A IMPROVED PRIVACY PRESERVING ALGORITHM US-ING ASSOCIATION RULE MINING IN CENTRALIZED DATABASE

Mr. Chirag Bhabhor[1]
[1]M.E. Student
[1]Department Of Computer Science & Engineering
[1]Kalol Institute Of Technology and Research, Kalol, Ahmedabad, Gujarat.

Abstract--- The main problem is that to hide sensitive information, including personal information, fact or even patterns which are generated by any algorithm of data mining from the others. In order to focusing on privacy preserving association rule mining, the simplistic solution to address the problem of privacy is presented. To overcome these problems, we propose a algorithm named improved Privacy Preserving Algorithm using Association Rule Mining which is based on the random Perturbation technique which is best in efficiency and performance. This method is suitable to the any type of data. Our algorithm is a good way to apply data mining techniques with security that hides our logical in-stances from others.

Keywords : Privacy Preserving Association Rule Mining

I. INTRODUCTION

Association rule techniques are used for data mining if the goal is to detect relationships or associations between specific values of categorical variables in large data sets. There may be thousands or millions of records that have to be read and to extract the rules for, but the question is what will happen if there is new data, or there is a need to modify or delete some or all the existing set of data during the process of data mining. In the past user would repeat the whole procedure, which is time-consuming in addition to its lack of efficiency. Data mining is the task of discovering interesting and hidden patterns from large amounts of data where the data can be stored in databases, data warehouses, olap (on line analytical process) or other repository information . It is also defined as knowledge discovery in databases (KDD) .

II. VARIOUS NOVEL ALGORITHMS

A. *Data perturbation based association rule*

The algorithms can be described as the following one. Let D be the source database, R be a set of significant association rules that can be mined from D, and let Rh be a set of rules in R.How can we transform database D into a database D', the released database, so that all rules in R can still be mined from D', except for the rules in Rh. The heuristic proposed for the modification of the data was based on data perturbation, and in particular the procedure was to change a selected set of 1-values to 0-values, so that the support of sensitive rules is lowered in such a way that the utility of the released database is kept to some maximum value. Therefore, the key question of this algorithm is how to put D into D' with the use of heuristic thought. describes a technique that uses a queue and a random number generator to generate the items so that each item has an approximately equal frequency of being added to transactions. And the work avoids the question that it is hard for to utilize existing tools for association rule mining.

B. *Data blocking based association rule*

1) *Replacement-based techniques*

After original data is replaced the value of some data with the unknown value, the support and confidence of sensitive association rules will not be able to determine, which may be a range of arbitrary values. The paper in [17] discusses specific examples with the use of an uncertain symbol used in association rule mining, in which case the support and confidence interval are used to support and confidence interval to replace.

2) *Anonymity techniques*

Agrawal et al. improve on the distribution reconstruction technique presented in [18] by using the Expectation Maximization (EM) method. The authors claim that EM is more effective than the currently available technique in terms of the level of information loss. Finally, they propose novel metrics for the quantication and measurement of privacy preservingdata mining algorithms.

C. *Data hiding*

The detail of the method is described as follow:

Encrypt step:

(1) Data Miner produce public encryption key C-e;
(2) Data Miner produce key-pair of each site (ei,di);
(3) Send ei to each site, di to Disturb Center;
(4) Disturb Center send ID to each site;
(5) Each site Encrypt their frequent itemset to ID ei(Ce(frequent itemset));

Decrypt step:

(1) Each site send their data ID ei(C-e(frequent itemset)) to Disturb Center;
(2) Disturb Center use di decrypt the data from each site and remove the ID;
(3) Disturb Center disrupts the order of the data ,send C-e(frequent itemset) to Data Miner;
(4) Data Miner decrypts the data, getting frequent.

D. *Rule hiding algorithms*

This is the technique which we are going to tackle for our proposed work for providing privacy in association rule mining. Rule hiding helps in hiding the sensitive information of the large data sets having large itemsets. The algorithm provides two approaches. The first one focuses on hiding the rules by reducing the minimum support of the itemsets that generated these rules (i.e., generating itemsets).

The second one focuses on reducing the minimum confidence of the rules.

III. THE ALGORITHM

The entire system architecture consists of five phases: 1) Check for Authentication. 2) Encoded the data by using the random Perturbation technique 3) On the basis of decryption key we read the transaction 4) Perform Pruning 5) On the basis of decryption key we generate association rules Our algorithm is a good way to apply data mining techniques with security that hides our logical instances from others. Our algorithm shows good performance in different operating environment.

Algorithm: IPPM (Improved Privacy Preserving Mining)

Input: A. Set of rules to hide the data values

B. The source database

C. A Key for visualizing the authentication.

Output: D. The database (DB) transformed so that the set of rules are properly applied and produce the result with security.

