IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Issue-5, May 2019 E-ISSN: 2347-2693

Heuristic Approach in Association Rule Hiding- A Study

S. Sharmila[1], S. Vijayarani[2]

[1,2]Department of Computer Science, Bharathiar University, Coimbatore, Tamilnadu, India -641046

DOI: https://doi.org/10.26438/ijcse/v7i5.300305 | Available online at: www.ijcseonline.org

Accepted: 09/May/2019, Published: 31/May/2019

***Abstract*-**Privacy preserving data mining Extracts relevant knowledge from large amount of data and at the same time protect sensitive information from the data miners. People in business, hospitals, educational institutions, and banks need a secure and safe transaction of their data. To serve this need Privacy Preserving Data Mining (PPDM) was created. PPDM solves the problem related to designing accurate models about combined data without requiring the access to exact information in individual data record. PPDM is the most important research area for protecting the perceptive data or knowledge. The important technique of PPDM is Association rule hiding that protects the association rules generated by association rule mining. This study presents a survey of association rule hiding approach for preserving privacy of the user data. Association rule hiding methodology consists of five approaches namely Heuristic, Border, Exact, Cryptography and Reconstruction The study has briefly explained the Heuristic approach.

***Keywords*-**Privacy preserving Data Mining, Association Rule Hiding approaches. Heuristic Approach

I. INTRODUCTION

There are several data mining techniques available to protect the privacy. Here, privacy preserving techniques are classified on the basis of data distribution, data distortion, data mining algorithms, anonymization, data or rules hiding, and privacy protection. Intensive research findings over the decades have revealed that the existing privacy preserving data mining search approaches are still suffering from major inadequacies which include distributed clients' data to multi semi honest providers, overhead of computing global mining, and incremental data privacy issue in cloud computing, integrity of mining result, utility of data, scalability and overhead performance [1]. Thus, a robust, scalable model is essential to overcome these shortcomings. Furthermore, to protect the privacy of each client, proper anonymization of data is essential prior to publishing it. The connection between personal data and personal identification should be dispelled.

Privacy preservation data mining techniques are divided into two areas—, data hiding and knowledge hiding [2]. Data hiding is modification of sensitive data before disclosing to others. Knowledge hiding means hiding of sensitive knowledge after extracting it from database. In Data hiding confidential or private information is removed from the data before its disclosure. Knowledge hiding is concerned with the sanitization of confidential knowledge from the data [3].Generally, security protects the data against unauthorized access when transmitted across a network [4] [5]. However, reaching an authorized user, no additional constraints on revealing the personal information of an individual are imposed [6].

II. LITERATURE SURVEY

Vassilios S. Verykios [7] Author had discussed about deep study of knowledge hiding, privacy of information that was hidden in large databases. Author had analyzed the use of data mining techniques author had investigated about methods of hiding sensitive association rules by categorize, some open challenges were also discussed. It was found that ruling an optimal solution for sanitizing database (to defend privacy of sensitive information) is NP-Hard. Existing move toward provided only the estimated solution to hide responsive knowledge. There is need of judgment exact explanation to the privacy trouble in database disclosure

Mohammad azam chhipa [8] Discussed about data mining, association rule mining and techniques to hide sensitive rules which were mined through data mining. Author has studied about association rule and algorithms to mine association rule in data set, and also studied about studied different approaches to hide sensitive rules.

S.Vijayarani [9] had discussed about how privacy was preserved by removing the sensitive rules from the rules to reconstruct the new dataset from the non-sensitive rules. The efficiency of each algorithm was determined based on the Accuracy and

Time factor. The author had concluded that privacy was preserved for the classification rules by reconstructing the new dataset by removing the sensitive rules.

Dr. S. Vijayarani [10] research work was Concentrated on how the traditional algorithms were used for generating association rules in data streams. A number of rules generated by an algorithm and execution time were considered for performance Factors.

Mahtab Hossein Afshari [11] had introduced cuckoo optimization algorithm for hiding rules. Hiding was performed using the distortion technique. Author had defined three fitness functions which had made possible to achieve a solution with the fewest side effects. Introduced an efficient immigration function in this approach has improved the ability to escape from any local optimum. The efficiency of proposed approach was evaluated by conducting some experiments on different databases.

