IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 6, Ver. III (Nov – Dec. 2014), PP 60-65
www.iosrjournals.org

Privacy Preservation by Using AMDSRRC for Hiding Highly Sensitive Association Rule

Mr. Pravin R. Ponde[1], Prof. Chetan V. Andhare[2], Dr. S. M. Jagade (Ph.D) [3]
[1]*M.E, Department of Computer Science and Engineering,*
TPCT's College of Engineering, Osmanabad, Maharashtra, India
[2]*Asst Prof, Department of Information Technology,*
Dr. D. Y. Patil College of Engineering, Ambi, Maharashtra, India
[3]*Principal, TPCT's College of Engineering, Osmanabad, Maharashtra, India*

Abstract: *Researchers are needed for settling on the choice of information mining. In any case a few associations to help with some external counsellor for the procedure of information mining on the grounds that they don't have their consultant. At the time of getting counsel from the outside counsellor it may have to face risks. The loss of business intelligence and client information security and security related issues are emerging. In mining it is the main challenging issue. The owner of the data has some private information or property like association rule contained outsources database. However the mining results integrity can influence gravely if the administration supplier is not reliable. To overcome this problem we are planning to use AMDSRRC. In this paper, we propose a heuristic based algorithm named AMDSRRC (Advanced Modified Decrease Support of R.H.S. item of Rule Clusters) to conceal the exceedingly association rules with different items in consequent (R.H.S) and antecedent (L.H.S). The sensitive rule is the one having helped more prominent than or equivalent to MST (Minimum Support Threshold) and certainty more noteworthy than or equivalent to MCT (Minimum Confidence Threshold) & close to 100%. The planning is to improve efficiency and adaptability of the proposed algorithm than existing standard algorithm MDSSRC.*

Keywords: *AMDSRRC, MDSRRC, Hiding association rule, privacy preserving, Apriori algorithm.*

I. Introduction

Today, with the unpredictable development in internet, storage and data, preparing innovations, privacy preservation is the key concern in fields of businesses. If the freely available systems are to be made secure, then it must be ensured that private delicate information have been spruced out and that certain inference channels have been blocked too. The privacy preserving association rule mining has been a broadly utilized methodology concerning concealed private information by cleaning the original database. Association rule related algorithm can be separated into three principle classes, specifically heuristic methodology, border based approach, and exact methodology [6]. Heuristic methodology has been the hotly debated issue of algorithm recently. Heuristic methodologies hide delicate association rules by specifically altering, or we say, purifying the original database D, and get the discharged database D' straightforwardly from D. Heuristic based approaches can be further partitioned into two gatherings focused around data modification methods: data distortion strategies and data blocking systems. The proposed algorithm essentially focused around distortion method and data blocking technique to rundown support or confidence in order to hide sensitive rules. Association rule mining methodology is mostly used in data mining to establish relationship between item sets. Many organization opens out their private information to the outsourced for their mutual benefit to find some useful information for any decision making purposes and improve their business schemes. But this database may contain some private data and which the organization does not want to open up. The main problem of privacy plays an 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 technique are helpful to enhance the security of the database.

The proposed work represents the working of two algorithms which are MDSRRC and AMDSRRC algorithm. The MDSRRC algorithm uses in the hiding the sensible rules which are generated in the database. The sanitized database is generated by the minimum confidence threshold and minimum support threshold algorithms. The sanitized database which is generated from this algorithm is successful to hide all sensitive rules. This database maintains the quality of the database. But this algorithm has some limitations like database owner specifies the sensitive rules. Next it does not mention the number of sensible rules and again what are those sensitive rule is not known. The proposed algorithm overcomes all the above limitations and increases the efficiency of MDSSRC by applying a novel idea of selecting highly sensitive rules to hide sensitive data. In this paper, knowledge about the existing work is mentioned in the literature survey section, i.e. in Section II. Some background details and formulation is mentioned in Section no III. Proposed system is mention in section IV.

Experimental result is shown in Section V. Final Conclusion and the future works are mentioned in the Section VI.

II. Literature Survey

Methods designed or implemented for association rule mining is as follows:

1. Heuristic Based Approaches:

They are divided in to two techniques:

A. Data Distortion Technique

In this type we replace the values from 1 to 0 or 0 to 1. Again this has two basic approaches for rule hiding. First it reduces the support of rules and second reduces the confidence of rules. Verykios. studied this concept and implement and proposed new five algorithms, this algorithms used for hiding the sensitive knowledge of database, this can be possible by reducing the support or confidence of the sensitive rules. Hiding of association rues is done by first three algorithms and hiding of large item sets are related to algorithms 2.b and 2.c. Oliveira was responsible for improving the balance between the two, sensitive knowledge and discovered pattern, which provided better privacy. A method that reduced the side effects on sanitized database introduced by Y H Wu in this method two algorithms are described, in the first if the item is present in the left side then algorithm increases the support of the sensitive item. In the second algorithm if the sensitive item present in the right hand side then algorithm decreases the support of the sensitive item.

