International Journal of Engineering Research and General Science Volume 2, Issue 6, October-November, 2014
ISSN 2091-2730

DATA PRIVACY USING MDSRRC

Priyanka k. Dhongade[1] ,Prof. Yogesh Nagargoje[2]
CSE Department
Everest Educational Scociety's Group of Iinstitutions,
Dr. Seema Quadri Institute of Tech, Dr.B.A.M.U, ,Aurangabad, India

priyankadhongade18@gmail.com, yogeshvcet1@gmail.com

Abstract—*:* The use of the data mining techniques and its related application is increased in recent years to extract important knowledge from large amount of data. This has increased the disclosure risks to sensitive information when the data is released to outside parties. Database containing sensitive knowledge must be protected against unauthorized access. Seeing this it has become necessary to hide sensitive knowledge in database. In this paper, we propose a heuristic based algorithm named MDSRRC (Modified Decrease Support of R.H.S. item of Rule Clusters) to hide the sensitive association rules with multiple items in consequent (R.H.S) and antecedent (L.H.S). This algorithm overcomes the limitation of existing rule hiding algorithm DSRRC. Proposed algorithm selects the items and transactions based on certain criteria which modify transactions to hide the sensitive information. Experimental result shows that proposed algorithm is highly efficient and maintains database quality.

KEYWORDS— ASSOCIATION RULE, SENSITIVE PATTERN, PRIVACY PRESERVING DATA MINING (PPDM), SENSITIVITY

INTRODUCTION

Association rule mining technique is widely used in data mining to find relationship between item sets. Many organizations disclose their information or database for mutual benefit to find some useful information for some decision making purpose and improve their business schemes. But this database may contain some private data and which the organization does not want to disclose. The issue of privacy plays important role when several organizations share their data for mutual benefit but no one wants to disclose their private data. Therefore before disclosing the database, sensitive patterns must be hidden and to solve this issue PPDM techniques are helpful to enhance the security of database.

Mining association rule techniques are wide employed in data mining to and relationship between item sets. The corporate and many government organizations reveals their data or information for mutual benefit to search out some useful data for some decision making purpose and improve their business schemes. But this database may contain some confidential information and which the organization does not need to reveal.

The problem of concealment plays important role once the corporate share their data for mutual profit however there is no one need to leak their rivet data. So before revealing the information, sensitive patterns should be hidden and to resolve this issue PPDM (Privacy preserving data mining) techniques are helpful to boost the safety of database. These approaches have in general the advantage to require a minimum amount of input (usually the database, the information to protect and few other parameters) and then a low effort is required to the user in order to apply them. The selection of rules would require data mining process to be executed first. For association rules hiding, two basic approaches have been proposed. The first approach hides one rule at a time. First selects transactions that contain the items in a give rule. It then tries to modify transaction by transaction until the confidence or support of the rule fall below minimum confidence or minimum support. The modification is done by either removing items from the transaction or inserting new items to the transactions. The second approach deals with groups of restricted patterns or association rules at a time. In our work we are concern of hiding certain association rules which contain some sensitive information which are on the Right hand side or left hand side of the rule, so that rules containing confidential item can't be reveal. Our approached is based on modifying the database in a way that confidence of the association rule can be reduce with the help increase or decrease the support value of RHS or LHS correspondingly. As the confidence of the rule is reduce below a specified threshold, it is hidden or we can say it will not be disclosed. The proposed formula is that the improved version of DSRRC. DSRRC could not hide association rules with multiple items in antecedent (L.H.S) and resultant (R.H.S.). To overcome this limitation, we proposed an algorithmic rule MDSRRC which uses count of things in resultant of the sensitive rules. It modifies the minimum number of transactions to cover most sensitive rules and maintain data quality. [1][2][3]

II. LITERATURE REVIEW AND THEORETICAL BACKGROUND

Association rule hiding techniques can be classified into heuristic based approaches, reconstruction based approaches, border based approaches, exact approaches, and cryptography based approaches. Proposed algorithm use heuristic based approach which is widely used.

