., A border-based approach for hiding sensitive frequent itemsets, in Data Mining, Fifth IEEE International Conference. 2005. p. 1550-4786.

[9] Xingzhi Sun , Y., P.S., Hiding sensitive frequent itemsets by a border-based approach. Journal of Computing Science and Engineering, 2007. Vol.1(No.1): p. 74-94.

[10] Shyue-Liang Wang, D.P., Ayat Jafari Hiding collaborative recommendation association rules. Applied Intelligence, 2007. Vol.27(No.1): p. 67-77.

[11] T. Berberoglu and M. Kaya, Hiding Fuzzy Association Rules in Quantitative Data, in The 3rd InternationalConference on Grid and Pervasive Computing Workshops. 2008. p. 387-392.

[12] Chih-Chia Weng ; Dept. of Comput. Sci., N.D.U., Taoyuan ; Chen, Shan-Tai ; Hung-Che Lo, A Novel Algorithm for Completely Hiding Sensitive Association Rules, In Intelligent Systems Design and Applications, 2008. ISDA'08. Eighth International Conference. 2008. p. 202-208.

[13] C. N. Modi, U.P.R., and D. R. Patel, Maintaining privacy and data quality in privacy preserving association rule mining, In Computing Communication and Networking Technologies (ICCCNT), International Conference. 2010. p. 1-6.

[14] Yogendra Kumar Jain, V.K.Y., Geetika S. Panday, An Efficient Association Rule Hiding Algorithm for Privacy Preserving Data Mining. International Journal on Computer Science and Engineering (IJCSE), 2011. Vol.3: p. 2792 -2798.

[15] Komal Shah , A.T., Amit Ganatra Association Rule Hiding by Heuristic Approach to Reduce Side Effects & Hide Multiple RHS Items. International Journal of Computer Applications (0975 – 8887) 2012.

[16] D. Jain, P.K., R. Soni, and B. B. K. Chaurasia, Hiding Sensitive Association Rules without Altering the Support of Sensitive Item (s). Advances in Computer Science and Information Technology. Networks and Communications, 2012. Vol.3(No.2): p. 500–509.

[17] Rao, N.H.D.a.U.P., Hiding sensitive association rules to maintain privacy and data quality in database, in Advance Computing Conference (IACC), 2013 IEEE 3rd International. 2013. p. 1306–1310.

[18] Niteen Dhutraj, S.S., Vivek Kshirsagar Hiding Sensitive Association Rule For Privacy Preservation. IEEE Transactions on Knowledge and Data Engineering 2013.

[19] Peng Cheng, J.-S.P., Chun-Wei Lin, Use EMO to Protect Sensitive Knowledge in Association Rule Mining by Removing Items, in IEEE Congress on Evolutionary Computation (CEC). 2014: Beijing, China.

[20] T. Jahan, G.N.a.C.V.G.R. Data Perturbation and Features Selection in Preserving Privacy. In IEEE 2012 proceedings 9781-4673-1989-8/12. 2012.

[21] J. Liu, J.L.a.J.Z.H., Rating: Privacy Preservation for Multiple Attributes with Different Sensitivity requirements, In proceedings of 11th IEEE International Conference on Data Mining Workshops, IEEE 2011.

[22] A. Parmar, U.P.R., D. R. Patel, . Blocking based approach for classification Rule hiding to Preserve the Privacy in Database. In proceedings of International Symposium on Computer Science and Society, IEEE 2011.

[23] Animesh Tripathy , M.P. A novel framework for preserving privacy of data using correlation analysis. In Proceedings of the International Conference on Advances in Computing, Communications and Informatics (ICACCI '12). ACM. 2012. NY, USA.

[24] S. Lohiya and L. Ragha. Privacy Preserving in Data Mining Using Hybrid Approach. In proceedings of Fourth International Conference on Computational Intelligence and Communication Networks, IEEE 2012.

[25] Luciano Bononi, M.B., Gabriele D'Angelo, Lorenzo Donatiello. Concurrent Replication of Parallel and Distributed Simulations. In Proceedings of the 19th Workshop on Principles of Advanced and Distributed Simulation (PADS '05). IEEE Computer Society. 2005. , Washington, DC, USA.

[26] G. V. Moustakides , V.S.V., A max-min approach for hiding frequent itemsets., In Workshops Proceedings of the 6th IEEE International Conference on Data Mining (ICDM2006). 2006. p. 502–506.

[27] S. Menon, S.S., and S. Mukherjee, Maximizing accuracy of shared databases when concealing sensitive patterns. Information Systems Research,16(3), 2005: p. 256–270.

[28] Dragos N. Trinca, Fast and Cost-Effective Algorithms for Information Extraction in some Computational Domains. 2008, University of Connecticut, Storrs: CT, USA. .

[29] Yongcheng Luo, Y.Z.a.J.L. A Survey on the Privacy Preserving Algorithm of Association Rule Mining. In Proceedings of the IEEE Second International Symposium on Electronic Commerce and Security (ISECS '09). 2009. Washington, DC, USA, 241245.

[30] T. Sirole, J.C., A Survey of Various Methodologies for Hiding Sensitive Association Rules. International Journal of Computer Applications (0975 – 8887), 2014. Vol. 96(No.18).