IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-6, Issue-11, Nov 2018 E-ISSN: 2347-2693

A Review Approaches for Hiding Sensitive Association Rules in Data Mining

Janki Patel [1*], Priyanka Shah [2]

[1,2] Information Technology Engineering Department, Sardar Vallabhbhai Patel Institute of Technology, Vasad, Gujarat, India

Available online at: www.ijcseonline.org

Accepted: 17/Nov/2018, Published: 30/Nov/2018

***Abstract*—** Nowadays, Data Mining is a popular tool for extracting hidden knowledge from huge amount of data. To find hidden knowledge in the data without revealing sensitive information is one of the major challenges in data mining. There are many strategies have been proposed to hide the sensitive information. Association rule mining is one of the data mining techniques used to extract hidden knowledge from large datasets. This hidden knowledge contains most of the times confidential information that the users want to keep private or do not want to disclose to public. Therefore, privacy preserving data mining (PPDM) techniques are used to preserve such confidential information or restrictive pattern from unauthorized access. In this paper, all the approaches for hiding sensitive association rules in PPDM have been compared theoretically and points out their pros and cons.

***Keywords*—** Data Mining, Association rule mining, privacy preserving data mining (PPDM)

I. INTRODUCTION

Privacy preserving data mining (PPDM) is a fruitful research area in Data Mining. In PPDM, data mining algorithms are analyzed and compared the impacts which occur in data privacy. The goal of PPDM is to transform the existing dataset in some way that the confidentiality of the data and knowledge remains intact even after the mining process. Methods that allow the knowledge extraction from data, while preserving privacy, are known as privacy-preserving data mining (PPDM) techniques. Association Rules are widely used to analyze retail basket or transaction data. Association Rules are intended to identify strong rules discovered in transaction data using measures of interestingness, based on the concept of strong rules. Numerous techniques are used to hide sensitive association rules by performing some modification in the original dataset. Most of the existing techniques are based on support and confidence framework. In addition, we identified that most of the techniques are suffering from the side effects of lost rules, ghost rules and other side effect, such as number of transaction modified and hiding failure.

Mining association rules can be reduced to two sub-problems. First one is: to find all frequent item sets for a predetermined minimum support and second is: to generate the association rules from the large item sets found. Most of the times frequent item sets are required for decision making process. Association Rules generated from such item sets are called as positive rules, but sometimes there is a need to find infrequent item sets to discover negative association rules.

Association rule hiding has been the topic of a number of surveys and review articles, as well as books, where the goal was to collect and classify association rule hiding algorithms. In this survey, we attempt to provide an analytical review of major directions in association rule hiding and present our own insights into this topic. Unlike other surveys that describe the algorithms in terms of an approach-based classification, our research does not intend to provide a detailed description of association rule hiding algorithms because some decent surveys already exist. The main objective is to introduce new concepts and trends about different perspectives of data privacy process.

Section I contains the introduction of basic approach for weather forecasting. II contain the related works of basic literature papers. Section III contain the methodology and algorithms section IV explain the comparative study between different algorithms and at last conclusion and future scope.

II. RELATED WORK

Vaishali Patil, Ramesh Vasappanavara, Tushar Ghorpade(2016) proposing secure multi-party algorithm based on secure sum technique to simplify the operation of mining association rule when the database is horizontally partitioned among multiple sites. We are using a Frequent-Pattern (FP) growth algorithm to find frequent itemsets and try to reduce total computation time. Items in the transactions are encrypted to provide security from unauthorized access. Association rule mining can be performed in centralized as

well as partitioned database. In the case of partitioned database, security of the frequent itemsets is a very important aspect. Various techniques are proposed to achieve this goal which uses a third party or homomorphic encryption techniques. But sometimes security may get violated in the case of a trusted third party and the cost of encryption and decryption is higher in case of homomorphic encryption..As a part of future scope further work can be done on using different algorithm for frequent itemset mining and security on different datasets in other partitioning scheme of database.

