

**Hiding Rare Association Rules to Maintain Privacy in Distributed Database**Ms. Shweta Gandevia¹, Mr. Mohit Patel²^{1,2}Department of Computer Science and Engineering, PIT Vadodara,

Abstract - Innovation of technology in our world has expanded drastically. People like to use the applications that can easier their work in every day's life. A Distributed Database is a database in which portions of the database are stored at multiple locations within a network. In such Situations, data owners may be concerned with the misuse of data; hence they do not want their data to be mined, especially when these contain sensitive information. Privacy Preserving Data Mining (PPDM) aims to protect data privacy in the course of data mining. In this proposed work Mining the item patterns (Association Rules) using the MIS Tree generation in a distributed manner, thus there is a generation of both frequent as well as rare itemsets from those selecting sensitive itemsets and provide privacy to them to protect the information by providing privacy preserving technique

Keywords - Privacy preserving data mining, Distributed environment, Frequent item sets, Rare Itemsets.

I. INTRODUCTION

A great amount of data are now available in Science, Business, Industry, and many other areas. Due to the rapid advances in Computerization, such data may provide a rich resource for knowledge discovery and Decision Support. For example, when we shop at a Supermarket, the cashier scans the bar codes of items and stores our shopping transaction into a database. The supermarket can find valuable information for marketing by analyzing the sales data shopping transaction database[1].

In order to understand, analyze and eventually make use of the huge amount of data. A multidisciplinary approach, Data Mining is proposed to meet the challenge. Data mining is the process of identifying interesting patterns from large databases[1].

Today there is large amount of data proceed in every day from different sources. that large amount of data stored in different database. this data store in storage devices in from of row data. Data mining is the process of discovering interesting pattern and knowledge from large amount of data. Following Example where data mining techniques are used are Direct mail marketing, bioinformatics, credit card fraud detection , text analysis and market basket analysis. Extracting knowledge from row data, There is some technique to deal with security .Privacy preserving in data mining is one of the technique that deal with security of the knowledge that extracted by data mining technique.

There are various Data Mining Tasks:

- Classification
- Clustering
- Association Rule Mining
- Sequential Pattern Mining
- Regression

II. PRIVACY PRESERVING DATA MINING

The basic idea of PPDM is to modify the data in such a way so as to perform data mining algorithms effectively without compromising the security of sensitive information contained in the data.

PPDM mainly focus on how to reduce the privacy risk brought by data mining operations, while in fact, unwanted disclosure of sensitive information may also happen in the process of data collecting, data publishing, and information (i.e., The data mining results) delivering.

The objective of PPDM is to safeguard sensitive information from unsolicited or unsanctioned disclosure, and meanwhile, preserve the utility of the data.

Techniques for Privacy Preservation**(1) Data Distribution**

Whenever data is in distributed manner, They are Exchanging support counts is enough for mining association rules.

(2) Data Modification

Data modification is used in order to modify the original values of a database that needs to be released to the public and in this way to ensure high privacy protection. It is important that a data modification technique should be in concert with the privacy policy adopted by an organization.

(3) Data or Rule Hiding

Association rule hiding is a new technique on data mining, which studies the problem of hiding sensitive association rules from within the data. Association rule hiding has been playing a vital role in sensitive knowledge preservation when sharing data between enterprises. The aim of association rule hiding is to remove sensitive association rules from the released database such that side effects are reduced as low as possible[14].

(4) Cryptographic methods

Cryptography technique protects data/information from others accessing in a distributed database environment for semi honest model[4].

III. ASSOCIATION RULE HIDING

Association rule mining is a technique in data mining that identifies the regularities found in large volume of data. Such a technique may identify and reveal hidden information that is private for an individual or organization. Privacy preserving association rule mining needs to prevent disclosure not only the confidential or a personal information from the original or aggregated data also to prevent data mining techniques from discovering sensitive knowledge.

Association Rule Hiding

Association rule hiding is widely researched along two principal directions.

- (1) The first variant includes approaches that aims at hiding specific association rules among those mined from the original database.
- (2) The second variant includes approaches that hides specific frequent itemsets from those frequent itemset found by mining original database.

