K.Srinivasa Rao *et al*, International Journal of Computer Science and Mobile Applications,

Vol.2 Issue. 4, April- 2014, pg. 8-13 **ISSN: 2321-8363**

Support and Confidence Based Algorithms for Hiding Sensitive Association Rules

K.Srinivasa Rao

Associate Professor, Department Of Computer Science and Engineering,
Swarna Bharathi College of Engineering, Khammam
ksrao517@gmail.com

Dr. CH. Suresh Babu

Professor Dept.of CSE, Sree Vaanmayi Inst of Science & Technology, Nalgonda
sureshbabuchangalasetty@gmail.com

Dr. A. Damodaram

Director – Academic Audit Cell & Prof. of CSE Dept, JNTUH, Hyderabad
damodarama@rediffmail.com

ABSTRACT

We propose an algorithm to hiding association rules on data mining. We make a target data table without joining the multiple tables using the hiding association rules. Hiding association rules joined data table and all dimension tables, it reduces support and confidence in multi relational data mining. Association analysis is a powerful and popular tool for discovering relationships hidden in large data sets. We can modify transaction data item sets.

KEYWORDS

Data Mining; Association rules hiding; Minimum confidence; Minimum support

1. INTRODUCTION

Data mining is the process of extracting useful information or knowledge from large databases. Data mining has developed an important technology for large Database. Data mining applications like business, marketing, medical analysis, products control and scientific etc [1], [2]. Association rule mining is one of the important problems in the data mining domain. Association rules analysis is a popular tool for discovering useful association from large database. Some difficult and sensitive hidden information has a very large critical problem to resolved [3]. The association hiding rules divided into two parts sensitive association rules, sensitive association items.

Association rule hiding proposed the two approaches. The first approach hides one rule at a time. The first approach select the transaction items in database and then modification transaction items that means inserting items and removing items in database. The second approach hides group of rules at a time. In this approaches also allow useful access to only a subset of data item [2]. If we apply association rules hiding in data mining, it is reduces the support association rules and modifying the relationship in the database.

However, a database is typically contains of multiple tables. For example, there are multiples dimension tables and a fact table in a database, Although efficient mining techniques have been proposed to discover frequent item sets and multi-relational association rules from multiple tables [8,9].

In this paper, we purposed a novel algorithm for hiding association rules in multi-relational data mining Based on the association rules of multiple table reduction of confidence in database. The rest of the paper is organized as follows. Section 2 presents the problem description. Section 3 proposed algorithm for hiding sensitive association rules from multiple tables. Section 4 shows an example of the proposed Conclusion. Section 5 shows the result of the proposed algorithm and compares it with joining table approach.

©2014, IJCSMA All Rights Reserved, www.ijcsma.com

2. PROBLEM DESCRIPTION

The main idea of Multi-relational data mining is to extract hidden rules and useful unused data from large database. We proposed an algorithm for sensitive association rules. This algorithm used to modify transaction data and inserting new data in database and removing data from database. More specify that given a transaction database D, a minimum support, a minimum confidence and a set of items S to be hidden [7,6,5]In this paper, we assume that only sensitive are given and purpose an algorithm to modify data in database. So that sensitive items cannot be informed through association rules mining algorithm. Based on two strategies, we propose two data mining algorithm for hiding sensitive items in association rules namely Increase support of LHS first (ISLF) and Decrease support of RHS first (DSRF) [6]. Note while the support measure the frequency of association rules and the confidence is a measure of the strength of the relationship between set of items in mining sensitive association rules that are greater than the minimum support threshold and minimum confidence threshold. In this transaction database D, MST and MCT, In the database the sensitive association rule R and mined rules Rn. Then we apply the MST and MCT. In MST the sensitive rules Rn ⊆ R to be find a new database. In MCT R-Rn is hidden in rules and database [3]. With this assumption hiding them one data time or all together will not make any difference. Hiding a sensitive rule will not affect any other sensitive rule [1].