1) Pros:
- It can be scaled to very large data sets.
- It is easy to utilized the new distance function rather than matching the old one.

2) Cons:
- The main problem we are facing is on with transactional sets of binary data, flipping entry values.
- These values have the number of side effects on the non-sensitive rules.

B. Data Blocking Technique:

First Y. Saygin in purpose blocking technique in order to decrease or increase items support, replaced 0's or 1's by the sign "?". So it is difficult for anyone to finding the value which is stored behind the "?". This technique provides some privacy, the more efficient approaches were proposed by Wang and Jafari. At the time of hiding many rules at a time, they require few numbers of scans for databases and cut more number of rules.

1) Pros
- One of the best attractive thing about the blocking approach is that it maintains the truthfulness of the data.
- With the use of hiding process it reduce the disclosure of sensitive entries.

2) Cons
- For the non-sensitive rule they have the number of side effects.
- This approach is restricted up to low dimensional data sets and binary based dataset.

2. Border based Approaches:

Border based approach was proposed by Sun and Yu, this approach used to modify the borders of the original database, it modify the lattice of the frequent and infrequent item sets. Hides sensitive association rules by a border are formed to separating the infrequent and frequent item sets. In this approach by using border value of non sensitive item, separation of positive and negative value of the border from the item set. Then the value which is Minimum affected is selected. For minimum side effect purpose modification is done by greedy selection.

In this approach by using the opposite values, i.e. by using opposite negative and positive borders values of the database, after that try to cut down the item set with sensitive data and with the negative border. In this approach the positive border value with highest support and maximum distance from the border is selected.

1) Pros:
- From the result database the quality of database can be well maintained by the controlling modification.
- For selecting the modification with the minimum side effect is one of the application of the border-based approached.

2) Cons

Border based approach used to separated data along border have bad support.

3. Cryptographic based techniques:

Most of the times multiple organizations want to share the private data, but without losing their sensitive data. So this technique demand for cryptographic protocol which can divide into

a. Vertically partitioned distributed data

For the secure calculation of the idea of secure sum is the use of this technique. This technique also includes secure calculation of the union set and size of the scalar product and the interactions sets. These techniques used the vertical and horizontal partitioning technique. This technique describes the use of scalar dot product, for counting the frequent item sets.

b. Horizontal partitioned distributed data.

Finds global frequent item sets while ensuring no loss of inter-site information. It calculates support degree inter-sites secure sum.

4. Recent work

a. Ling Qiu for outsourcing association rule mining at the time of protecting BI and the privacy of the customer the approached is proposed. They proposed Bloom filter based approach. It can outsource the mining task for protecting business intelligence and the customer data privacy, and simultaneously maintain the result for precision mining which save storage space requirement without any running time and the mining process.

b. Mohammad A. Ouda represents the PPDM method for horizontally partitioned of the data. The proposed algorithm used RSA encryption and homomorphism technology which is same time secured. No any global computation carrying the data at the centralized site but the algorithm named as KNN has need to be conduct locally for every site.

c. C N Modi proposed an algorithm named as DSRRC. This algorithm maintain privacy and quality of database. This algorithm used to improve the quality of database.

d. Laks V. S. Lakshmanan proposed the model for association rule for privacy preserving from the outsourced Database Transaction. This method solves the problem for preserving the mining of frequent pattern on an encrypted outsourced transaction database placed at cloud. Where they assume a traditional model from which the advisor knows the exact frequency of the item and the domain of the item that where it is located. For identifying the cipher items they can used this knowledge.

III. Background And Problem Definition

An association rule is an implication of the form $X \rightarrow Y$, where X, Y are Item sets, and $X \cap Y = \emptyset$. We say the rule $X \rightarrow Y$ holds in the database D with confidence c if $|XUY|/|X| \geq c$. It can also be said that the rule $X \rightarrow Y$ has support s if $|XUY|/|D| \geq s$. Note while the support is a measure of the frequency of a rule, the confidence is a measure of the strength of the relation between sets of items. The most popular association rule mining problem aims to find all significant association rules. A rule is significant if its support and confidence is no less than the user specified Minimum Support Threshold (MST) and Minimum Confidence Threshold (MCT). To calculate the significant rules, an association rule mining algorithm first finds all the frequent item sets and then derives the association rules from them. On the contrary, the association rule hiding problem aims to prevent some of these rules, which is referred as "sensitive rules", from being mined. Given a data set D to be released, a set of rules R mined from D, and a set of Highly Sensitive Rules $HSR \subseteq R$ to be hided, how can we get a new data set D', such that the rules in HSR cannot be mined from D', while the rules in R- HSR can still be mined as many as possible.