By ensuring that the itemsets that lead to the generation of a sensitive rule become insignificant in the disclosed database, the data owner can be certain that his or her sensitive knowledge is adequately protected from untrusted third parties.

Advantages of Association Rule Hiding

- (1) support for the sensitive item is unchanged. Instead, only the position of the sensitive item set is changed.
- (2) It provide the use of a different technique for modifying the database transaction ,to reduce the confidence of sensitive rules without making any change in the sensitive item.

DisAdvantages of Association Rule Hiding

This approach tries to hide every single rule from a given set of rules without checking if some of the rules could be pruned after modification of some transactions from the set of all transactions. This approach hides rules having sensitive items either in the right side or in the left side.

The common approaches used in association rule hiding algorithms are :

(1) Heuristic approaches

The Heuristic approaches are used to modify the selected transactions from the database for hiding the sensitive data.

(2) Border-based approaches

The Border-based approaches is the sensitive rule hiding can be done through the modification of the origin a borders in the lattice of the frequent and the infrequent patterns in the data set.

(3) Exact approaches

The Exact approaches are non-heuristic algorithms which envisage the hiding process as a constrain satisfaction problem that may be solved using integer programming or linear programming.

IV. RELATED WORK

There are following paper refer as the literature survey. This all paper included the advantages and limitation.

A. Mining Rare Association Rules in a Distributed Environment using Multiple Minimum Supports[6]

This paper proposes an abstract algorithm for discovering rare association rules in a distributed environment. It utilized the idea of using statistic percentile to produce multiple minimum supports to mine rare association rules. The proposed algorithm has been implemented using JAVA and evaluated by comparing with the Optimized Distributed Association rule Mining (ODAM) algorithm and the Apriori with Multiple Support Generating by statistic Percentile

threshold (Apriori MSG-P) algorithm to show that the algorithm can discover more rare association rules with an optimized communication cost. The Proposed Algorithm is AprioriMSD (Apriori with multiple minimum support in a Distributed Environment). Specifically, AprioriMSD produces a higher number of frequent itemsets than Apriori MSG-P. Additionally, more frequent itemsets can be found on a higher size of itemsets. These results imply that frequent itemsets derived by each size of itemsets can be either symmetrical or skewed. Therefore, AprioriMSD is an optimal distributed data mining algorithm of rare association rules because it recognizes skewed datasets and handles such datasets appropriately. And After Generating Association Rules there is no providing privacy to protect the information.

B. A Novel Approach for Finding Rare Items Based on Multiple Minimum Support Framework[13]

In this paper, an effort is made to analyze the complete set of rare items for finding almost all possible rare association rules from the dataset. The Proposed approach makes use of Maximum constraint model for extracting the rare items. A new approach is efficient to mine rare association rules which can be defined as rules containing the rare items. In this Paper the approach utilizes a tree structure to ascertain the rare items.

C. An Improved Multiple Minimum Support Based Approach to Mine Rare Association Rules[11]

In this paper, Authors propose an improved approach in which minsup is fixed for each item based on the notion of “support difference”. The proposed approach assigns appropriate minsup values for frequent as well as rare items based on their item supports and reduces both “rule missing” and “rule explosion” problems.

D. Privacy preserving heuristic approach For Association Rule Mining in Distributed Database[14]

In this paper, There is mining of Association rule for the distributed database using Apriori Approach. The challenges of data mining is to secure the confidentiality of sensitive patterns when releasing database of third parties. Privacy Preserving in this paper is used as hide association rule. Association rule hiding algorithm sanitize database such that certain sensitive association rule cannot be discovered through Association rule mining techniques. In this Paper there is used the Heuristic approach in Data Distortion Technique. In proposed algorithm there is a use of MDSRRC algorithm, which hides multiple R.H.S items for association rules those are generated for distributed database.

V. PROPOSED WORK

In the proposed work there is a generation of both frequent as well as rare itemsets. Main objective is mining the rare itemsets from the different locations and provide the privacy for protect the association rules.

There are following steps in Propose algorithm in Distributed Environment:

Step 1 : Take Distribute Database(i.e. D1, D2, D3)

Step 2 : Calculate SC of each items

Step 3 : Calculate MIS of each item using Following Equation(2) [7]

$$\begin{aligned} \text{MIS}(ij) &= S(ij) - SD \text{ when } (S(ij) - SD) > LS \\ &= LS \text{ otherwise} \end{aligned} \quad \dots\dots\dots(1)$$

Step 4 : Construction of MIS-Tree.