Lichun Li, Rongxing Lu, Kim-Kwang Raymond Choo, Anwitaman Datta, and Jun Shao(2016) proposing privacypreserving mining on vertically partitioned databases. In such a scenario, data owners wish to learn the association rules or frequent itemsets from a collective dataset, and disclose as little information about their (sensitive) raw data as possible to other data owners and third parties. To ensure data privacy, we design an efficient homomorphic encryption scheme and a secure comparison scheme. We then propose a cloud-aided frequent itemset mining solution, which is used to build an association rule mining solution. Our solutions are designed for outsourced databases that allow multiple data owners to efficiently share their data securely without compromising on data privacy. Our solutions leak less information about the raw data than most existing solutions. Demonstrating the utility of the proposed homomorphic encryption scheme and outsourced comparison scheme in other settings will be the focus of future research.

Yaoan Jin, Chunhua Su, Na Ruan, and Weijia Jia (2016) proposing a new scheme which is a secure and efficient association rule mining (ARM) method on horizontally partitioned databases. We enhance the performance of ARM on distributed databases by combining Apriori algorithm and FP-tree in this new situation. To help the implement of combining Apriori algorithm and FP-tree on distributed databases, we originally come up with a method of merging FP-tree in our scheme. We take advantage of Homomorphic Encryption to guarantee the security and efficiency of data operation in our scheme. More speficially, we use Paillier's homomorphic encryption method which only has addition homogeneity to encrypt items' supports. At last, we perform experimental analysis for our scheme to show that our proposal outperform the existing schemes. In the future work, our scheme can be more efficient by improving our encryption methods.

Golnar Assadat Afzali, Shahriar Mohammadi Jia (2017) proposing data anonymisation which is used to fit the proposed model for big data mining. Besides, special features of big data such as velocity make it necessary to consider each rule as a sensitive association rule with an appropriate membership degree. Furthermore, parallelization techniques which are embedded in the proposed model, can help to speed up data mining process. In this research, new big data association rule hiding technique is presented, which uses fuzzy logic approach, tries to decrease undesired side effect of sensitive rule hiding on non-sensitive rules in data streams. Features such as parallelism and scalability are embedded in the proposed model to provide the facility of implementing this model for huge volume of data. Results show that the proposed model can be more effective in big data mining than existing rule hiding approaches. As future work, we will try to decrease undesired side effect of the proposed model to gain less information loss.

Shabnum Rehman and Anil Sharma (2017) proposing Association rule is one of the most used data mining techniques that discover hidden correlations from huge data sets. There are several mining algorithms for association rules Apriori is one of the most popular algorithm used for extracting frequent item sets from databases and getting the association rule for knowledge discovery. The time required for generating frequent item sets plays an important role. Based on this algorithm we are performing comparison of sanitized data and existing data based on number of iterations and the execution time. The experimental results shows that the number of iteration is reduced in sanitized data than that of existing data also the time is reduced in sanitized data. The association rule generation leads to ensure privacy of the dataset by creating items so, in this way privacy of association rules along with data quality is well maintained.

III. METHODOLOGY

A. Apriori Algorithm:

Apriori algorithm find frequent itemset and then base on support and confidence rules are generated. It is easy to implement. Also easy to understand. It can use to large dataset – Scalable. Sometimes not able to find a large number of candidates rules which can be computationally expansive. Calculating support every time –time consuming.

