ISSN 2319 - 5975

Volume 4, No.3, April - May 2015

International Journal of Networks and Systems

Available Online at http://www.warse.org/ijns/static/pdf/file/ijns01432015.pdf

Hiding Of Sensitive Association Rules With MDSRRC Algorithm for Preserving Privacy In Database

Sangram Kakade [1], Dipak Gaikwad [2], Ajay Mahadik [3], Pralhad Jadhav [4]

[1] Bachelor of Engineer(computer), India, sskakade5549@gmail.com
[2] Bachelor of Engineer(computer), India, gaikwadipak555@gmail.com
[3] Bachelor of Engineer(computer), India, ajaymahadik60@gmail.com
[4] Bachelor of Engineer(computer), India, pralhadjadhav777@gmail.com

ABSTRACT

In day to day life decision making data mining plays an important role. In concern with privacy problem by implementing an association rule hiding algorithm for maintaining data privacy in that mining which would be well organized in providing secrecy and look up the performance at the time where massive database are in used. This paper focuses on the research in hiding sensitive association rule to maintain privacy and data quality in database. In the paper we have proposed heuristic methods based algorithm named MDSRRC(Modified Decrease Support of RHS item of Rule Clusters)to hide sensitive association rules with multiple items in consequent (RHS) and antecedent (LHS). The algorithm selects the items and transactions based on certain criteria which modify transactions to hide sensitive information

Key words : sensitive patterns, privacy preserving data mining, MDSRRC.

1. INTRODUCTION

Association rule mining techniques are wide employed in data mining to find the relationship between items and items sets. The corporate and many government organizations share their historical data for mutual benefit to find out some useful relations or information for some decision making purposes and improve their business schemes. But this database contain some private information and which the organization does not need to reveal.

So before revealing database some sensitive patterns[1] should be hidden and to resolve this issue PPDM(Privacy Preserving Data Mining)[3]. plays an important role in preserving data privacy in database. The proposed algorithm use hiding strategies which are based on decrease support and confidence of the sensitive rule. The proposed algorithm is enhanced variety of DSRRC[2]. DSRRC could not hide the association rule with multiple items in LHS and RHS. To overcome this limitation we have proposed an algorithm MDSRRC which hides the count of items in RHS of the sensitive pattern. It modifies the minimum number of transaction to hide maximum sensitive rules and maintains data quality.

2. PROBLEM DESCRIPTION

Hiding of sensitive patterns as: modify original database D to sanitized database D’ so mining technique fails to mine sensitive patterns from database while all non-sensitive patterns remains visible. A general definition of problem is as:

Given transactional Database D, MCT(Minimum Confidence Threshold), MST(Minimum support Threshold), generate all association rules R from database D, $SR \subseteq R$ sensitive rule , which database owner wants to hide. Problem is to find sanitized database D’ such that any mining technique fails to mine SR set while all non-sensitive rules can be mined.

The goal of association pattern hiding is satisfy the following conditions

1. Sanitized database must hide all sensitive rules
2. Sanitized database must support mining of all non-sensitive rules
3. Sanitized database must not generate any new rules, not present In D

Here we try to find optimized solution to problem means we can’t remove all side effect rather we try to decrease the side effects on database by affecting minimum number of transaction.

3. LITERATURE REVIEW AND THEROTICAL BACKGROUND

Table 1. Notation And Definition

D	Original Database
D’	Sanitized Database
R	Association rule

	generated from original database
SR	Sensitive association rule SR∊R
Ti	The ith transaction in database
I	Set of distinct item in database
IS	IS={is0,is1...isk} k=<n, Set of items present in consequent of sensitive rules with decreasing order of their frequency in consequent of sensitive rules
IS0	Items with highest count in consequent of sensitive rule
MCT	Minimum confidence threshold
MST	Minimum support threshold
RHS	Antecedent of an association rule
LHS	Consequent of an association rule

