An Enhanced Approach of Sensitive Information Hiding

Kaushal Bhatt[1] Ankit Dongre[2]
[1]P.G. Scholar [2]Assistant Professor
[1,2]JIT, Borawan, Khargone (M.P)

Abstract— Organizations involving in similar businesses often share their databases to find out useful business logics. The process of finding is called data mining. But serious concern occur when the database, shared by organization contain some sensitive information and organization what to hide this information before sharing the database. Then the concept of privacy preserving data mining is comes in spot light where the sensitive information is hide in a such a way that database does not lost its integrity and business logics derived from this databases does not contain sensitive information. In this research we introduce new technique for privacy preserving data mining, which hides sensitive information effectively as compared to previous research work.

Key words: PPDM, Hybrid algorithm, Sensitive information, Association rules

I. INTRODUCTION

Association rules are in the form of A -> B where A,B are subset of I are the sets of items called Item sets and A ∩B = Φ. Association rules show attributes value conditions that appear frequently together in a transaction data base. A mostly example are used of association rule data mining is Market Basket Analysis [2]. The set of items is- **I = {Milk, Bread, Butter}** A rule derived from the shopping market database could be **{Butter, Bread} => {Milk}** meaning that if butter and bread are bought, customers also buy milk. **Association rules** [4, 2] provide information on the basis of “if then” statements. These rules are computed from the data and, unlike the “if then” rules of logic, the association rules are probabilistic. If 90% of transactions that purchase bread and butter, then also purchase milk.

- **Antecedent**: bread and butter
- **Consequent**: milk
- **Confidence factor:** 90%

In addition to the antecedent (the “if” part) and the consequent (the “then” part), an association rule has two numbers that express the degree of uncertainty about the rule. Associations rule analysis the collection of antecedent and consequent are sets of items (called item sets) that are also known as disjoint. It means that they do not have any item in Common. **Support** for an association rule X->Y is the percentage of transaction in database

That contains X U Y. The second big parameter is called the **Confidence** of the rule. Strength for an association rule X U Y is the ratio of number of transactions that contains X U Y to number of transaction that contains X.

Support A=>B Common item in any giving table / Total no transaction in any table

Confidence A=>B Total Support in Number (A U B) / Total support in Number (A)

Association Rule Hiding

The problem of association rule hiding was first probed in 1999.After that, many Approaches were proposed. They are categorized as - data sanitization data modification approaches and knowledge sanitization data reconstruction approaches. The data modification approaches [7,10] are also the so-called data sanitization. They generally hide sensitive association rules by directly modifying sanitizing the original data D, to the database D’ directly from D. As the sanitization is performed on data level, data modification approaches cannot control the hiding effects intuitively. Moreover, it is found that the data sanitization can produce a lot of I/O operations [6,9].

II. RELATED WORK

Many researchers proposed research on association rule hiding, due to vast concern about privacy of data. In 2008 belwal [1] introduce modified definition of support and confidence in their research by introducing hiding counter which is used to reduce the confidence and support of association rule below minimum threshold. The ISL (increase support at left hand side) algorithm [] and DSR (Decrease Support at Right hand side) is achieved through modify the database table from 1 to 0 or from 0 to 1 in a selected transaction. The hybrid approach of association rule hiding suggested by research [ ] in order to hide an association rule, either decrease its support or its Confidence to be below than pre-specified minimum support and minimum confidence threshold . this method utilize the both ISL and DSR approaches of privacy preserving data mining This algorithm first tries to hide the rules in which item to be hidden i.e. A is in right hand side and then tries to hide the rules in which A is in left hand side. For this algorithm t is a transaction, T is a set of transactions, AR is used for rule, RHS(AR) is Right Hand Side of rule AR, LHS(AR) is the right hand side of the rule AR, Confidence(C) is the confidence of the rule AR.

III. PROPOSED WORK

A. *Algorithm:-*

1) Input:
   - A source database D_o,
   - A minimum confidence threshold value min_confidence,
   - A set of hidden items X_h.