IV. Proposed Approach

The implementation of this method is divided into five different modules. The modules are Binarization, Apriori algorithm, sensitive rule generation, AMDSRRC and MDSRRC algorithm and creation of sanitized database. The general architecture of method is shown in Fig 1.To understands AMDSRRC following example is illustrated. In Table.1, transactional database D is shown. With 3 as MST and 40% as MCT, the possible frequent item sets using Apriori algorithm are:

Table.1 Original Dataset D

TID	Items
T1	a b c d e
T2	a c d
T3	a b d f g
T4	b c d e
T5	a b d
T6	c d e f h
T7	a b c g
T8	a c d e
T9	a c d h

Frequent 1-itemsets:
a, b, c, d, e 1st Level
Frequent 2-itemsets:
ab, ac, ad, bc, bd, cd, ce, de 2nd Level
Frequent 3-itemsets:
abd, acd, cde 3rd Level

A. Apriori Algorithm

Apriori algorithm is a classic algorithm basically this algorithm is used in the data mining for learning the association rules. Matrix Apriori algorithm is the result of analysis of basic two association algorithm which are the result of analysis of two algorithms named as Apriori algorithm and FP-growth algorithm. The algorithm has the successfully generates the frequent patterns and after that from the generated patterns it created the association rules this are the two steps of the matrix Apriori algorithm. the algorithm performs same as the Apriori algorithm in the second step but different in the first step.

- **System Architecture:**

The sorting of the transaction is done in decreasing order of their sensitivity, only if transaction has the value is0. Selecting the first transaction form the sorted transaction with higher sensitivity, deleted item is0 from the transaction it is the process of initialization of rule hiding. After that all the sensitive rule which contain support and confidence update it. If any rule is remaining and it has below the MST and MCT respectively then delete it from SR. Continue this process by selecting transaction with higher sensitivity and deleting is0 from it. When all sensitive rule is hidden this process is terminated, means this process is continue until the entire sensitive rule is hidden. The sanitized database is generated by updating, modifying updated transaction into new database. Sanitized database D' preserves the privacy of sensitive information and maintains database quality. The fig 1 shows system architecture.

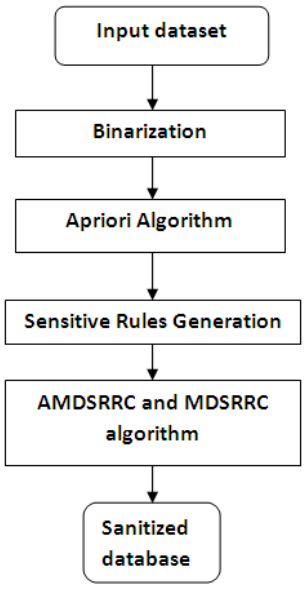

Fig1 .System Architecture

B. Sanitized Database Generation:

The possible generated association rules by Matrix Apriori algorithm are as follows: . Let the database owner specify rule a →bd, a→cd and d→ac as sensitive rules. Then select transaction with the highest sensitivity and delete is0 item from that transaction. Update confidence and support of all the sensitive rules. Sort transactions which support is0, and delete the is0 from transaction with highest sensitivity, then delete the is0 from transaction with highest sensitivity. Finally all the sensitive rules are hidden.

V. Experimental Result

We have tested this application on standard transactional database set which is shown below in Table II. The Binarization technique is applied on it which is shown in table III. IN table III transaction with its sensitivity is shown. In table IV sanitized database is generated with its entire input item, after first deletion of item from its first transaction. Table V final sanitize database, with all sensitive rules are hidden.

Parameter	Existing algorithm (MDSRRC)	Proposed (AMDSRRC Algorithm)
HF	0%	0%
MC	26.6%	10%
AP	0%	0%
DISS(D, D')	5.4%	5.4%
SEF	26.66%	10%

