Comparison of ISL, DSR, and New Variable Hiding Counter Algorithm of Association Rule Hiding

Kirtirajsinh Zala

Abstract— The security of the large database that contains certain crucial information, it will become a serious issue when sharing data in network against unauthorized use. Here data mining is to discover new mining association rules from large data repositories. Association rules are powerful tools for discovering new rules and relationships among the items. Data Modification and Rule hiding is one of the most important features of Data mining. Main approach of rule hiding is to protect sensitive data from disclosure. The main approached of association rule hiding algorithms to hide some generated association rules, by increase or decrease the support or the confidence of the rules. The association rule items whether in Left Hand Side (LHS) or Right Hand Side (RHS) of the generated rule, that cannot be deduced through association rule mining algorithms. The concept of Increase Support of Left Hand Side (ISL) algorithm is decrease the confidence of rule by increase the support value of LHS. Another techniques of Heuristic approach is by introducing new variable hiding counter without editing or reduction of support and confidence by introducing two new terms Confidence(modified confidence),Msupport(modified support) and Hiding counter. In this paper we will try to analyze and compare ISL,DSR and NEW HIDING COUNTER algorithm and try to find out Pro and Cons of it.

Index Terms - Data Mining, Association Rules, modified support, modified confidence, hiding counter, ISL, DSR

INTRODUCTION

Securing information against unauthorized access is an important goal of data base security and privacy. Before releasing the data sensitive information and important data should be hidden from public use so that privacy of sensitive data is manitained.many researchers have proposed several approaches for knowledge hiding, m.attallah et al.[3] was first to proposed heuristic approach for preventing data from public use.oliveria et al. presented taxonomy of attacks against sensitive data. Here many approaches for hiding sensitive data classification and clustering .in this paper we are only concern about comparing isl,dsr and new variable hiding counter and try to find out pro and cons of all three algorithm with help of example.

Here how all the three algorithms works and hide sensitive data with their approach and see how they modified transaction in database and hide more number of rules with less modification. all three algorithm related to the heuristic approach which use support and confidence to hide a rule and introduce a new variable hiding counter which will be added in number of transaction to decrease the support and confidence in order to hide a rule. In order to hide association rules three are many strategies but we will discuss here three basic strategies.

1: increase the support of the item which is in the left hand side of the rule(ISL)[1].

2: decrease the support of the item which is in the right hand side rule (DSR)[1].

3: introducing new variable hiding counter in transaction[6].

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

BACKGROUND AND RELATED WORK

There is a large amount of work related to rule hiding. Maximum researchers have worked on the basis of reducing support and confidence of sensitive rules. in data mining sensitive data is modified or trimmed out so that the sensitive data could not be identified by the data mining algorithm. Rule hiding focuses on rule hiding and frequent item set hiding. Data hiding can be divided into three subgroups: perturbation based techniques [7][8], cryptographic techniques [7][8] and anonymization based techniques .

- *Author Kirtirajsinh Zala is currently pursuing masters degree program in INFORMATION TECHNOLOGY engineering in Ganpat University, India, PH-09898980047.*
E-mail: kirtirajzala@gmail.com

Exact approaches give no side effects with optimal solution but have computational cost. Heuristic approaches uses heuristics for modifications in the database. These techniques are efficient, scalable and fast algorithms however they do not give optimal solution and may have side effects. These techniques based on support and confidence decreasing. There are two types of techniques: distortion and blocking. Distortion techniques select transactions which hold sensitive item sets and then selected items are deleted from transaction and database is modified. Blocking techniques replaces items with unknown values instead of deletion of items to modify database. The first algorithm is based on support reduction [7]. In [8] five algorithms are proposed based on hiding strategies. Not only item sets but also rules are considered through hiding in the algorithms.

INTRODUCING NEW VARIABLE HIDING COUNTER ALGORITHM

Let I = {l1, l2, l3….lm} be a set of literals, called items. Given a set of transactions D, where each transaction T is a set of items such that T ⊆ I, an association rule is an expression P→Q, where P ⊆I, Q⊆I and P ∩ Q=Ø.

$$\text{Confidence} = \frac{IPUQI}{IPI}$$

$$\text{Support} = \frac{IPUQI}{N} \quad \text{(N is number of transactions) (2)}$$

in other words confidence measures degree of correlation between the items and support measures correlation between itemsets.In association rule mining we have to find all minimum support and minimum confidence related to given support and confidence.

Here in this algorithm [6] we are introducing a new terms called Hiding counters, Mconfidence (modified confidence), Msupport (modified support).

$$\text{Confidence} = \frac{IP\cup QI}{IPI} \qquad (1)$$

The problem of mining association rule is to find all rules that have support and confidence greater then user specified minimum support threshold (MST) and minimum confidence threshold (MCT)[6].As an example[l], for a given database in following table, 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%).

$$\text{Support} = \frac{IP\cup QI}{N} \quad \text{(N is number of transactions) (2)}$$

New modified confidence for the rule P→Q is as below

$$\text{Mconfidence } (P\rightarrow Q) = \frac{IP\cup QI}{I\ P\ I + \text{hiding counter of rule } P\rightarrow Q.} \quad (3)$$

New modified support for the rule X→Y is as below

$$\text{Msupport } (X\rightarrow Y) = \frac{IP\cup QI}{N+\text{hiding counter of rule } P\rightarrow Q.} \quad (4)$$

T1	ABD
T2	B
T3	ACD
T4	AB
T5	ABD

The goal of hiding rule is to hide maximum sensitive items and prevent from disclosure as well as public use. In given transactions we assume one rule that is sensitive and hide a rule by adding hiding variable counters in it.