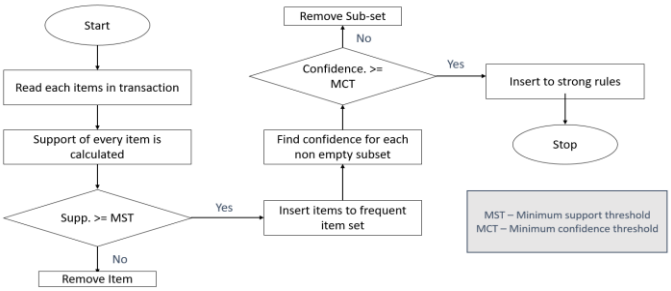

Figure 1. Basic Flow of Apriori Algorithm

B. FP Growth Algorithm:

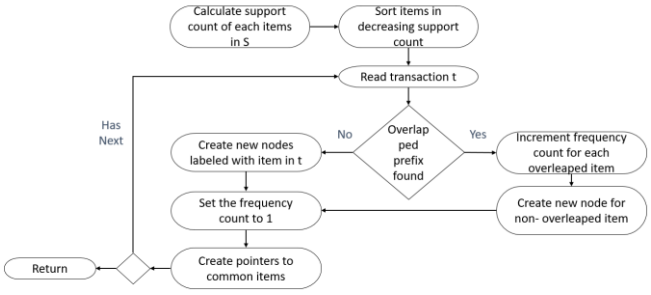

Figure 2. Basic Flow of FP Growth

FP growth algorithm use divide and conquer strategy to generate association rules. This algorithm mainly works in two phase construct tree and extract frequent items. It is used to construct tree algorithm use bottom up approach. It is compressed data structure. It is less expensive computation. It is Scanning whole data only twice.

C. Genetic Algorithm

A genetic algorithm (GA) is a method for solving both constrained and unconstrained optimization problems based on a natural selection process that mimics biological evolution. The algorithm repeatedly modifies a population of individual solutions. At each step, the genetic algorithm randomly selects individuals from the current population and uses them as parents to produce the children for the next generation. Over successive generations, the population "evolves" toward an optimal solution.

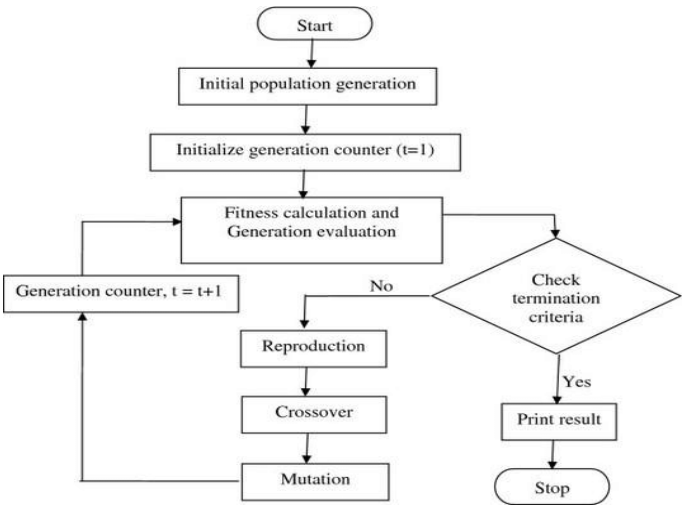

Figure 3. Basic Flow genetic algorithm

To generate strong association rules, Genetic Algorithm operators like selection, crossover and mutation have been applied on association rule generated by Apriori algorithm.

IV. COMPARATIVE STUDY

Table I. Comparison between Methods

Proposed Method	Pros	Cons
Apriori algorithm[8]	Number of iterations in sanitized data is less than original data	An improved Apriori can be implemented to deal with the scalability of data
FP growth algorithm[3]	Works without candidate item generations and avoids multiple scans of database	Security on different datasets changes
Homomorphic encryption[13]	It can calculate the association rules correctly	It can be more efficient by improving our encryption methods.
Fuzzy logic[6]	Decrease undesired side effect of sensitive rule hiding on non-sensitive rules in data streams	Information loss
Paillier additive cryptosystem[7]	Use of PC Tree for frequent item set calculation	Synchronization is required. Key generation is not robust.
Secure multi-party computation algorithm[9]	No use of any cryptosystem.	Complicated inequality verification. Candidate set generation

Genetic Algorithm[12]	The best point in the population approaches an optimal solution and Selects the next population by computation which uses random number generators	Efficiency can be improved

V. CONCLUSION

This paper shows comparative study of different algorithms for association rule hiding techniques on basis of various parameters and for future analysis FP growth is more efficient than other algorithms as it can handle large dataset easily in less time complexity. A detailed study of genetic algorithm used in hiding rues. Genetic and its various application are explained in detail. Performing a fair comparison of rule hiding is inherently a challenging task, since every proposed algorithm uses different settings and metrics. The performance of the algorithms might vary among different combinations of datasets and input parameters. Genetic Algorithm allow for the same level of privacy as any other algorithm, while also allowing for operations to be performed on the data without the need to see the actual data. We observe that Genetic Algorithm is always better than other algorithms.