In table 1, we show the notation used in this paper, Mining an association pattern with support and confidence is follow: The support of rule A→B is calculated using the following formula: support(A→B)=|A∪Y|/|D|, where |D| define the total number transaction in the database D and |A∪B| is the number of transaction which support item set AB. The confidence of rule is calculated using formula: Confidence (A→B) =|A∪B|/|A|, where |A| is number of transaction which support item set A. The rule A→B is mined from database if support (A→B)>=MST and confidence (A→B)>=MCT.Association rule hiding method can be classified into heuristic based approaches, reconstruction based, border based, exact approach and cryptography based tactics

Data Distortion: changes the item value by new value in database matrix. It '0' to '1' or '1' to '0' for selected items in selected transaction

Data Blocking: in its place of injecting or removing items from database it replaces '1' to '0' with '?' in selected transaction.

4. PROPOSED MDSRRC ALGORITHM

Some important terms are use in the proposed algorithm is as follow:

1.Sensitivity of the items: how many sensitive rules contain these items?

2.Sensitivity of transaction: is the total number of sensitivities of all sensitive items contains in that transaction. The proposed algorithm start with mining the association rule from original database 'D' using association rule mining technique e.g. Apriori algorithm. Then user identifies some rules as sensitive rules from then rules generated by the association rule mining algorithm. Then algorithm calculates occurrences of each item in R.H.S of sensitive rules. Now algorithm finds IS={is0,is1......isk} k<=n ,by arrange those items in decreasing order of their counts. After that sensitivity of each item is calculated then sensitivity of each transaction is calculated. Then transactions which support is0 are store in descending order of their sensitivities.

Now rule hiding process starts by selecting first transaction from the sorted transaction with higher sensitivity, delete item is0 from that transaction. Then update support and confidence of all sensitive rules and if any rules have support and confidence bellow MST and MCT value respectively then delete it from SR. Then update sensitivity of each item, transaction and IS. Again select transaction with higher sensitivity and delete is0 from it. This process continues until all sensitive rules are hidden.

MDSRRC Algorithm

INPUT: MCT(Minimum confidence threshold)MST(Minimum support threshold), Database D.

OUTPUT:
Database D with all sensitive rules are hidden.

```
1.ApplyApriori[2]. algorithm to given database D. Find out all possible association rule R.
2.Choose set of rules SR∈R as sensitive rules.
3.Determine sensitivity of each item J∈D.
4. Find out sensitivity of each Transaction.
5.Count occurrences of each item in R.H.S of sensitive rule, find IS={is0,is1...isk} k>=n, by arranging those item in R.H.S of sensitive rules, Find same count then sort these in descending order of their actual support count.
1.Select the transactions which support is0, then sort them in descending order of their sensitivity. If two transactions have same sensitivity then sort those in increasing order of their length.
While (SR is not empty)
{
```

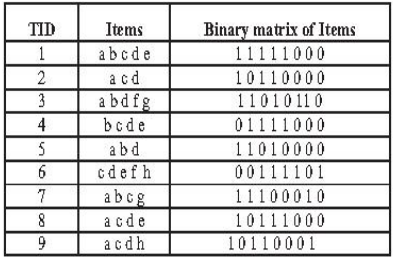
```
Start with first transaction from sorted transaction,
Delete item is0 from that transaction.
For each rule r ∈ SR
}
Renew support and confidence of the rule r.
If(support of r<MST or confidence of r<MCT)
{
Remove Rule r from SR
Update sensitivity of each item.
Update IS (This may change is0).
Update the sensitivity of each transaction.
Select the transaction which are support is0,
Sort those in descending order of their sensitivity.
}
Else
{
Take next transaction from sorted transactions,
}
}
}
End.
```

K.