3. PROBLEM FORMULATION

Association rules using support and confidence can be defined as follows. Let I={I1,I2,…….Im} be a set of items. Let D={T1,T2,……...Tn} be a set of transactions. where each transaction T in D is a set of items such that $T \subseteq I$ an association rules of implication in the form of $X \rightarrow Y$, where
$X \subset I$, $Y \subset I$ and $X \cap Y = \emptyset$. We can say the rule $X \Rightarrow Y$ holds in database D with confidence C if $|X \cup Y| / |X| \geq C$,
We also say that the rule $X \Rightarrow Y$ has support S if $|X \cup Y| / |D| \geq S$[3,2,6,9].

In X and Y represent the body (left hand side) and head (right hand side) according the rule such as 95% customers buy the toothpaste and also buy the toothbrush. Hera the confidence rule is 95%. It means that the 95% of transaction that contains the X (toothpaste) and also Y (toothbrush) [6]. In other words, the confidence of a rule measures the degree of the correlation between item sets, while the support of a rule measures the significance of the correlation between item sets. The problem of mining association rules is to find all rules that are greater than the user-specified minimum support and minimum confidence [2,10,11].

As an example, for a given database in Table 1, a minimum support of 33% and a minimum confidence of 70%, nine association rules can be found as follows: B ⇒A (66%,100%), C⇒A (66%,100%), B ⇒C (50%,75%), C⇒ B (50%,75%), AB⇒ C (50%,75%), AC ⇒B (50%,75%), BC⇒A (50%, 100%), C⇒AB (50%, 75%), B ⇒ AC (50%,75%), where the percentages inside the parentheses are supports and confidences respectively[6,8,12,13].

Table 1: Large item sets obtained from D

Item Sets	Support
A	100%
B	66%
C	66%
AB	66%
AC	66%
BC	50%
ABC	50%

K.Srinivasa Rao *et al*, International Journal of Computer Science and Mobile Applications,
Vol.2 Issue. 4, April- 2014, pg. 8-13 ISSN: 2321-8363

Table 2: The rules derived from the large item set of Table 3

Rules	Confidence	Support
$B \Rightarrow A$	100%	66%
$B \Rightarrow C$	75%	50%
$C \Rightarrow A$	100%	66%
$C \Rightarrow B$	75%	50%
$B \Rightarrow AC$	75%	50%
$C \Rightarrow AB$	75%	50%
$AB \Rightarrow C$	75%	50%
$AC \Rightarrow B$	75%	50%
$BC \Rightarrow A$	100%	50%

Before presenting the solution strategies, we introduce some notation. Each database transaction is a triple: t=<TID, list_ of _elements, size>.

Where TID is the identifier of the transaction t and list_0f_elements is a list with one element for each item in the database. Each element has value 1. if the corresponding item is supported by the transaction and 0 otherwise. Size is the number of elements in the list of elements having value 1 (e.g., the number of elements supported by the transaction). For example, if 1= {A,B,C,D} a transaction that contains the items {A,C} would be represented as t=< TI,[1010],2>. According to this notation, a transaction t supports an item set S if the elements of t list_ of elements corresponding to items of S are all set to 1. A transaction t partially supports S if the elements of t list_ of _ elements corresponding to items of S are not all set to 1. For example, if S = {A,B,C} = [1110] and p=< T1, [1010],2>, q=< T2, [1110], 3 > then we would say that q supports S while p partially supports S [12,13].

Table 3: The sample database that uses the proposed notation

TID	Items	Size
T1	111	3
T2	111	3
T3	111	3
T4	110	2
T5	100	1
T6	101	2

Definition 1.1: Association rule mining

1. Sup $_X\cup_Y(=C_X\cup_Y / |D|) \geq MST$
2. Conf $_{X\to Y}$ $(=CX\cup Y/C_X) \geq MCT$

Definition 1.2: A class of modification. Given two transaction sets ∑1 and ∑2, a class of modification is a function φ: (∑1, I, O) →∑2 that transforms ∑1 to∑2, where I is the item(s) to be modified and O is the modification scheme.

Definition 1.3: Association rule hiding. Let D' be the database after applying a sequence of modification to D. A strong rule X→Y in D will be hidden in D' if one of the following condition holds in D'.