REFERENCES

[1] Shubhra Rana, Dr. P. anthi Thilagam, "Hierarchical Homomorphic Encryption based Privacy Preserving Distributed Association Rule Mining". IEEE 13th International Conference on Information Technology, 2014.

[2] Rachit V. Adhvaryu, Nikunj H. Domadiya, "Privacy Preserving in Association Rule Mining On Horizontally Partitioned Database". International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), Vol. 3, Issue 5, May 2014.

[3] Vaishali Patil, Ramesh Vasappanavara, Tushar Ghorpade, "Securing association rule mining with FP growth algorithm in horizontally partitioned database". International Conference on Control, Computing, Communication and Materials (ICCCCM), 2016.

[4] Lichun Li, Rongxing Lu, Kim-Kwang Raymond Choo, Anwitaman Datta, and Jun Shao, "Privacy-Preserving Outsourced Association Rule Mining on Vertically Partitioned Databases". IEEE, 2016.

[5] Yaoan Jin, Chunhua Su, Na Ruan, and Weijia Jia, "Privacy-Preserving Mining of Association Rules for Horizontally Distributed Databases Based on FP-Tree". Springer International Publishing AG, 2016.

[6] Golnar Assadat Afzali, Shahriar Mohammadi Jia, "Privacy preserving big data mining: association rule hiding using fuzzy logic approach". The Institution of Engineering and Technology, 2017.

[7] Umesh Kumar Sahu, Anju Singh, "Approaches for Privacy Preserving Data Mining by Various Associations Rule Hiding Algorithms – A Survey". International Journal of Computer Applications, 2016.

[8] Shabnum Rehman and Anil Sharma, "Privacy Preserving Data Mining Using Association Rule Based on Apriori Algorithm". Springer Nature Singapore Pte Ltd, 2017.

[9] Narges Jamshidian Ghalehsefidi,, Mohammad Naderi Dehkord, "A Hybrid Algorithm based on Heuristic Method to Preserve Privacy in Association Rule Mining". Indian Journal of Science and Technology,2016.

[10] D. Menaga, S. Revathi, "Least lion optimisation algorithm (LLOA) based secret key generation for privacy preserving association rule hiding". The Institution of Engineering and Technology, 2018.

[11] Chun-WeiLin, Tzung-PeiHong, Hung-ChuanHsu, "Reducing Side Effects of Hiding Sensitive Itemsets in Privacy Preserving Data Mining". Hindawi Publishing Corporation the Scientific World Journal, 2014.

[12] Bettahally N. Keshavamurthy, Asad M. Khan, Durga Toshniwal, "Privacy preserving association rule mining over distributed databases using genetic algorithm". Neural Comput & Applic, 2013.

[13] Baocang Wang, Yu Zhan, and Zhili Zhang, "Cryptanalysis of a Symmetric Fully Homomorphic Encryption Scheme". IEEE, 2017.

[14] P. Amaranatha Reddy, MHM Krishna Prasad, "Challenges to find Association Rules over various types of data items: a Survey". International Conference on Computing, Communication and Automation(ICCCA), 2017.

[15] Chan Man Kuok, Ada Fu, Man Hon Wong, "Mining Fuzzy Association Rules in Databases".

[16] Harendra chahar, B N keshavamurthy, chirag modi, "Privacy-preserving distributed mining of association rules using Elliptic-curve cryptosystem and Shamir's secret sharing scheme". Indian Academy of Sciences, 2017.

Authors Profile

Miss. Janki Patel pursed Bachelor of Computer Science and Engineering from Institute of Technology and Management Universe, Vadodara in 2017. She is currently pursuing Master of Computer Engineering from Sardar Vallabhbhai Patel Institute of Technology,Vasad. She has published 2 research papers in reputed international journals. Her main research work focuses on Cryptography Algorithms, Network Security, Cloud Security and Privacy, Big Data Analytics, Data Mining and Computational Intelligence based education.

Mrs. Priyanka Shah pursed Bachelor of Information Technology and Master of Computer Engineering. She is currently working as Assistant Professor in Information Technology Department in SVIT, Vasad. Her main research work focuses on specialized area Cloud Computing, Mobile Computing, Wireless Communication Enterprise resource planning and Computer Organization. She has 10 years of academic experience.