- Initially, The items in the transactional dataset are sorted in descending order of their MIS values.
- Create a root node labeled “null” is created .
- Each transaction in the transactional dataset is scanned and MIS-tree is updated by 1 according to items’ SC.

Step 5 : Each site send their MIS Tree to Server.

Step 6 : Each MIS tree is converted into transactional dataset

Step 7 : Combine all transactional dataset into single transactional dataset.

Step 8 : Further, Calculate SC and MIS of each items

Step 9 : Construct global MIS-tree from transactional dataset.

Step 10 : Compare each items’ SC with MIN of MIS Tree

Step 11 : Delete Item’s Node whose SC < MIN

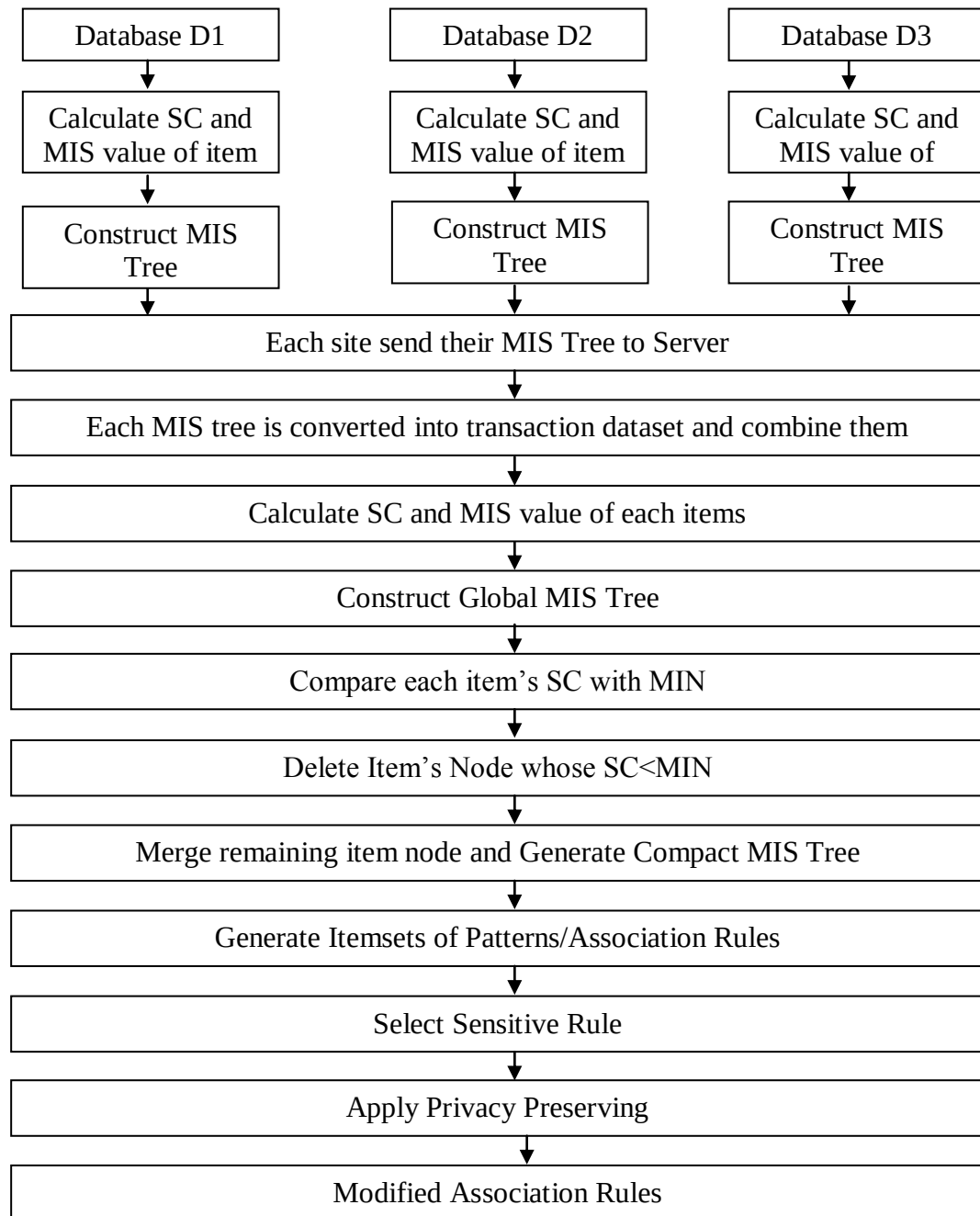
Step 12 : Merge remaining Item’ node and generate Tree called Compact MIS Tree.