Syed Shujauddin Sameer [12] Analyzed three different techniques for securing the sensitive data from the user First technique increases the support of left-hand side rule it had required changing in the database. In second approach query analyst required some effort by the user. The third technique was likely to block the access to certain type of data

III. CLASSIFICATION OF HEURISTIC APPROACHES

I. Data Partitioning Techniques

Data Partitioning techniques have been applied to scenarios in which the databases are accessible for mining are distributed across a number of sites, each sites are willing to share only data mining results, not the source data. In these cases, the data are distributed either horizontally or vertically. In horizontal partition the transactions are distributed in multiple partitions while in vertical partition the attributes are split across multiple partitions. Data partitioning techniques can be classified into two sub category; Cryptography-Based Techniques and Generative-Based Techniques [13].

a. Cryptography-Based Techniques

Cryptography-Based techniques are used to solve the secure multiparty computation (SMC) problem [14][15]. Similarly, secure multiparty computation (SMC) exist; when two or more party want to communicate but neither party want to disclose confidential data to third one.

b. Generative-Based Techniques

.This approach shares just a small portion of its local model in each party that is used to construct the global model. The existing solutions are built over horizontally partitioned data and privacy preserving distributed clustering uses generative model [16].

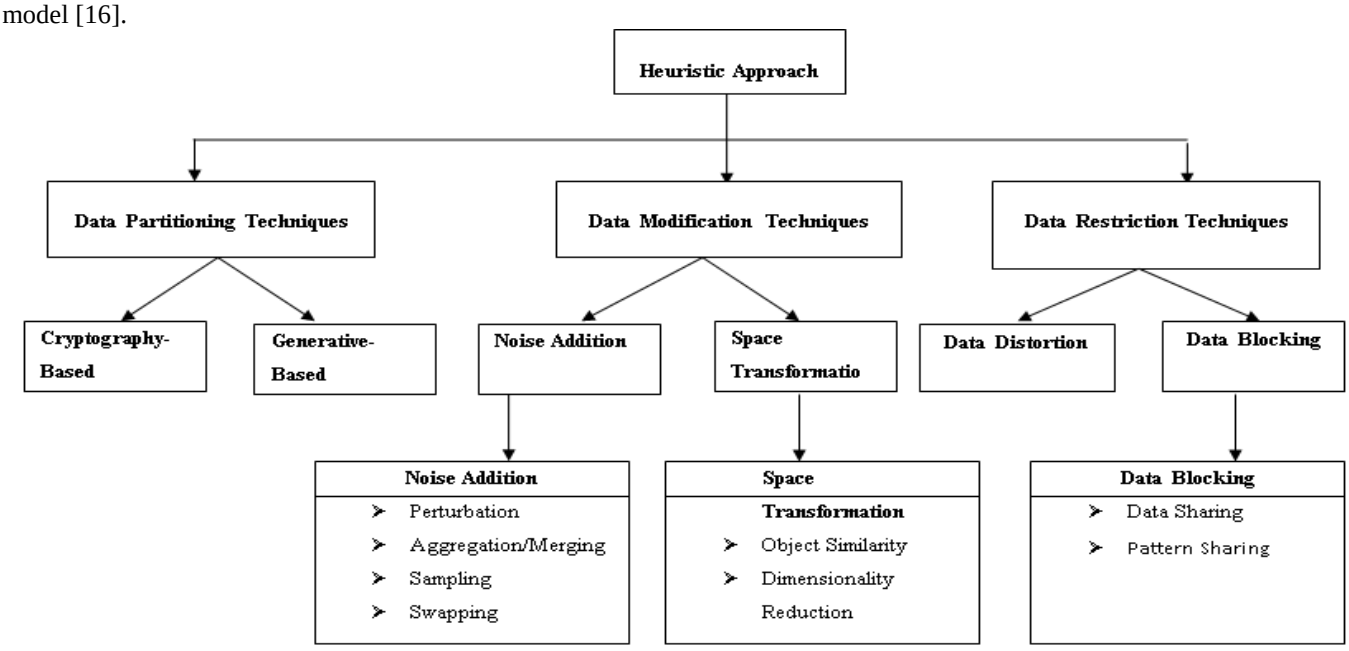

Figure 1: Classification of Heuristic approaches

Figure 2 describes the classification of Heuristic approaches. It can be further classified into three broad categories namely, data partitioning] data modification and data restriction [17].

