Volume 5, Issue 12, December 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Privacy Preservation by Hiding Sensitive Association Rules

Shital Mane, Diksha Saykad, Mrunali Sorte, Shraddha Solkar, Prof. Geeta Navale
SITS, Narhe, Pune, Maharashtra,
India

Abstract— - Now-a-days there is a huge amount of data present in the database which holds a lot of information. This information can be extracted from it using data mining process. One of the techniques of data mining is association rule mining. Hiding of Association rules is one of the major problems in the data mining domain. The association rules hold many secrets. So before publishing, these rules must be hidden. Sensitive information must be hidden, since revealing the secret or important association information may cause problems. In our paper, Privacy preservation is done by hiding sensitive association rules. During hiding of sensitive association rules, false rules are not generated and minimum modification degree is achieved and information is not lost.

Keywords— Data mining, Association rules, Privacy Preservation, Sensitive Association Rules.

I. INTRODUCTION

In recent years, privacy-preserving data mining has been studied extensively, because of the large amount of sensitive information on the internet. A number of algorithmic techniques have been designed for privacy-preserving data mining.
Data mining techniques have been developed in many applications. However it also causes a threat to privacy. We investigate to find an appropriate balance between a need for privacy and information discovery on association patterns.
We propose an innovative technique for hiding sensitive patterns. In our approach, different algorithms are explained for hiding sensitive information. Also, a set of experiments is performed to show the effectiveness of our approach.
Data mining technology has given us new capabilities to identify correlations in large data sets. This introduces risks when the data is to be made public, but the correlations are private. We introduce a method for selectively removing individual values from a database to prevent the discovery of a set of rules, while preserving the data for other applications. It maintains the privacy preservation when the data shares between the different organizations. And also it hides the sensitive information.

II. LITERATURE SURVEY

Data mining is used to extract useful and knowledgeable information from large amount of data. It is the process of finding patterns in large relational databases. In 1989 the term KDD means Knowledge discovery in databases is developed by Gregory Piatetsky-Shapiro. The term KDD became more famous in artificial intelligence and machine learning community. [1]
One of the techniques of data mining is association rule mining. The association rule concept was popularized in 1993 because of article of Agrawal. That article was published in 1993. This article is one of the most cited papers in data mining. [3]
Association rules are represented by if/then statements. These statements help to find out relationships between seemingly unrelated data in a relational database or other information repository. An example of an association rule would be "If a customer buys milk, he is 85% likely to also purchase sugar." [2]. There are various methods or algorithms are available to generate association rules and to hide sensitive association rules. These algorithms are as follows.
Apriori algorithm is used to generate the association rules. In association rule hiding method the single antecedent and consequent are selected. Association rules are useful in market basket database. This rules show customer behavior in market basket database. [4]
MDSRRC stands for Modified Decrease Support of R.H.S item of Rule Clusters. This algorithm hides sensitive association rules with multiple items in consequent and antecedent. This algorithm is better than DSRRC algorithm. [5]
PPARM algorithm and IMBA algorithm is combine to generate new algorithm. With the help of this newly generated algorithm, a greater degree of hiding can be made while a less degree impact for the non sensitive rules made true. This algorithm provides high privacy protection. [6]
Exact border based approach is used to hide sensitive frequent item set. This approach gives optimal solution. [7]
By using secure computation and RSA encryption technology, avoid data leakage which cause by data sharing. For hiding sensitive rules ISL and DSR algorithm is used. [8]
A number of heuristic techniques are developed for both effective and efficient hiding sensitive association rules. Border revision theory that uses the anti monotone property of the frequent item sets. This property used to define the border line between what must be hidden and what to be protected. [9]

The Apriori algorithm is improved to update classical apriori algorithm. In classical Apriori algorithm may need to generate a large number of candidate generations. Each time algorithm has to check newly generated candidate generations are frequent sets. The manipulation with redundancy lead to high frequency in querying so large amount of resources expended in time or space. The improved apriori algorithm decrease the number of database scanning. This algorithm requires less time to generate frequent item sets as compared to classical Apriori algorithm. Strong rules which are generated by this algorithm are hidden by applying decreasing support confidence algorithm. [10]

DSRRC stands for Decrease Support of R.H.S Item of Rule Clusters. This algorithm hides many sensitive association rules at a time while maintaining database quality. This proposed algorithm hides only rules that contain single item on R.H.S of the rule. But it is more efficient than other heuristic approaches. [11]

Branch and Bound algorithm is used to hide sensitive association rules. This algorithm hides sensitive association rules without undesired side effects. The hiding of sensitive association rules is done without generating spurious rules [12]

