

# Analysis of Different Privacy Techniques in Distributed Database

Divyarth Rai, Raghvendra Kumar

Dept of Computer Engineering LNCT Group of College, Jabalpur, M.P., India

divyarthrai7@gmail.com, Raghvendraagrawal7@gmal.com

**Abstract:** - Privacy consideration often constraint data mining data mining paper. This paper addresses the problem of Privacy where transaction is distributed across multiple sites. Each site holds the some transaction, and the sites wish to collaborate to identify globally valid rule. However the sites must not revel individual data. We presents the multi party transaction data who discovering rule sets either sites reveling the individual data.

**Keywords:** - Vertical Partitioning, Secure multi party computation, Privacy Techniques Distributed Database

# 1. Introduction

Data mining could be a method of discovering fascinating information from great amount of information holds on in information, information warehouse or different data repositories. It analyzes information from completely different views and summarizing it into helpful data, data which will be accustomed increase revenue, cut costs, or both. Therefore it's conjointly unremarkably called information Discovery of information (KDD). Data processing involves associate degree integration of techniques from various disciplines like information and warehouse technology, statistics, machine learning, high performance computing, pattern recognition, neural network, information visualization, etc [1].

# 1.1. Distributed Database

In fashionable days distributed information has become an important space of data process. It eradicates several of the short comings of the centralized information and additional work naturally in several organizations that follow suburbanized structure. Distributed information could be a cluster of information that logically belongs to identical system however is meet the sites of a network. It's going to be hold on in multiple computers situated within the same physical location, or is also distributed over a network of interconnected computers. A distributed information system consists of loosely coupled sites that share no physical parts. There are chiefly two approaches to



store a relation R in a very distributed information system:

# 1.2.1Replication

In replication, the system maintains many indistinguishable replicas of identical relation r in several sites and information is accessible during this theme. Similarity is enlarged once scan request is served. Update operations could increase overhead as every website containing the reproduction required being progressive so as to take care of consistency.

## 1.2.2Fragmentation

The relation r is fragmented into many relations r1, r2, r3....rn in such how that the particular relation may be reconstructed from the fragments so the fragments ar scattered to completely different locations. There an essentially two schemes of fragmentation:

Horizontal Fragmentation - Splits the relation by distribution every tuple of R to at least one or additional fragments.

**Vertical Fragmentation -** Splits the relation by rotten the schema R of relation r vertically. During this world each organization's knowledge gets increased exponentially day by day. These large amounts ought to be processed so necessary knowledge remains and unnecessary data ought to be free. Currently a day's technology is increasing and distributed information that's horizontal and vertical information has get image. Since the information or tuples is split into two databases, to use data processing formula they have to be along. Take into account an image wherever two parties owning their classified information ought to run data processing formula while not revealing their whole information. Here arise security or privacy problems.

## 1.3Vertical Database

A vertical database (Shows Figure 1) or vertical partitioning of database is that the method of subdividing the attributes of a relation or world information making fragments. These fragments area unit known as vertical fragments. If the database has r attributes then the information will be vertically splinted into r fragments. These fragments have cluster of attributes and area unit keep in several sites in several region and every cluster of attributes has common geographical properties. If every attribute of world information is mapped into not but



one attribute of the fragment, it's doable to construct the initial relation by connection the fragments along. Party's area unit obligatory to exchange each record or send their knowledge to a committed central place to conduct the information mining computation. Here arises the privacy downside. The matter has been explained thoroughly in [2].



Figure 1: Shows as Vertical Database

# 1.4 Classification Rule Mining

Classification [3] may be data processing or machine learning techniques accustomed predict cluster membership for knowledge instances. Classification rule mining is additionally known as rule based mostly classification. It uses IF-THEN RULES. The IF a part of the rule is thought as rule antecedent and also the THEN part is thought as rule resultant. These rules will be directly being strip-mined from coaching set or indirectly by changing models and extracting the rule.

## 1.5 Decision Tree Classifier

It is a supervised learning methodology that constructs Decision trees from coaching set knowledge. Decision Tree Classifiers [3] area unit effectively utilized in areas like measuring device signal classification, character classification, diagnosis, consultants system etc. a Decision tree [1] may be a flow diagram like tree structure wherever internal node denotes a check on the attribute, every branch represents associate degree outcome of the check and every leaf node represents a category node. The height node in a very tree is that the root node from wherever the tree construction starts. Decision tree classification is in style as a result of the choice tree construction



doesn't need any domain information or parameter settings. Throughout Decision tree construction attribute choice measures area unit used for the attribute that best partitions the tuples into completely different categories. Throughout Decision tree construction, any of the branches could reproduce noise or outliers within the coaching knowledge. Thus tree pruning is finished to boost the formula. It breaks the complicated method process into a set of less complicated choices. Decision tree algorithms have adopted greedy approach within which Decision trees area unit created in top-down algorithmic divide-andconquer manner. Most typically used Decision tree algorithms area unit repetitive Dichotomise ID3, C4.5 and CART. Privacy conserving Decision tree formula has been wont to solve distributed computation drawback wherever the parties altogether build a choice tree over the information set by sharing the mandatory knowledge and preventing the exposure of personal sensitive data.