1. Sup XUY < MST
2. Conf X→Y < MCT

4. PROPOSED ALGORITHM

We propose two data mining algorithms for hiding sensitive association rules, namely Increase Support of LHS (ISLF) and Decrease Support of RHS (DSRF). The first algorithm tries to increase the support of left hand side of the rule. The second algorithm tries to decrease the support of the right hand side of the rule. The details of the two algorithms are described as follow.

Algorithm (ISLF)

Input:

1. A source database D,
2. A min_ support,
3. A min_ confidence,
4. A set of predicting items X

Output: a transformed database D', where rules containing X on LHS will be hidden

1. Find large 1-item sets from D;
2. for each predicting item x X
3. If x is not a large 1-itemset, then X: = X _ {x};
4. If X is empty, then EXIT;
5. Find large 2-itemsets from D;
6. For each x X {
7. For each large 2-itemset containing x {
8. Compute confidence of rule U, where U is a rule like x → y;
9. If conf (U) < min_ conf, then
10. Go to next large 2-itemset;
11. Else {//Increase Support of LHS
12. Find T_L = {t in D/t does not support U};
13. Sort T_L in ascending order by the number of items;
14. While {conf (U) ≥ min_ conf and T_L is not empty {
15. Choose the first transaction t from T_L;
16. Modify t to support x, the LHS (U);
17. Compute support and confidence of U;
18. Remove and save the first transaction t from T_L; 19.}; // end While
20.}; // end if conf (U) < min_ conf
21. If T_L is empty, then {
22. Cannot hide x →y;
23. Restore D;
24. Go to next large-2 item set;
25. } // end if T_L is empty
26. } // end of for each large 2-itemset
27. Remove x from X;
28. } // end of for each x X
29. Output updated D, as the transformed D':

Algorithm (DSRF)
Input:
1. A source database D,
2. A min_ support,
3. A min_ confidence,
4. A set of predicting items X

Output: a transformed database D', where rules containing X on LHS will be hidden

1. Find large 1-item sets from D;
2. for each predicting item x X
3. If x is not a large 1-itemset, then X :=X _ {x};
4. If X is empty, then EXIT;
5. Find large 2-itemsets from D;
6. For each x X {
7. For each large 2-itemset containing x {
8. Compute confidence of rule U, where U is a rule like x → y;
9. If conf (U) < min_ conf, then
10. Go to next large 2-itemset;
11. Else {//Decrease Support of RHS
12. Find T_R = {t in D/t fully support U};
13. Sort T_R in ascending order by the number of items;
14. While {conf (U) P min_ conf and T_R is not empty} {
15. Choose the first transaction t from T_R;
16. Modify t so that y is not supported;
17. Compute support and confidence of U;

©2014, IJCSMA All Rights Reserved, www.ijcsma.com

18. Remove and save the first transaction t from T_R;
19. }; // end While
20. }; // end if conf(U) < min_ conf
21. If T_R is empty, then {
22. Cannot hide x → y;
23. Restore D;
24. Go to next large-2 item set;
25. } // end if T_R is empty
26. } // end of for each large 2-itemset
27. Remove x from X;
28. } // end of for each x X
29. Output updated D, as the transformed D';

Table 4: Database D using specified notation

TID	Items	ABC	Size
T1	ABC	111	3
T2	ABC	111	3
T3	ABC	111	3
T4	AB	110	2
T5	A	100	1
T6	AC	101	2

The all possible rules with confidence are: A→B (66.6%), A→C (66.6%), B→A (100%), B→C (75%), C→A (100%), C→B (75%). Suppose we first want to hide item A, first take rule in which A is in RHS. These rules are B→A and C→A both has greater confidence from MCT. First take rule B→A search for transaction which support both B and A, B=A=1. There are four transactions T1, T2, T3, T4 with A=B=1. Now update table put 0 for item A in all four transactions. Now calculate confidence of B→A, it is 0% which is less than MCT so now this rule is hidden. Now take rule C→A, search for transaction in which A=C=1, only transaction T6 has A=C=1, update transaction by putting 0 instead 1 in place of A. Now take the rules in which A is in LHS. There are two rules A→B and A→C but both rules have confidence less than MCT so there is no need to hide these rules. So Table 5 shows the modified database after hiding item A[12].