A. *Heuristic Approaches for Hiding The Sensitive Rule* These approaches use mainly two techniques for hiding sensitive rule: data distortion which permanently deletes some items from database and data blocking which put '?' instead of deleting items from database.

Data Distortion changes the item value by a new value in database matrix. It alter '0' to '1' or '1' to '0' for selected items in selected transactions to decrease the confidence, by decreasing or increasing support of items in sensitive rules. Heuristic algorithms cannot give an optimal solution because of side effects to non sensitive rules. [4] Presented heuristic algorithm for hiding sensitive rules. They have also provided proof of NP-Hardness of optimal solution of sanitization problem. [2] Proposed five different algorithms with five assumptions to hide sensitive rules in database, among them three are based on reduced support of item set and two are based on reduced confidence of the rule below the minimum threshold. [5] Considers all side effect parameters and based on that modify the selected transactions to reduce the side effects on sanitized database. [6] Proposed two algorithms to automatically hide sensitive association rules without pre mining and selection of hidden rules [7]. [3] Proposed algorithm using clustering to reduce the side effects on sanitized database but it can hide rules only with single antecedent and single consequent.

Data Blocking instead of inserting or deleting item from database it replaces '1' and '0' with '?' in selected transactions.

So after applying this technique, adversary will not know the original value of '?'. [8] And [9] proposed algorithm which uses data blocking technique to hide sensitive rules. [10] Proposed more efficient algorithm using clustering than presented in [8] [9]. We have proposed heuristic based approach which is efficient than other approaches presented in above section.

III. PROPOSED MODIFIED DECREASE SUPPORT OF R.H.S ITEM OF RULE CLUSTER ALGORITHM

In order to hide the sensitive rule like $A \rightarrow B$, we can decrease either confidence or support of the rule below the user specified minimum threshold. To decrease the confidence of the rule, we can choose two methods like (1) increase the support of A (L.H.S. of the sensitive rule) but not support of $A \cup B$, or (2) decrease the support of $A \cup B$ by decreasing support of B (R.H.S of the sensitive rule) because it decrease the confidence of the rule faster than simply decreasing the support of $A \cup B$. Proposed algorithm hides rules with multiple items in L.H.S and multiple items in R.H.S. So the rule is like $xA \rightarrow yB$ where $x, y \in I$ and $A, B \subset I$. Here y is an item selected by proposed algorithm to decrease the support of the R.H.S. and decrease the confidence of the rule below MCT. We replace '1' to '0' in some transaction to decrease the support of selected items.

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

1. Sensitivity of Item: is number of sensitive rules which contain this item.
2. Sensitivity of Transaction: is the total of sensitivities of all sensitive items which are presented in that transaction.

A detail description of sensitivity is given in detail in [12]. The proposed algorithm starts with mining the association rule from the original database D using association rule mining algorithm e.g. Apriori algorithm [11]. Then some rules as sensitive rules (SR) are specified by user from the rules generated by the association rule mining algorithm. Then algorithm counts occurrences of each item in R.H.S of sensitive rules. Now algorithm finds IS= {is0, is1… isk } $k \leq n$, by arranging 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 sorted in descending order of their sensitivities.

Now rule hiding process starts by selecting first transaction from the sorted transactions 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 below MST and MCT 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. As a result, modified transactions are updated in the original database and new database is generated which is called sanitized database D', which preserves the privacy of sensitive information and maintains database quality. Proposed algorithm MDSRRC is shown below, which is used to hide the sensitive rules from database. Given a database D, MCT (minimum confidence threshold) and MST (minimum support threshold) algorithm generates sanitized database D'. Sanitized database hides all sensitive rules and maintains data quality.

MDSRRC Algorithm

INPUT:

MCT (Minimum Confidence Threshold), Original database D,

MST (Minimum support threshold).