IPPM(R, DB, Key)

Begin

A. Check the Authentication

a. Enter uid & pwd

b. If (uid == udb && password == pdb)

{

c. Welcome in the database

SIPM (DB)

User(entry)

{

Log(id)

}

}

d. Else

{

Not an authorized user

}

Exit (0)

B. IPPM (DB)

a. While (object. read ()! =-1)

{ [Start Reading]

[Generate Tokens]

TK1, Tk2…..Tkn

[Token is generated according to the alphabet entered]

If(,)

{ TK1, Tk2……Tkn }

Else

{

[Enter the character]

String a=Object.nextLine();

STK1, STk2……STkn

}

26

}

b. [compute the occurrences]

For i=1 to n iterations do

{

Itemset[i]=count;

Count++;

}

c. [Enter the minimum support]

Check for authentication again

Enter the min-sup key If(min-sup==msdb)

{

Rule hiding by support reduction

}

Else

{

[Enter the value again]

}

C. rule hiding by support reduction

Begin

1. Sort Lh in descending order of size and minimum support of the large

itemsets Foreach Z in Lh

2. Sort the transactions in Tz in ascending order of transaction size

3. N_ iterations = |TZ|- (MST - SM) x |D|

For k = 1 to N_iterations do

{

4. Place a ? mark for the item with the largest minimum support of Z in the

next transaction in Tz

5. Update the supports of the affected itemsets

6. Update the database, D

}

}

End

IV. PERFORMANCE ANALYSIS

TID	Items	Encrypted Items
1	abcde	12345
2	acd	134
3	abdfg	12467
4	bcde	2345
5	abd	124
6	cdefh	34568
7	abcg	1237
8	acde	1345
9	acdh	1348

Frequent Itemsets with Support Count
1:7,2:5,3:7,4:8,5:4, 12:4,13:5,23:3, 14:6, 24:4, 34:6, 35:4, 45:4,124:3, 134:4, 345:4

Fig.1: performance analysis

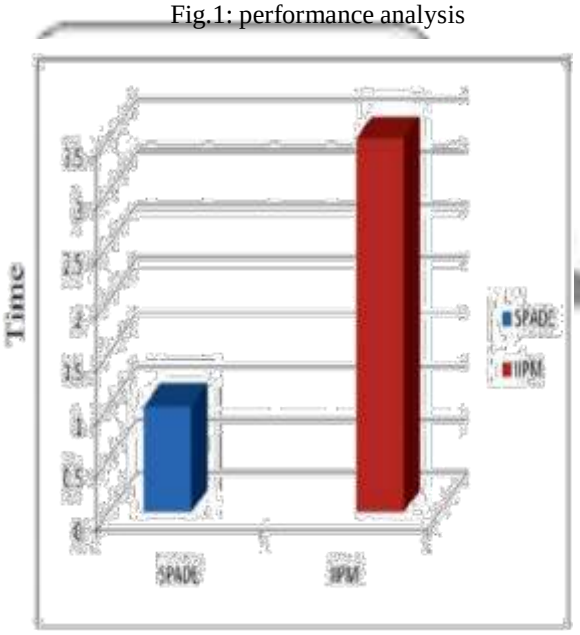

Fig.1: graph

V. CONCLUSION

In this article, a brief description of classification rule algorithm and associated privacy is defined. Performance study shows that IIPM outperforms all other algorithms.

REFRENCES

[1] IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING,VOL. 18, NO. 11, NOVEMBER 2006.

[2] CAEP:CLASSIFICATION BY AGGREGATING EMERING PATTERN;Guozhu Dong,Xiuzhen Zhang ,Limsoon Wong and Jilyan Li,In Discovery Science 99,November 5,1999.

[3] CMAR: ACCURATE AND EFFICIENT CLASSIFICATION BASED ON MULTIPLE CLASS-ASSOCIATION RULES; Wenmin Li Jiawei Han Jian Pei School of Computing Science, Simon Fraser University.

[4] USING GENERAL IMPRESSIONS TO ANALYZE DISCOVERED CLASSIFICATION RULES[Department of Information Systems and Computer Science National University of Singapore Lower Kent Ridge Road, Singapore]

[5] SCALABLE MINING FOR CLASSIFICATION RULES IN RELATIONAL DATABASES[Data Management Department IBM T. J. Watson Research Center 19 Skyline Drive Hawthorne, NY 10532, USA, IBM Silicon Valley Lab 555 Bailey Avenue San Jose, CA 95141, USA Purdue University 150 North University Street West Lafayette, IN 47907, USA]

[6] UNDERSTANDING THE CRUCIAL DIFFERENCES BETWEEN CLASSIFICATION AND DISCOVERY OF ASSOCIATION RULES-A POSITION PAPER[*SIGKDD Explorations,July 2000*]

[7] INTEGRATING CLASSIFICATION AND ASSOCIATION RULE MINING[Department of Information Systems and Computer Science National University of Singapore Lower Kent Ridge Road, Singapore 119260]

[8] AN EFFICIENT ALGORITHM FOR CLOSED ASSOCIATION RULE MINING[Mohammed J. Zaki and Ching-Jui Hsiao Computer Science Department Rensselaer Polytechnic Institute, Troy NY 12180]