© 2018 IJSRSET | Volume 4 | Issue 1 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section : Engineering and Technology

Survey on Privacy-Preserving Mining of Association Rule and Double Encryption Technique

Rutuja Thite, Dr. M. U. Kharat
Department of Computer Engineering, Bhujbal Knowledge City, Nashik, Savitribai Phule Pune University, Pune, India

ABSTRACT

Data mining can extract important knowledge from large data collections, but sometimes these collections are split among various parties. Privacy concerns may prevent the parties from directly sharing the data and some types of information about the data. Such data is available is huge amount so it is very difficult to find out the data and relationship among items. For this problem, association rule mining with improved cryptographic technique is one of the solutions of data mining techniques, which can efficiently correlate the items. The output of such technique can be used in many real time applications to take the proper decisions. But the data owner, who shares their data for mutual advantages, wants to secure their data in association rule mining process. Because it can reveal the sensitive data, which might be harmful. Therefore it becomes very challenging task to achieve the security of data while mining the knowledge from it. This paper represents the core idea of privacy preserving association rule mining on vertically partitioned data with use of improved cryptographic technique.

Keywords : Privacy-Preserving Data Mining, Association Rules Mining, Vertically Partition Data, Encryption Techniques.

I. INTRODUCTION

Data mining innovation has developed as methods for identifying patterns and patterns from large quantities of data. Mining incorporates different algorithms, for example, clustering, classification, association rule mining and sequence detection and succession location. Traditionally, every one of these algorithms have been developed within a centralized model, with all information being assembled into a central site, and algorithms being run against that information.

Association rule mining discovers all rules in the databases that fulfill some minimum support and minimum confidence requirements. Numerous algorithms are utilized to improve the privacy and security of information. By vertically partitioned, we imply that every site contains some elements of a transaction. Utilizing the traditional —market basket illustration, one site may contain basic grocery purchases, while another has clothing purchases. Utilizing a key, for example, credit card number and date, we can join these to recognize relationships between purchases of clothing and groceries. Notwithstanding, this reveals the individual purchases at every site, possibly violating consumer privacy agreements.

There are more sensible illustrations. In the sub-assembly manufacturing process, distinctive manufacturers give components of the finished product. Cars incorporate several subcomponents; tires, electrical equipment, and so forth; made by independent producers. Once more, authors have proprietary data collected by several parties, with a single key joining all the data sets, where mining would help detect/predict malfunctions. The recent

trouble between Ford Motor and Firestone Tire give a real-life illustration. Ford Explorers with Firestone tires from a particular factory had tread separation issues in specific circumstances, resulting in 800 injuries. Since the tires did not have issues on different vehicles, and different tires on Ford Explorers did not represent an issue, neither one of the sides felt responsible. The delay in recognizing the genuine issue prompted to a public relations nightmare and the eventual replacement of 14.1 million tires. Many of these were probably fine – Ford Explorers represented just 6.5 million of the replaced tires. Manufacturers had their own data – early era of association rules based on all of the data may have enabled Ford and Firestone to resolve the safety problem before it became a public relations nightmare.

Casually, the issue is to mine association rules across two databases, where the columns in the table are at various sites, splitting each row. One database is assigned the essential, and is the initiator of the protocol. The other database is the responder. There is a join key present in both databases. The rest of the attributes are present in one database or the other, but not both. The goal is to find association rules involving attributes other than the join key.
In business application associations are especially concerned with privacy issue. Most association gather their data from individual party ad their particular need. They find it essential to share data to each other. In such cases every unit need to make sure that privacy of individual party must not disregarded. If Government wants to survey about some medical disease. Individual hospital don't want to revel personal information of its patient, it is against law. This information is regarded private and we want to avoid exposing confidential information of patient. So we must design the system which solve this issue and provide security of data owner.

The Main objective of this project is to provide computational complexity, communication complexity and less storage cost of our association rule mining and frequent item set mining solutions and with the help of Eclat algorithm. With the help of Double Encryption we are processing frequent Itemset mining and association rule mining for high privacy requirements.

The Input for Project is real time dataset such as retail dataset from UCI Machine Learning Repository, the dataset includes transactions and while the single transaction contain Item set, in which each item is comma or space separated. The output of project is Association Rule based on the Mining query sent by the user. The mining query is threshold value which varies from 0 to 1. Frequent item set mining and association rule mining have been employed in applications such as market basket analysis, health care, web usage mining, bioinformatics and Prediction. So the owners of Chains of retail shop or retail shop, Chief Website Developer of an E-commerce business etcetera could make use of this very project.

Section I describes introduction about association rule mining, section II describes literature review, section III includes proposed work where we see the system architecture, and section V concludes the paper.