Suppose in **above table** we have also given a MST of 60% and a MCT of 70% .We can see four association rules can be found as below

A→B (60%, 75%)
B→A (60%, 75%)
A→D (60%, 75%)
D→A (60%, 100%)

Now we have to hide D and B [6].

	Msupport	Mconfidence	HIDING COUNTERS
A→B	(60%	75%	0)
B→A	(60%	75%	0)
A→D	(60%	75%	0)
D→A	(60%	100%	0)

First we hide *B*

T1	ABD
T2	*B*
T3	*ACD*
T4	*AB*
T5	*ABD*

Hiding B

	Msupport	Mconfidence	hiding counters
A→B	(60%	75%	0)
B→A	(60%	60%	1) ←rule is hidden
A→D	(60%	75%	0)
D→A	(60%	100%	0)

Now we Hide D

	Msupport	Mconfidence	hiding counters
A→B	(60%	75%	0)
B→A	(60%	60%	1) ←rule is hidden
A→D	(60%	75%	0)
D→A	(60%	60%	2) ←rule is hidden

ISL(INCREASE THE SUPPORT OF LEFT HAN SIDE RULE)APPROACH ALGORITHM

If we want to hide D and B, we check it by modifying the transaction T2 from B to BD (i.e. from 0100 to 0101) we cannot hide the rule D→A.

T1	ABD	1101
T2	B	0100
T3	ACD	1011
T4	AB	1100
T5	ABD	1101

Hiding D→A (ISL approach) [1]

T1	ABD	1101
T2	BD	0101 ←
T3	ACD	1011
T4	AB	1100
T5	ABD	1101

So by above explanation [1] we can see that rule D→A can not be hidden by ISL approach because by modifying T2 from B to BD (Le. from 0100 to 0101) rule D→A will have support and confidence 60% and 75% respectively.
Now we will check it by DSR approach....

T1	ABD	1101←
T2	B	1
T3	ACD	1011
T4	AB	1100
T5	ABD	1101

(Hiding D→A by DSR approach)[1]

T1	BD	0101 ←
T2	B	1
T3	ACD	1011
T4	AB	1100
T5	ABD	1101

We see bye DSL approach rule D→A is hidden as its support and confidence is now 40% and 66% respectively, but as a side effect the rule A→D is also hidden. so it affects other rule also.

The concept of Increase Support of Left Hand Side (ISL) algorithm is decrease the confidence of rule by increase the support value of LHS. It doesn't work for both side of rule; it works only for modification of LHS. In Decrease Support of Right Hand Side (DSR) algorithm, confidence of the rule decrease by decrease the support value of RHS. It works for the modification of RHS.

ANALYSIS AND CONCLUSION OF THREE ALGORITHM COMPARISONS

a) Introducing new variable hiding counters

Advantage

Here in this approach there no side effects of other rule, we just have to add hiding counter in total transactions
We just have to make modification according to minimum support and confidence in order to hide a rule

Disadvantage

This is just a theoretical approach which is in its primary stage. Further research work can be done on theoretical basis.

b) ISL and DSR approch

Advantage

Here in both approach required less number of databases scanning and prune more number of hidden rules. More number of rules can be find by database scanning.

Disadvantage

Both this approach does not hide the entire rule and goes for number of modification to hide certain rule. It doesn't work for both side of rule. if we want to hide left hand side rule we have to go for ISL algorithm and if we want to hide right hand side rule we have to go for DSR algorithm.

ACKNOWLEDGMENT

We express our profound gratitude to all faculty members Of UVPCE-GANPAT UNIVERSITY, Mehsana and Hitesh.R.Chhinkaniwala (Assist.Prof,CE) for his help whose critical suggestions, discussions and guidance cannot be valued in words to the logical conclusion of this work.

REFERENCES

1. Shyue-Liang Wang, Yu-Huei Lee, Steven Billis, Ayat Jafari "Hiding Sensitive Items in Privacy Preserving Association Rule Mining" 2004 IEEE International Conference on Systems, Man and Cybernetics.
2. R. Agrawal and R. Srikant, "Privacy preserving data mining", In ACM SIGMOD Conference on Management of Data, pages 439450,Dallas, Texas, May 2000.
3. M. Atallah, E. Bertino, A. Elmagarmid, M. Ibrahim, and V. S. Verykios"Disclosure limitation of sensitive rules,".In Proc. of the 1999 IEEE Knowledge and Data Engineering Exchange Workshop (KDEX'99), pp.45–52, 1999.
4. Wu, Y.H., Chiang, C.M., and Chen, A.L.P. Hiding sensitive Association rules with limited side effects. IEEE Transactions on Knowledge and Data Engineering, 2007,19(1):29-42
5. Text Book -Data Mining Concepts & Techniques- Jiawei Han,Micheline Kamber - Morge Kaufmann Publisher.
6. Ramesh Chandra Belwal, Jitendra Varshney, Sohel Ahmed Khan, Anand Sharma, Mahua Bhattacharya,"Intoducing New variable Hiding counter
7. J.Vaidya, C.Clifton and Y.Zhu, Privacy Preserving Data Mining. Springer.2006
8. C.Aggarwal,P.Yu,Privacy reserving Data Mining:Models & Algorithms.Springer.2008