## 1.6 Privacy Techniques

Privacy has emerged data processing has emerged as a really active analysis space in data processing. Discovering information through a mixture of various databases raises security issue. Though data processing results sometimes don't violate privacy of people, it can't be assured that associate unauthorized person won't access the information is divided over completely different sites and data isn't encrypted, it's not possible to derive new information concerning the opposite sites. Data processing techniques try and determine regularities in knowledge, that area unit unknown and laborious to find by people. Regularities or patterns area unit to be unconcealed over the complete knowledge, instead of on people. But to search out such speech act of patters, the mining method needs to access and use individual data. Techniques mentioned here area unit.

**1. Secure Sum Protocol:** The main aim of this protocol is that over two parties will reason the add of the values every party holds in such the way that no party will learn something concerning the values of the opposite parties.

**2. Secure Intersection Protocol:** Once there are a unit solely two parties, computing the dimensions of the intersection of two sets happiness to every of the parties will result in security issues. The goal is that each one



parties ought to learn the dimensions of set intersection, while not learning concerning the item sets of alternative parties.

**3. Yao Circuit:** Yao introduced in [4] the thought of secure two party computations. He showed that any perform f(x,y), wherever x is that the input of party one and y is that the input of party two is evaluated in secured manner.

**4. X (lnX) Protocol:** X (lnX) protocol is usually used for conserving privacy for two parties. Suppose we've got two parties A and B having price xa and xb severally. The goal of X (lnX) protocol is to offer A and B each a share of ya and yb atomic number severally such

ya + yb = (xa + xb)ln(xa + xb)

**5. Scalar Product Protocol:** This protocol permits quite two parties for computation. The most goal of this protocol is to secure the personal information of different parties such a celebration will apprehend its own result and information solely.

## Conclusion

This paper addresses the problem of computing global support within a scenario of

heterogeneous database. We assume that all sites have the same representation, but each site does not have information on different entities. The goal is to produce rules that hold its input globally while limiting the information shared about each site. Many proposals have been cited to implement secure multi party computation. Secure multi party computation being used in large scale databases which extends to preserve privacy to the private data of different sites. In this paper our focus is based on heterogeneous partitioned distributed data through a popular association rule mining technique.

# REFERENCES

[1]Agrawal, R., et al "Mining association rules between sets of items in large database". In: Proc. of ACM SIGMOD'93, D.C, ACM Press, Washington, pp.207-216, 1993.

[2].Agarwal, R., Imielinski, T., Swamy, A. "Mining Association Rules between Sets of Items in Large Databases", Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-210, 1993.

[3]Agrawal, R., Srikant, R, "Privacy-Preserving Data Mining", In: proceedings of the 2000 ACM SIGMOD on management of data, pp. 439-450, 2000.

[4]Clifton C, Vaidya J, "privacy preserving Association rule mining in vertically partitioned data", ACM SIGKDD, pp. 1-6, 2002.

[5] Franklin, M., Galil, Z. & Yung, M.,"An overview of Secured Distributed Computing". Technical Report CUCS-



00892, Department of Computer Science, Columbia University.

[6] Goldreich, O., Micali, S. & Wigerson, A. ,"How to play any mental game", In: Proceedings of the 19th Annual ACM Symposium on Theory of Computing, pp.218-229, 1987.

[7] Han, J. Kamber, M, "Data Mining Concepts and Techniques". Morgan Kaufmann, San Francisco, 2006.

[8]Kantarcioglu, M., Clifton, C, "Privacy-Preserving distributed mining of association rules on horizontally partitioned data", In IEEE Transactions on Knowledge and Data Engineering Journal, IEEE Press, Vol 16(9), pp.1026-1037, 2004.

[9] Lindell, Y., Pinkas, B, "Privacy preserving Data Mining", In: Proceedings of 20th Annual International Cryptology Conference (CRYPTO), 2000.

[10] N V Muthu Lakshmi, Dr. K Sandhya Rani ,"Privacy Preserving Association Rule Mining without Trusted Site for Horizontal Partitioned database", International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, pp.17-29, 2012.