II. LITERATURE REVIEW

In paper [1] authors proposed a privacy-preserving outsourced frequent item set mining solution for vertically partitioned databases. This allows the data owners to outsource mining task on their joint data securely without compromising on data privacy. Their proposed solutions leak less information about the raw data than most existing solutions. Based on experimental results findings using different parameters and datasets the runtime in each of solutions is only one order higher than that in the best non-privacy-preserving data mining algorithms. Since both data and computing work are outsourced to the cloud servers, the resource consumption at the data owner end is very low.

In paper [2] authors presented an efficient result integrity verification approach that can provide deterministic guarantee for outsourced frequent item set mining. The key idea of the approach is to construct cryptographic proofs of all (in) frequent item sets. They discussed how to optimize the number of proofs to improve the performance. For experimental results generation IBM generator is used to generate four synthetic datasets S1, S2, S3, and S4 of various sizes. Also real-world retail dataset is used and results are calculated on various performance parameters such as time, scalability, etc.

In paper [3] authors tackled issues of privacy preserving association rule mining are addressed here. In particular, privacy preserving algorithms over horizontal and vertical partitioned databases are discussed and results are compared. For vertically partitioned of data more than one attribute is required in transaction while for horizontally partitioned dataset only one attribute is sufficient which generates the association rule which is discussed in the paper also how to partition and merge data horizontally and vertically is discussed by authors.

In paper [4] authors have proposed CRYPPAR, a full-fledged framework for privacy preserving association rule mining based on cryptographic approach over vertically partitioned data. The authors also conducted empirical evaluation on CRYPPAR. The results indicated that the method of building it is efficient and may become a general way to do PPDM in real life. In paper various methodology are used to achieved better performance such as secure scalar product for two parties, partial topology generator for association rule mining , support computation of an Itemset.

In this paper [5], authors concentrate on the situation when the database is distributed vertically, and propose an effective multi-party protocol for evaluating itemsets that preserves the privacy of the individual parties. The proposed protocol is algebraic and recursive in nature, and depends on a recently proposed two-party protocol for the same issue. It is not only appeared to be much faster than similar protocols, additionally more secure. They likewise show a variation of the protocol that is resistant to collusion among parties.

Privacy considerations [6] often constrain data mining projects. This paper addresses the issue of association rule mining where transactions are distributed across sources. Every site holds some attributes of each transaction, and the sites wish to collaborate to identify globally valid association rules. Notwithstanding, the destinations must not reveal individual transaction information. Authors show a two-party algorithm for efficiently discovering frequent itemsets with minimum support levels, without either site revealing individual transaction values.

This paper [7] proposes a protocol for secure mining of association rules in horizontally distributed databases. The present leading protocol is that of Kantarcioglu and Clifton. Their protocol, like theirs, depends on the Fast Distributed Mining (FDM) algorithm of Cheung et al., which is an unsecured distributed variant of the Apriori algorithm. The fundamental ingredients in their protocol are two novel secure multi-party algorithms — one that computes the union of private subsets that each of the associating players hold, and another that tests the inclusion of an element held by one player in a subset held by another.

In paper [8] authors aim is to find frequent items and to develop a global association rules model based on the genetic algorithm (GA). The GA is used due to its inherent features like robustness with respect to local maxima/minima and domain-independent nature for large space search technique to find exact or approximate solutions for optimization and search problems. For privacy preservation of the data, the concept of trusted third party with two offsets has been used. The data are first anonymized at local party end, and then, the aggregation and global association is done by the trusted third party.

In paper [9] authors have focused on the problem of privately mining association rules in vertically distributed Boolean databases. First of all, they proposed an efficient multiparty protocol for evaluating item sets that preserves the privacy of the individual parties. The recommended protocol is algebraic and recursive in nature, and is depend on a recently proposed two-party protocol for the same problem. It is not only displayed to be much faster than similar protocols, but also more protected. Secondly, presented a variant of the extended protocol that is resistant to collusion among parties.

In paper [10] authors have to say, because of the encouraged development in the various fields, such as Cloud Computing. A third party service provider, the server, comes in the frame, when a company, the data owner, who lacks in expertise or the resources, outsources its mining needs. However the data owner thinks both the items and the association rules of the outsourced database as a confidential property. The server stores the data and ships transformed by the data owner. Then the data owner sends mining queries to the server, and the server returns the extracted patterns. From these patterns, the owner recovers the true patterns. Within corporate privacy-preserving frameworks, the problem of outsourcing the association rule mining responsibilities in the outsourced environment is studied. In this paper to improve the security rob frugal algorithm is used in which one to one substitution is performed & then fake transaction are added before sending the original dataset to third party server.