B. *Procedure:*

1) *Find all possible rules from given items X_h;*
2) *Compute confidence of all the rules.*
3) *For each hidden item H*
4) *For each rule AR in which H is in RHS*
   - If confidence (C) < min_confidence, then Go to next large 2-itemset;
   - Else go to step 5
5) *Decrease Confidence of RHS i.e. item h.*
   - Find T = t in D | t ;
   - While (T is not empty)

```
      –   Choose the first transaction t from T;
      –   Modify t by putting 0 instead of 1 for RHS item;
      –   Remove and save the first transaction t from T;
End While
6) Compute confidence of AR;
7) If T is empty, then H cannot be hidden;
      –   End For
      –   End
```

C. Output:

Updated D_O, as the transformed D_M

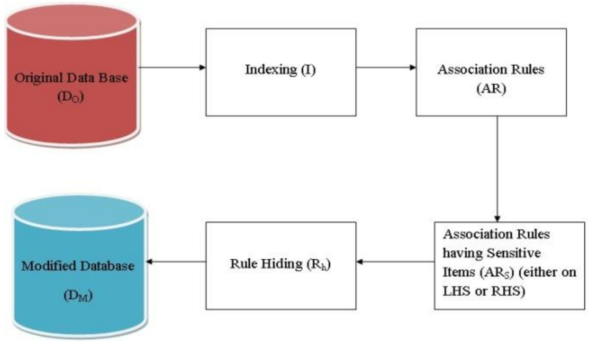
Original Data Base (D_O)
Indexing (I)
Association Rules (AR)
Association Rules having Sensitive Items (AR_S) (either on LHS or RHS)
Rule Hiding (R_h)
Modified Database (D_M)

Diagram [3.1] Working Model of Algorithm

IV. IMPLEMENTATION

To hide any sensitive item X algorithm works on the basis of confidence (X → Y) or support (X → Y). To hide any sensitive item in rule X → Y, this algorithm first finds the value of confidence (conf) in the available set of rules if the confidence value is above the user specific minimum confidence level , then algorithm modify the database in such a way that, after calculating the confidence from modified database rules containing sensitive item either on left or right side always have confidence value below than user specific confidence value.

Dataset: Suppose our transactional database is as follow-

Tid	Milk	Butter	Bread
T1	1	1	1
T2	1	1	1
T3	1	1	1
T4	1	1	0
T5	1	0	0
T6	1	0	1

Table [4.1] Transaction Table

L	R	C
Milk	Butter	67
Butter	Bread	75
Bread	Butter	75
Milk	Bread	100
Bread	Milk	67
Butter	Milk	100

Table [4.2] Calculated confidence table

A. Suppose Minimum threshold value of confidence is 60%

Suppose we first want to hide item Milk, for this, first take rules in which Milk is in RHS. These rules are Butter–>Milk and Bread–>Milk and both have greater confidence. Choose Butter–>Milk and search all transactions where Milk and Butter is equal to 1.There are four transactions T1, T2, T3, T4 with Milk = Butter = 1. Put 0 for item Milk in all the four transactions. After this modification, we get Table 1 as the modified table.

Tid	Milk	Butter	Bread
T1	0	1	1
T2	0	1	1
T3	0	1	1
T4	0	1	0
T5	1	0	0
T6	1	0	1

Table [4.3] Modified table1

Now calculate confidence of Butter–>Milk, it is 0% which is less than minimum confidence so now this rule is hidden. Now take rule Bread–>Milk, search for transactions in table 1 where Milk and Bread is equal to 1, only transaction T6 has Milk= Bread = 1, update transaction by putting 0 instead of 1 in place of Milk. Now calculate confidence of Bread–>Milk, it is 0% which is less than the minimum confidence so now this rule is hidden. After this modification, we again check the table 1 and it clearly shows that in T6 Milk value is changed to 0.

Tid	Milk	Butter	Bread
T1	0	1	1
T2	0	1	1
T3	0	1	1
T4	0	1	0
T5	1	0	0
T6	0	0	1

Table [4.4] Modified table 1

After that we again calculated confidence from modified transaction table.

L	R	C
Milk	Butter	0
Butter	Bread	75
Bread	Butter	75
Milk	Bread	0
Bread	Milk	0
Butter	Milk	0

Table [4.5] Modified confidence table derived from table 4

We clearly see that association rules having sensitive information either on left hand side or right hand side having confidence below minimum confidence value which is 60%.