5.EXAMPLE

To know the MDSRRC algorithm below Example is illustrated. The Transactional Database D is shown in table 2.with 3 as MST and 40% as MCT, the possible generated association rule by apriori algorithm a→b ,b→a, a→c, c→a, a→d, d→a ,b→c, c→b, b→d, d→b, c→d, d→c, c→e, e→c, d→e, e→d, a→cd, c→ad, a→d,d→ac, ad→c, cd→a, c→de, d→ce, cd→e, e→cd, ce→d, de→c, a→bd, b→ad, ab→d, ad→b and bd→a let, the Database user specify the rule a→bd,a→cd, and d→a as a sensitive rule.The transaction also with its sensitive is shown in below table 3. The sensitivity of a=3,b=2,c=1,d=3. Now algorithm identify the frequency of each items are present in R.H.S(right hand side) of sensitive rules. Here frequency of d=3,a=1,b=1,c=1 so, IS={a,b,c,d} in this example item d is selected is**0.**Then it sorting transaction which supports is**0** in descending order of there sensitivity after that select transaction with maximum sensitivity, then truncate is**0** from give transaction with maximum sensitivity. Now total sensitive rules are hidden then final sanitized Database is shown In table 5.

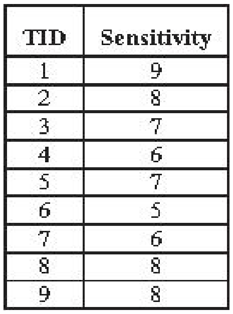
Table 2:- Transactional Database (D)



TID	Items	Binary matrix of Items
1	abcde	11111000
2	acd	10110000
3	abdfg	11010110
4	bcde	01111000
5	abd	11010000
6	cdefh	00111101
7	abcg	11100010
8	acde	10111000
9	acdh	10110001

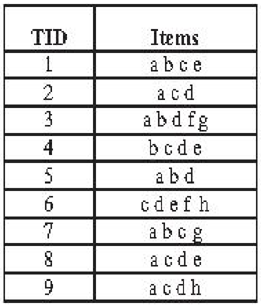
In the above table 2 is represent the present items and 0 is represent the absent items.

Table 3: Transaction with sensitivity shows in that table.



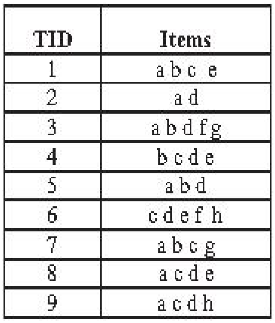
TID	Sensitivity
1	9
2	8
3	7
4	6
5	7
6	5
7	6
8	8
9	8

Table 4: Sanitized Database



TID	Items
1	abce
2	acd
3	abdfg
4	bcde
5	abd
6	cdefh
7	abcg
8	acde
9	acdh

Table 5: Final Sanitized Database



TID	Items
1	abc e
2	ad
3	abdfg
4	bcde
5	abd
6	cdefh
7	abcg
8	acde
9	acdh

6. CONCLUSION

The use of the association rule hiding techniques for privacy preserving data mining is to hide certain private Information so they cannot discovered through association rule. in this paper, we proposed an algorithm named MDSRRC which hides sensitive association rules with fever modifications on database to maintain data quality and to reduce the side effect of database. MDSRRC algorithm can be extended to increase the efficiency and reduce the side effects by decreasing the modifications on database.

REFERENCES

[1] M. Atallah, A. Elmagarmid, M. Ibrahim, E. Bertino, and V. Verykios, “Disclosure limitation of sensitive rules,” in Proceedings of the 1999 Workshop on Knowledge and Data Engineering Exchange, ser. KDEX ’99. Washington, DC, USA: IEEE Computer Society, 1999, pp. 45–52.

[2] C. N. Modi, U. P. Rao, and D. R. Patel, “Maintaining privacy and data quality in privacy preserving association rule mining,” 2010 Second Iternational conference on Computing, Communication and Networking Technologies, pp. 1–6, Jul. 2010.

[3] S. Wu and H. Wang, “Research on the privacy preserving algorithm of association rule mining in centralized database,” in Proceedings of the the 2008 International Symposiums on Information Processing, ser. ISIP’08. Washington, DC, USA: IEEE Computer Society, 2008, pp. 131–134.