III. PROPOSED SYSTEM

The following Figure 1 shows the proposed system architecture.The system model is included two or more data owners and a server. Every data owner have a private database, the data owners encrypt their private databases prior outsourcing to the third party server. Data owner can also request the server to mine association rules or frequent Itemset from the joint database for their behalf. The (honest but curious) server is tasked with the compiling and storing of databases got from various data owners, at server side data is partially decrypted & then association rules are generated over those data , then at the end those encrypted association rules are send to relevant data owners. At owner side the owner has to decrypt those encrypted rules to get original association rules.

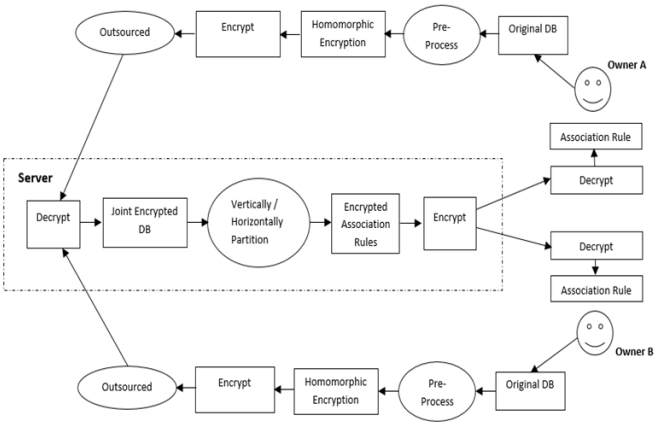

Fig.1. System Architecture

IV. CONCLUSION

From this survey we conclude that, Privacy becomes great challenge for researchers while applying data mining technique on data in various real time applications. Because data owner share their data for mining the knowledge to obtain the mutual benefits. In this survey various recent approaches are studied, which are based on association rule mining for knowledge discovery. Such methods also preserving the privacy of distributed data. Some methods perform the association rule mining on horizontal and vertical distributed model of data. Efficiency and accuracy are challenging task along with privacy. Also recent approaches are compared along with their limitations, which may be useful for further research study.

V. REFERENCES

[1]. Lichun. Li, R. Lu, K. K. R. Choo, A. Datta and J. Shao, "Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases," in IEEE Transactions on

Information Forensics and Security, vol. 11, no. 8, pp. 1847-1861, Aug. 2016.

[2]. B. Dong, R. Liu and W. H. Wang, "Integrity Verification of Outsourced Frequent Itemset Mining with Deterministic Guarantee," 2013 IEEE 13th International Conference on Data Mining, Dallas, TX, 2013, pp. 1025-1030.

[3]. M. N. Kumbhar and R. Kharat, "Privacy preserving mining of Association Rules on horizontally and vertically partitioned data: A review paper," Hybrid Intelligent Systems (HIS), 2012 12th International Conference on, Pune, 2012, pp. 231-235.

[4]. D. H. Tran, W. K. Ng and W. Zha, "CRYPPAR: An efficient framework for privacy preserving association rule mining over vertically partitioned data," TENCON 2009 - 2009 IEEE Region 10 Conference, Singapore, 2009, pp. 1-6.

[5]. D. Trinca and S. Rajasekaran, "Towards a Collusion-Resistant Algebraic Multi-Party Protocol for Privacy-Preserving Association Rule Mining in Vertically Partitioned Data," 2007 IEEE International Performance, Computing, and Communications Conference, New Orleans, LA, 2007, pp. 402-409.

[6]. Vaidya, Jaideep, and Chris Clifton. "Privacy preserving association rule mining in vertically partitioned data." Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2002.

[7]. Tassa, Tamir. "Secure mining of association rules in horizontally distributed databases." IEEE Transactions on Knowledge and Data Engineering 26.4 (2014): 970-983.

[8]. Bettahally N. Keshavamurthy, Asad M. Khan,Durga Toshniwal "Privacy preserving association rule mining over distributed databases using genetic algorithm" Springer-Verlag London 2013 Neural Computing & Application (2013) 22 (Supp 1): S351-S364 DOI 10.1007/s00521-013-1343-9.

[9]. D. Trinca and S. Rajasekaran "Towards a Collusion-Resistant Algebraic Multi-Party Protocol for Privacy-Preserving Association Rule Mining in Vertically Partitioned Data" in IEEE, Orleans, LA, pp. 402-409, 2007.

[10]. F. Giannotti, L. Lakshmanan, A. Monreale, D. Pedreschi, and H. Wang, "Privacy-preserving mining of association rules from outsourced transaction databases," IEEE Systems Journal, vol. 7, no. 3, pp. 385-395, 2013.