Step 13 : Generate Itemsets of Patterns or Association Rules.

Step 14 : Select Rare itemsets as a sensitive Rule

Step 15 : Apply Privacy Preserving to sensitive rules

Step 16 : get modified Rules



In Distributed Database System , Datasets are stored at more than locations and from them we have to generate Frequent and rare itemsets (association rule) by locally generated and from all the sites we have to combine them at any one location at that time we have to protect them so we can provide privacy preserving technique at global site and protect the generated association rule by association rule hiding.

References

- [1] Yongjian Fu, "Data mining : Tasks, Techniques, Applications", Department of Computer Science, University of Missouri-Rolla.
- [2] Jiawei Han and Micheline Kamber, "Data Mining: Concepts and Techniques, 2nd Edition".
- [3] G.R. Bamnote, Himanshu Joshi, " Distributed Database: A Survey", International Journal Of Computer Science And Applications. Vol. 6, No.2, ISSN: 0974-1011, Apr 2013.
- [4] N V Muthu lakshmi, " Privacy Preserving Association Rule Mining in Horizontally Partitioned Databases Using Cryptography Techniques". (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (1), 2012, 3176 – 3182.
- [5] K. Saranya, K. Premalatha, S.S. Rajasekar, "A Survey on Privacy Preserving Data Mining", IEEE SPONSORED 2ND INTERNATIONAL CONFERENCE ON ELECTRONICS AND COMMUNICATION SYSTEM (ICECS 2015), 2015 IEEE.
- [6] Jutamas Tempaibookkul, "Mining rare association rules in a distributed environment using multiple minimum supports "IEEE 2013.
- [7] R N Yadawad, RBV Subramanyam, U P Kulkarni "Mining of Rare Itemsets in Distributed Environment". International Journal of Computer Applications (0975 – 8887) Volume 105 – No. 6, November 2014.
- [8] Tamir Tassa, " Secure Mining of Association Rules in Horizontally Distributed Databases " IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 4, APRIL 2014.
- [9] Kantarcioglu M, Clifton C, "Privacy-Preserving distributed mining of association rules on horizontally partitioned data", IEEE Transactions on Knowledge and Data Engineering Journal, IEEE Press, Vol 16 (9), pp.1026-1037, 2004.
- [10] Nikunj H. Domadiya, Udai Pratap Rao, " Hiding Sensitive Association Rules to Maintain Privacy and Data Quality in Database " 3rd IEEE International Advance Computing Conference IEEE 2013.
- [11] Uday Kiran, R., Krishna Reddy, P. 2009: An Improved Multiple Minimum Support Based Approach to Mine Rare Association Rules. In: IEEE Symposium on Computational Intelligence and Data Mining, pp. 340–347
- [12] Taweechai Ouypornkochagorn, "Mining Rare Association Rules on Banpheo Hospital (Public Organization) via Apriori MSG-P Algorithm". ECTI TRANSACTIONS ON COMPUTER AND INFORMATION TECHNOLOGY VOL.6, NO.2 November 2012
- [13] U Bhatt, P Patel . "A Novel Approach for Finding Rare Items Based on Multiple Minimum Support Framework". ELSEVIER, 3rd International Conference on Recent Trends in Computing 2015 (ICRTC-2015).
- [14] Bhoomika R Mistry, Amish Desai, "Privacy preserving heuristic approach For Association Rule Mining in Distributed Database", IEEE Conference, ICIECS'15,
- [15] <http://www.hindawi.com/journals/mpe/2013/210405/>