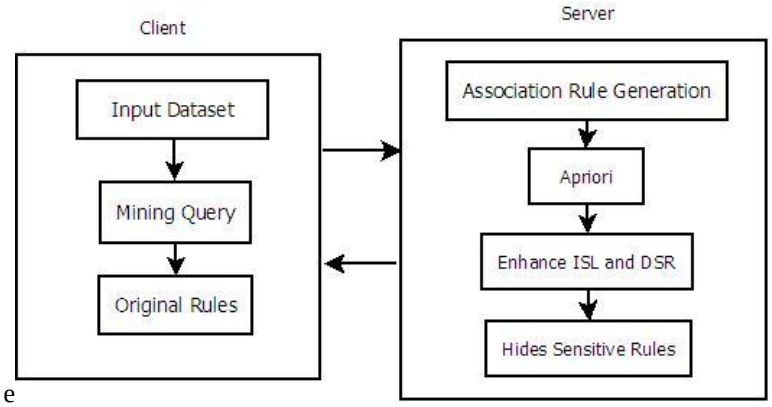
SRM stands for sensitive Rule Miner. This algorithm generates association rules with certain structure. These association rules are used to reveal and explain relationships. This algorithm detects threads. [13]

III. COMPARISON TABLE

Table 1: Comparison Table

YEAR	ALGORITHM	HIDING	INFORMATION LOSS	DEGREE OF MODIFICATION	FALSE RULES GENERATION
2013	MDSRRC	yes	-	-	-
2013	Apriori	yes	-	-	-
2013	SRM	yes	-	-	-
2013	Hiding Approaches	yes	-	-	-
2012	Branch and Bound	yes	yes	-	-
2012	Apriori	yes	-	yes	yes
2010	SWTA	yes	yes	-	-
2010	DSRRC	yes	-	-	-
2010	Apriori	yes	yes	-	yes
2009	Border Based Approaches	yes	-	-	-

IV. GENERIC FRAMEWORK



e

Figure1: Generic Framework

V. PROBLEM EXPLINATION

This project uses Apriori algorithm on static dataset to generate Association Rules. Then user give query to hide some association rules .Another algorithm is used to hide these association rules. Hiding sensitive association rules is done by considering information loss, degree of modification, false rules parameter. During hiding of sensitive association rules, information is not lost, minimum degree of modification is obtained, false rules are not generated.

VI. ALGORITHM

1. Problem Description: This Algorithm hides sensitive association rules.

Input: Static dataset which contain information.

Output: Dataset in which sensitive association rules are hidden.

Step 1: Apriori algorithm is applied on static dataset to generate association rules.

Step 2: User give mining query. Query tells which rules he want to hide. For example one thousand rules generate after applying apriori algorithm and user want to hide hundred rules then he give query that he want to hide hundred rules. These rules called as sensitive association rules.
Step 3: Another algorithm is applied to hide sensitive association rules.

2. Modified DSR Algorithm

1. Sort the given database according to Relevance count in descending order
2. Calculate dsr_count = CU – Cx × MCT + 1[2]
3. Find T = t in D | t fully support U;
4. Choose the first transaction t from T;
5. While (dsr_count > 0)
   5.1 Modify t by putting 0 instead of 1 for RHS item;
   5.2 Check for loss of rule if yes then go to step 5.4
   5.3 Remove and save the transaction t from T. Change the relevance count accordingly and decrease the value of dsr_count by 1
   5.4 Consider next transaction t End While
6. Compute the Confidence of U;
7. If dsr_count is not equal to 0, then h cannot be hidden;

VII. CONCLUSIONS

In our paper, static dataset is used to store information. Apriori algorithm is applied on static dataset to generate association rules. Then user gives query to hide some association rules that rules are called as sensitive association rules. Then hiding of sensitive association rules takes place by considering information loss, degree of modification, false rules parameter. While a hiding sensitive association rule, information is not lost, minimum degree of modification is achieved and false rules are not generated.

REFERENCES

[1] https://en.wikipedia.org/wiki/Data_mining#Process
[2] http://searchbusinessanalytics.techtarget.com/definition/association-rules-in-data-mining
[3] https://en.wikipedia.org/wiki/Association_rule_learning
[4] S. Kasthuri, T. Meyyappan, “Detection of Sensitive Items in Market Basket Database using Association Rule Mining for Privacy Preserving”,2013.
[5] Nikunj H. Domadiya and Udai Pratap Rao, “Hiding Sensitive Association Rules to Maintain Privacy and Data Quality in Database”,2013.
[6] Sun Wei, Wang Yonggu “Association rule mining algorithm based on Privacy preserving”,2010.
[7] Aris Gkoulalas-Divanis and Vassilios S. Verykios, “Exact Knoweledge Hiding by Database Extension”,2009.
[8] Tinghuai Ma, Sainan Wang and Zhong Liu ,“Privacy Preserving Based on Association Rule Mining”,2010.
[9] Vassilios S. Verykios,"Association Rule Hiding Methods",2013.
[10] PG Scholar Mr. Suraj P. Patil, Assoc Prof T. M Patewar, "A Novel Approach For Efficient Mining and Hiding of Sensitive Association Rule",2012.
[11] Chirag N. Modi, Udai Pratap Rao, Dhiren R. Patel, "Maintaining Privacy and Data Quality in Privacy Preserving Association Rule Mining",2010.
[12] Chieh-Ming Wu and Yin-Fu Huang, "Privacy Preserving Association Rules by Using Branch-and Bound Algorithm",2012.
[13] Irene Diaz, Luis J. Rodriguez-Muniz, Luigi Troiano, "On Mining Sensitive Rules to Identify Privacy Threats",2013.