OUTPUT:

Database D' with all sensitive rules are hidden.

1. Apply apriori algorithm [3] on given database D. Generate all possible association rules R.
2. Select set of rules SR⊂R as sensitive rules
3. Calculate sensitivity of each item j∈D.
4. Calculate sensitivity of each Transaction.
5. Count occurrences of each item in R.H.S of sensitive rules, find IS={is0, is1...isk} k≤n, by arranging those items in descending order of their count. If two items have same count then sort those in descending order of their actual support count
6. Select the transactions which supports 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.
7. While(SR is not empty)
8. {
9. Start with first transaction from sorted transactions,
10. Delete item is0 from that transaction.
11. For each rule r∈SR
12. {
13. Update support and confidence of the rule r.
14. If(support of r < MST or confidence of r < MCT)
15. {
16. Delete Rule r from SR.
17. Update sensitivity of each item.
18. Update IS (This may change is0).
19. Update the sensitivity of each transaction.
20. Select the transactions which are supports is0,
21. Sort those in descending order of their sensitivity.
22. }
23. Else
24. {
25. Take next transaction from sorted transactions, go to step 10.
26. }
27. }
28. }
29. End

MDSRRC select best items so that deleting those items hide maximum rules from database to maintain data quality.

IV. CONCLUSION

The proposed MDSRRC algorithm overcomes the limitations of DSRRC and provides Association rule hiding techniques for privacy preserving data mining to hide certain crucial information so they cannot discover through association rule MDSRRC hides sensitive association rules with fewer modifications on database to maintain data quality and to reduce the side effect of database. In future, MDSRRC algorithm can be extended to increase the efficiency and reduce the side effects by minimizing the modifications on database.

References

[1] Nikunj H. Domadiya and Udai Pratap Rao, "Hiding Sensitive Association Rules to Maintain Privacy and Data Quality in Database"*3rd IEEE International Advance Computing Conference (IACC) 2013.*

[2] V. S. Verykios, A. K. Elmagarmid, E. Bertino, Y. Saygin, and E. Dasseni, "Association rule hiding," *IEEE Transactions on Knowledge and Data Engineering, vol. 16, pp. 434–447, 2004.*

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

[4] 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.

[5] Y.-H. Wu, C.-M. Chiang, and A. L. Chen, "Hiding sensitive association rules with limited side effects," IEEE Transactions on Knowledge and Data Engineering, vol. 19, pp. 29–42, 2007.

[6] S.-L. Wang, B. Parikh, and A. Jafari, "Hiding informative association rule sets," Expert Systems with Applications, vol. 33, no. 2, pp. 316 – 323, 2007.

[7] S.-L. Wang, D. Patel, A. Jafari, and T.-P. Hong, "Hiding collaborative recommendation association rules," Applied Intelligence, vol. 27, pp. 67–77, 2007.

[8] Y. Saygin, V. S. Verykios, and A. K. Elmagarmid, "Privacy preserving association rule mining." in RIDE. IEEE Computer Society, 2002, pp 151–158.

[9] Y. Saygin, V. S. Verykios, and C. Clifton, "Using unknowns to prevent discovery of association rules," SIGMOD Rec., vol. 30, no. 4, pp. 45 - 54, Dec. 2001.

[10] C. N. Modi, U. P. Rao, and D. R. Patel, "An Efficient Solution for Privacy Preserving Association Rule Mining," (IJCNS) International Journal of Computer and Network Security, vol. 2, no. 5, pp. 79–85, 2010.

[11] J. Han, Data Mining: Concepts and Techniques. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2005.

[12] S. Wu and H. Wang, "Research on the privacy preserving algorithm of association rule mining in centralized database," in Proceedings of the 2008 International Symposiums on Information Processing, ser. ISIP'08. Washington, DC, USA: IEEE Computer Society, 2008, pp