Table 5: Update table after hiding item A

TID	ABC	Size
T1	011	2
T2	011	2
T3	011	2
T4	010	1
T5	100	1
T6	101	2

6. CONCLUSION

The purpose of the Association rule hiding algorithm for privacy preserving data mining is to hide certain crucial information so they cannot discovered through association rule. In this paper, we have proposed an efficient Association rule hiding algorithm for privacy preserving data mining. This is based on association rule hiding approach of previous algorithms and modifying the database transactions so that the confidence of the association rule can be reduced. In our proposed algorithm we can hide the generated crucial association rule on the both side (LHS and RHS) correspondingly, so it reduce the number of modification, hide more rule in less time. The efficiency of the proposed algorithm is compared with ISLF and DSRF approach. Our algorithm prunes more number of hidden rules with same number of transactions and modification.

7. EXAMPLE

This section shows an example of the proposed algorithm in hiding sensitive item in association rule mining. Consider Table 4 as a database, MST=33%, MCT=70%, each element has value 1 if the corresponding item is supported by the transaction and 0 otherwise. Size means the number of elements in the list having value 1.

Vol.2 Issue. 4, April- 2014, pg. 8-13 ISSN: 2321-8363

REFERENCES

[1] Yi-Hung Wu, Chia-Ming Chiang, and Arbee L.P. Chen "HidingSensitive Association Rules with Limited Side Effects" IEEE Transaction on Knowledge and data engineering, Vol.19, pp. 29-42, 2007.

[2] Shyue-Liang Wang, Bhavesh Parikh, Ayat Jafari "Hiding informative association rule sets" Science direct. 2006.

[3] Manoj Gupta and R. C. Joshi" Privacy Preserving Fuzzy Association Rules Hiding in Quantitative Data" International Journal of Computer Theory and Engineering, Vol. 1, No. 4, October, 2009

[4] Vassilios S. Verykios, Ahmed K. Elmagarmid,Elisa Bertino, Yucel Saygin,Elena Dasseni" Association Rule Hiding" IEEE Transaction on Knowledge and data engineering, Vol.16, April 2004.

[5] Guanling Lee, Chien-Yu Chang, Arbee L.P Chen" Hiding Sensitive Patterns in Association Rules Mining" Proceedings of the 28th Annual International Computer Software and Applications Conference (COMPSAC'04), 2004

[6] Shyue-Liang Wang, Yu-Huei Lee, Steven Billis, Ayat Jafari" Hiding Sensitive Items in Privacy Preserving Association Rule Mining" IEEE International Conference on Systems, Man and Cybernetics, 2004

[7] Xiaoming Zhang, Xi Qiao" New Approach for Sensitive Association Rule Hiding" 2008 International Workshop on Education Technology and Training, IEEE computer society, 2008

[8] Shan-Tai Chen, Shih-Min Lin, Chi-Yii Tang, and Guei- Yu Lin" An Improved Algorithm for Completely Hiding Sensitive Association Rule Sets" 2nd International conference on computer science and its application,pp.1-6,2010

[9] Shyue-Liang Wang1and Tzung-Pei Hong2 Yu-Chuan Tsai, Hung-Yu Kao" Multi-table Association Rules Hiding" 2nd International conference on Intelligent system design and application,pp.1298-1302,2010 .

[10] Chih-Chia Weng, Shan-Tai Chen, Hung-Che Lo" A Novel Algorithm for Completely Hiding Sensitive Association Rules" Eighth International Conference on Intelligent Systems Design and Applications,vol. 2,pp.202-208,2008.

[11] Yogendra Kumar Jain , Vinod Kumar Yadav, Geetika S. Panday" An Efficient Association Rule Hiding Algorithm for Privacy Preserving Data Mining" International Journal on Computer Science and Engineering (IJCSE), vol.3 ,pp.2792- 2798,2011

[12] Yuhong Guo" Reconstruction-Based Association Rule Hiding" IEEE computer society, 2007

[13] Elena Dasseni, Vassilios S. Verykios,. Ahmed K. Elmagarmid, Elisa Bertino" Hiding Association Rules by Using Confidence and Support" 4th International Workshop on Information Hiding, 2001.

©2014, IJCSMA All Rights Reserved, www.ijcsma.com