II. DATA MODIFICATION TECHNIQUES

In this approach, some of the values in original database are modified to preserve data. . In this technique, the dataset chosen is binary transactional dataset and the entry value is flipped only. The data is altering by replacing 1's to 0's and vice versa until the support or confidence of association rules is drop below certain threshold. The technique is further divided into noise addition techniques and space transformation techniques [18].

a. Noise Addition Techniques

In this approach, some noise for e.g., information not present in a particular record or transaction is added to the original data to prevent the discovery of confidential data or to preserve the privacy of confidential information. Noise is added to confidential attributes by randomly shuffling the attribute values to prevent the discovery of restrictive patterns that are not supposed to be discovered. The technique is further divided in to four sub categories: perturbation, aggregation or merging, sampling and swapping [19].

- Perturbation: Modify the original value of attributes, by changing 1 to 0 or b adding noise.
- Aggregation or Merging: The combination of several values into a broad category.
- Sampling: Modify data for only sample of a population.
- Swapping: Interchange values of individual records (transaction) [20].

b. Space Transformation Techniques

This technique not only meets privacy requirements but also guarantee valid cluster result. Two new space transformation techniques were described, called object similarity based-representation and dimensionality reduction-based transformation [21].

c. Object Similarity Based-Representation:

The idea of this approach is the similarity between objects. In this technique, if the data owners want to share data, first they compute dissimilarity matrix between object and then share such a matrix with third party [21].

d. Dimensionality Reduction-Based Transformation:

This technique is applicable when the attributes of object reside either in a central location or split across multiple site. Similarly, this approach is referred as privacy preserving clustering over partition data [21].

III. DATA RESTRICTION TECHNIQUES

The prime objective of data restriction technique is to limit access to mining results. Particularly, the techniques can be classified as generalization, suppression of information or by blocking the access to some pattern that are not hypothetical (imaginary) to be discovered. Moreover, this technique decrease the confidence of the rule with the minimum confidence threshold by placing question mark "?" in place of original value [22]. Furthermore, the technique produces uncertainty of the support and confidence of the rule without distorting the database. This technique is further divided in to two sub category: Data Blocking techniques and Data Distortion Technique [26].

a. Data Distortion Technique

In this technique the values are replaced from 1 to 0 or 0 to 1. Data Distortion Technique has two basic approaches for rule hiding. First it reduces the support of rules and second reduces the confidence of rules [49]. This technique used for hiding the sensitive knowledge of database, this can be possible by reducing the support or confidence of the sensitive rules [26] [27].

b. Blocking-Based Techniques

This technique modifies the original value of attributes by an unknown or question mark "?". Moreover, the technique is useful to hide confidential information if data are share for mining. Furthermore, the technique is applicable to preserve privacy in association rule and classification rule. It means that the private information remains private after hiding process. These technique can be divided into two groups: data-sharing techniques and pattern-sharing techniques [28].

- Data-Sharing Techniques hides the sensitive association rules that contain confidential information by performing some modification in the original data. To do so, only a small number of transaction that contain sensitive association rules will be modified by removing some item or by adding noise [28] [30].

- Pattern-Sharing Techniques: These techniques insure privacy preserving in association rules by removing the sensitive association rules before sharing the data. In addition, the technique acts on sensitive rule instead of the data itself [28][29].

Table 1. Comparative analysis on Heuristic Approach algorithms

Algorithm	**Items**	**Rules**	**Transaction**
Increase Support of LHS (ISL)	**Y**		
Increase The Support Of Left Hand Side Of The Rule (ISLF)		**Y**	
Decrease Support Of RHS (DSR)		**Y**	
SHR		**Y**	
Multi-Objective Method	**Y**	**Y**	
Tabu Search Optimization Technique	**Y**		
Genetic Algorithms	**Y**		
Cuckoo Optimization Algorithm	**Y**		
Artificial Bee Colony Algorithm			**Y**
Decrease The Support Of The Right Hand Side Of The Rules		**Y**	
New distortion based rule-hiding Method			**Y**
Data sanitization algorithm	**Y**		
FIH Using Intuition from Column Generation	**Y**		
Sliding Window Algorithm			**Y**
Deleting RHS rule		**Y**	
PPGA	**Y**		
data distortion technique		**Y**	
SIF-IDF (sensitive items frequency-inverse database frequency)	**Y**		
Algorithms RIPPER, PART, C4.5 Classification	**Y**		
Minimum Frequency Item Algorithm\ (MFI):Classification	**Y**		
MDSRRC (Modified Decrease Support of R.H.S. item of Rule Clusters)		**Y**	
Item Grouping Algorithm (IGA):Clustering	**Y**		

IV. CONCLUSION

Privacy preserving in data mining has a significant value in business operations. When data is shared through a network, data users can see all of the data. Solving this problem is a formidable challenge. Data mining techniques provide better results for the safe transaction of the data. This study presents a survey of Association Rule Hiding approaches to solve privacy problems. The main objective of PPDM is to incorporate the traditional data mining techniques to transform the data with the view to mask sensitive information, and a major challenge in this process is to efficiently transform the data and recover its mining outcome from the transformed one. Thus, the study examines the overhead for preserving privacy of growing data, and the integrity of mining result. The study discusses Heuristic-based approach in detail with the aim to solve the major problems associated with privacy preservation.