Table 2. Performance Results

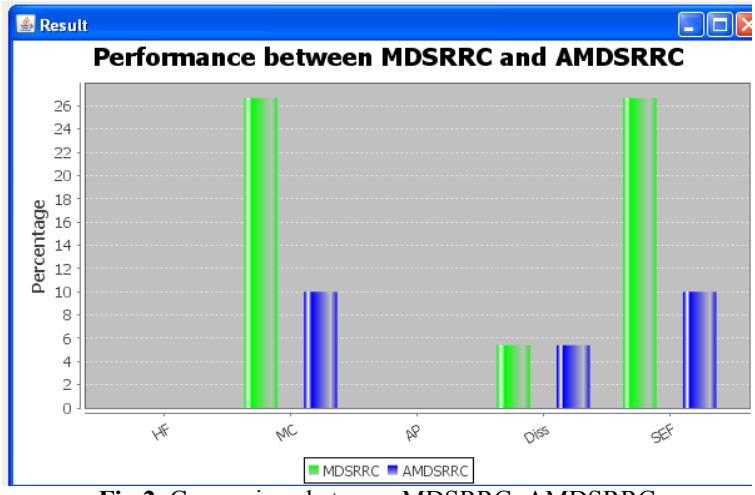

Fig 2. Comparison between MDSRRC, AMDSRRC

Fig2. Shows that performance of the AMDSRRC is better than MDSRRC in terms of database quality parameter. So MDSRRC hide sensitive rules with minimum modifications on database and maintain data quality.

VI. Conclusion

The MDSRRC algorithm hides the sensitive association rules with the modification on database for maintaining transaction database quality and the side effect on database reducing. In this model we outsourcing database on server or any service provider and security of sensitive data maintained by encryption policy. Our algorithm hides highly sensitive association rules with very few modifications on database ultimately maintain data quality. The Side Effect Factor (SEF) obtained using proposed algorithm is very much reduced than MDSRRC which means maximum non sensitive rules retain along with disappeared highly sensitive rules. In future, AMDSRRC algorithm can be extended to increase the efficiency and reduce the side effects by minimizing the modifications on database. We had to improve the AMDSRRC algorithm in the future like algorithm can help to reduced side effect of modification on datasets also increases the efficiency. In this paper we also discussed about the privacy preserving methodology.

References

[1]. Mr. Pravin R. Ponde and Dr. S. M. Jagade, "Privacy Preserving by Hiding Association Rule Mining from Transaction Database" In International Organisation of Scientific Research Journal. Vol-16, Issue 5, V2, Sep-Oct 2014, PP 25-31

[2]. Mr. Pravin R. Ponde and Dr. S. M. Jagade, "Maintaining Privacy and Data quality to hide sensitive items from Database" in International Journal of Application or Innovation and Engineering and Management, Vol-3, Issue-7, July 2014. PP. 355-361.

[3]. Nikunj H. Domadiya, U. P. Rao,"Hiding Sensitive Association Rules to Maintain Privacy and Data Quality in Database" in IEEE Third International Advace Computing Conference (IACC), PP 1306-1310, 2013.

[4]. 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.

[5]. X. Sun and P.S. Yu "A Border-Based approach for hiding the frequent item sets" In Proc. Fifth IEEE Int'I conf. data mining (ICDM '05), pp. 426-433 Nov 2005.

[6]. V. Verkios and A. Gkoulalas- Divanis, A Survey of association rule hiding method for privacy, ser. Advance in database systems. Springer US, 2008, vol. 34.

[7]. Charu C. Aggrawal, Philip S. Yu, privacy preserving data mining models and algorithm. springer publishing company incorporated, 2008, pp. 267-286.

[8]. Y. Guo, 'Reconstruction based association rule hiding', in proc. Of SIG<OD2007 Ph.D. Workshop on innovative database research 2007(IDA2007), 2007.

[9]. J. vaidya and C. Clifton, 'privacy preserving association rule mining in vertically partitioned data', In proc. Int'I Conf data mining pp. 639-644 july 2002.

[10]. M. Atallah, E. Bertino, A. Elmagarmid, M. Ibrahim and V. S. Verkios 'disclosure limitation of sensitive rules', in proceedings of the 1999 IEEE knowledge and data engineering exchange workshop (KDEX), pp. 45-52, 1999.

[11]. Han Jiawei and Kamber, Micheline. 'data mining concepts and tchniques' 2006. Morgon Kaufmann sanfransisco, C.A.
[12]. K. Wang, Y. He, J. Han, Pushing Support Constraints In: Association Rule Mining. IEEE Transactions on Knowledge and Data Engineering.
[13]. S.-L. Wang, D. Patel, A. Jafari, and T.-P. Hong, "Hiding collaborative recommendation association rules," Applied Intelligence, vol. 27, pp. 67–77, 2007.

Authors

Mr. Pravin R. Ponde, M.E, Department of Computer Science and Engineering, TPCT's College of Engineering, Osmanabad, Maharashtra, India
Email-id: pravinpondetpct@gmail.com

Prof. Chetan V. Andhare, Asst. Prof, Department of Information Technology, Dr. D. Y. Patil College of Engineering, Ambi, Talegaon, Maharashtra, India
Email-id: chetan.andhare@gmail.com

Dr. S. M. Jagade, Ph.D, Principal, TPCT's College of Engineering, Osmanabad, Maharashtra, India. Email-id : smjagade@yahoo.co.in