Result set:- Our proposed technique scan database less as compared to hybrid algorithm that scan data item first for right hand side and modify table and again left hand side for the same. Which we consider unnecessary scan because after first modification the database having value of sensitive information is 0 Below we show graphical comparison between proposed algorithm and hybrid algorithm.

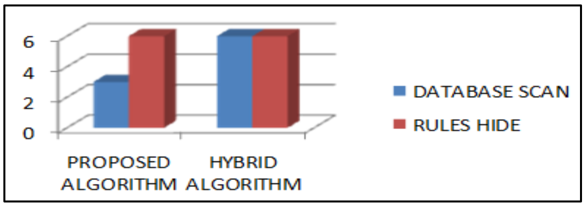

Chart [4.1] Comparison in terms of database scan and rule hidden

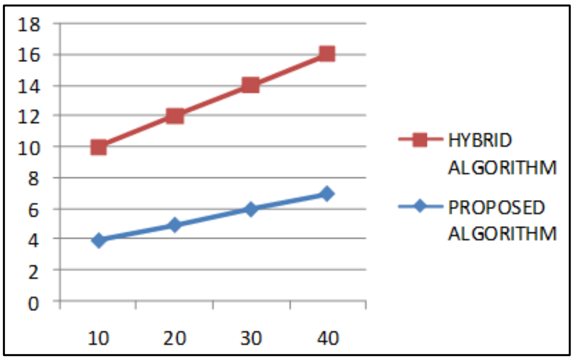

Chart [4.2] Number of Rules v/s Time

V. CONCLUSION

In this research, we have proposed an algorithm for hiding sensitive data in association rules mining which based on modifying the database transactions so that the confidence of the association rules can be reduced. The efficiency of the proposed approach is further compared with Hybrid approach and shown that this approach prunes more number of hidden rules with less number of times database scanned.

REFERENCES

[1] Ila Chandrakar, Manasa, Usha Rani, and Renuka. Hybrid Algorithm for Association Rule mining. Journal of Computer Science 6(12), pages 1494-1498, 2010.

[2] A. K. Pujari. Data Mining Techniques (book), 2001. University Press (India) limited.

[3] Poovammal, E. and M. Ponnavaikko, 2009. Utility independent privacy preserving data mining on vertically partitioned data. J. Comput. Sci., 9: 666-673. DOI: 10.3844/jcssp.2009.666.673

[4] Agrawal, R. and R. Srikant, 1998. Fast Algorithms for Mining Association Rules. In: Readings in Database Systems, Stonebraker, M. and J. Hellerstein (Eds.). Morgan Kaufmann, Massachusetts, ISBN: 1558605231, pp: 580-592

[5] Belwal, Varsheney, Khan, Sharma, Bhattacharya. Hiding sensitive association rules efficiently by introducing new variable hiding counter .IEEE 2008.

[6] Agrawal, R., and Srikant (2007), Privacy Preserving Data Mining‟, Proceedings of the 19th ACM International Conference on Knowledge Discovery and Data Mining, Canada, pp. 439-450.

[7] Bhatt, Kaushal, and Ankit Dongre. "A Survey of Sensitive Information Hiding Techniques." IJETAE-volume-4 , issue -1.

[8] R. Agrawal, T. Imielinski, and A. Swami. Mining association rules between sets of items in large databases. In Proceedings of the ACM SIGMOD Conference on Management of Data,pages 207–216, New York, NY, USA, May 1993. ACM Press.

[9] Aris Gkoulalas–Divanis;Vassilios S. Verykios ―Association Rule Hiding For Data Mining‖ Springer, DOI 10.1007/978-1-4419-6569-1, Springer Science + Business Media, LLC 2010

[10] E. Dasseni, V. S. Verykios, A. K. Elmagarmid, and E. Bertino. Hiding association rules by using confidence and support. In I. S. Moskowitz, editor, Proceedings of the 4th Information Hiding Workshop, volume 2137, pages 369–383, 2001. Springer Veralg Lecture Notes.