REFERENCES

[1]. S. Verykios, A. K. Emagarmid, E. Bertino, Y. Saygin, and E. Dasseni. Association rule hiding. IEEE Transactions on Knowledge and Data Engineering, **16(4):434–447**, 2004.

[2]. Inan and Y. Saygin. Privacy preserving spatio-temporal clustering on horizontally partitioned data. In Proceedings of the 8th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2006), **pages 459–468,** 2006.

[3]. Peng Cheng1,3 • John F. Roddick2 • Shu-Chuan Chu2 • Chun-Wei Lin1 Privacy preservation through a greedy, distortion-based rule-hiding method, Springer Science+Business Media New York 2015 , Appl Intell (2016) **44:295–306** DOI 10.1007/s10489-015-0671-0.

[4]. Telikani, A. Shahbahrami and R. Tavoli,Data sanitization in association rule mining based on impact factor, Journal of AI and Data Mining Vol **3, No 2**, **131-140.** 2015.

[5]. Syam Menon and Sumit Sarkar privacy and big data: scalable approaches to sanitize large transactional databases for sharing, big data & analytics in networked business MIS Quarterly **Vol. 40 No. 4, pp. 963-98**1/December 2016.

[6]. Afrah Farea, Ali Karcı Applications of Assoiation Rules Hiding Heuristic Approaches, SİU-2015: Sinyal İşleme Ve İletişim Uygulamalari Kurultayi.

[7]. R.Hemalatha 2 M.Elamparithi Privacy Preserving Data Mining Using Sanitizing Algorithm, International Journal of Computer Science and Information Technologies, **Vol. 6 (5)** , **4174-4179**, ISSN -0975-9646 2015.

[8]. Divya C. Kalariya , Association Rule Hiding based on Heuristic Approach by Deleting Item at R.H.S. Side of Sensitive Rule International Journal of Computer Applications **Volume 122 – No.8 (0975 – 8887),** July 2015

[9]. Saad M. Darwish, Magda M. Madbouly, and Mohamed A. El-Hakeem, A Database Sanitizing Algorithm for Hiding Sensitive Multi-Level Association Rule Mining, International Journal of Computer and Communication Engineering, **Vol. 3, No. 4**, July 2014.

[10]. Rahat Ali SHAH,Sohail ASGHAR Privacy preserving in association rules using a genetic algorithm , Turkish Journal of Electrical Engineering & Computer Sciences, (2014) **22: 434** { 450, doi:10.3906/elk-1206-66.

[11]. Dr.Vijayalakshmi M N1 S.Anupama Kumar2 Kavyashree BN3, Investigating Interesting Rules Using Association Mining for Educational Data, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume **3, Issue 2**, February 2014.

[12]. Tapan Sirole ,Jaytrilok Choudhary , A Survey of Various Methodologies for Hiding Sensitive Association Rules, International Journal of Computer Applications **(0975 – 8887) Volume 96– No.18**, June 2014.

[13]. Dhiren R. Patel, Ph.D Khyati B. Jadav Jignesh Vania , A Survey on Association Rule Hiding Methods, International Journal of Computer Applications **(0975 – 8887) Volume 82 – No 13**, November 2013

[14]. Suma B. Suma B. Association Rule Hiding Methodologies: A Survey, International Journal of Engineering Research & Technology (IJERT) ISSN: **2278-0181 Vol. 2 Issue 6,** June – 2013.

[15]. Dhyanendra Jain 1,Amit sinhal2,Neetesh Gupta3, Hiding Sensitive Association Rules without Altering the Support of Sensitive Item(s), International Journal of Artificial Intelligence & Applications (IJAIA), **Vol.3, No.2,** March 2012.

[16]. Shyue-LiangWang, Tzung-Pei Hong • Chun-Wei Lin •Kuo-Tung Yang, Using TF-IDF to hide sensitive itemsets, Appl Intell **(2013) 38:502–510** DOI 10.1007/s10489-012-0377-5.

[17]. Nidhi Porwal, Sunil Kumar Mahaveer Singh , An Algorithm for Hiding Association Rules on Data Mining, National Conference on Communication Technologies & its impact on Next Generation Computing CTNGC 2012 Proceedings published by International Journal of Computer Applications® (IJCA).

[18]. Gwadera GLR, Gkoulalas-Divanis A (2013) Permutation-based sequential pattern hiding. In: IEEE International Conference on Data Mining (ICDM**), pp 241–250**

[19]. Mariscal, G., Marban, O. & Fernandez, C. (2010).A survey of data mining and knowledge discovery process models and methodologies. The Knowledge Engineering Review, **vol. 5, no. 2, pp. 137-166**.

[20]. Vignani, B. & Satapathy, S. C. (2014). D-pattern evolving and inner pattern evolving for high performance text mining. Advances in Intelligent Systems and Computing**, vol. 247, pp. 501-507**.

[21]. Jena, L. K., Kamila, N. & Mishra, S. (2014). Privacy preserving distributed data mining with evolutionary computing. Advances in Intelligent Systems and Computing, **vol. 247, pp. 259-267.**

[22]. Lijffijt, J., Papapetrou, P. & Puolamäki, K. (2014).A statistical significance testing approach to mining the most informative set of patterns. Data Mining and Knowledge Discovery, **vol. 28, pp. 238-263**.

[23]. Verykios, V. 2013. "Association Rule Hiding Methods," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **(3:1), pp. 28-36.**

[24]. T. Tassa, "Secure Mining of Association Rules in Horizontally Distributed Databases", IEEE Transactions on Knowledge and Data Engineering, vol. **26, no. 4, (2014) April, pp. 970-983**.

[25]. R. Nilesh, and P. b. Nilesh, S. H. Krupali, "Privacy Preserving in Association Rule mining", international journal of advanced and innovative research (IJAIR), **Vol.2, No. 4, 2013, pp. 2278-7844**.
[26]. H.Hamilton, DBD: data mining projects", University of Regina Available at:http://www2.cs.uregina.ca/dbd/cs831/index.html, 2000{9, Last accessed: 15.03.2012
[27]. Maulesh R. Chhatrapati Shilpa Sherasiya Privacy Preserving Data Mining Using Heuristic Approach IJIRST –International Journal for Innovative Research in Science & Technology| **Volume 1 | Issue 10** | March 2015 ISSN (online): **2349-6010**.
[28]. Saiyed Wafa Ahsan* R. K. Gupta A Heuristic Approach to Association Rule Mining International Journal of Advanced Research in Computer Science and Software Engineering. Volume **6, Issue 2, February 2016** ISSN: 2277 128X.
[29]. Hitesh Chhinkaniwala 1 and Dr. Sanjay Garg2 Privacy Preserving Data Mining Techniques: Challenges & Issues. Proceedings of International Conference on Computer Science & Information Technology, CSlT – 2011.
[30]. Mrs. P.Cynthia Selvi, Dr. A.R.Mohamed Shanavas An Effective Heuristic Approach for Hiding Sensitive Patterns in Databases IOSR Journal of Computer Engineering (IOSRJCE) ISSN: 2278-0661 **Volume 5, Issue 1 (Sep-Oct. 2012), PP 06-11**

AUTHORS PROFILE

Mrs. S.Sharmila, is pursuing her Ph.D in Department of Computer Science, Bharathiar University, Coimbatore, and Tamilnadu, India. She has completed M.C.A in Bharathiar University, Tamilnadu, India. Her research area includes Association Rule Mining, Association Rule Hiding, Privacy Preserving and Optimization techniques. She has published papers in International Journals and Conferences.

Dr. S.Vijayarani Mohan is an Assistant Professor of Department of Computer Science at Bharathiar University, Coimbatore, India. She has obtained M.C.A., M.Phil., and Ph.D., in Computer Science. She has 10 years of teaching/research and 10 years of technical experience. Her research interests include data mining, privacy issues in data mining, text mining, web mining, data streams and information retrieval. She has published more than 95 research articles in national/international journals. She also presented research papers in international/national conferences. She has authored a book and guided more than 25 research scholars. She is a life member in professional bodies like CSI, ISCA, IAENG, IRED and UACEE