IIETA International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingenierie des Systemes d'Information
Vol. 24, No. 1, February, 2019, pp. 47-50
Journal homepage: http://iieta.org/Journals/isi

Efficient Method in Association Rule Hiding for Privacy Preserving with Data Mining Approach

Kurapati Praveena[1*], Gudla Sirisha[2], Satukumati Suresh Babu[3], Panchala Sambasiva Rao[4]

[1] CSE Department, Vignan's Nirula Institute of Technology and Science for Women, Peda palakaluru, Guntur 522009, India
[2] Vignan Institute of Technology and Science, Deshmukhi, Hyderabad, India
[3] Department of Information Technology, Vignan's Foundation for Science Technology and Research. (Deemed to be University), Guntur 522213, Andhra Pradesh, India
[4] Vignan's Lara Institute of Technology & Science, Vadlamudi, Andhra Pradesh 522213, India

Corresponding Author Email: kpraveena.511@gmail.com

https://doi.org/10.18280/isi.240106

Received: 11 November 2018
Accepted: 23 January 2019

Keywords:
confidence, support, association rules, item sets, data mining, association rules, privacy preservation, sensitive association rules

ABSTRACT

Security protecting and information mining is an inspection zone worried about the protection driven from identifiable data when measured for information mining. This paper tends to the security issue by allowing for the protection and algorithmic necessities at the same time. The target of this paper is to execute an Association hiding calculation for safeguarding information mining which would be proficient in giving secrecy and enhance the execution when the database stores and recovers immense measure of information. One of the procedures of information mining is association lead mining. Association rules Hiding. (ARH) is one of the real issues in the information mining space. The association rules hold numerous mysteries. So before distributing, these tenets must be hidden. Sensitive data must be covered up, since uncovering the mystery or critical association data may cause issues. In our paper, Privacy protection is finished by Association rule hiding. Amid hiding of delicate association rules, false guidelines are not created and least alteration degree is accomplished and data isn't lost.

1. INTRODUCTION

Information mining is a notable examination field to find profitable example from gigantic measure of information. These examples give significant data which is delineate regarding bunches, choice trees and association rules. So the presentation dangers of secret data are expanded when the information is discharged to the unknown gatherings [1]. Discovering obscure examples while not uncovering essential data is one of the greatest difficulties of information mining. Thinking about this, it winds up basic to shroud secret information in database. Protection saving and information mining system gives novel approach to take care of this issue [2]. Association rules hiding is one of the techniques for security maintaining to ensure the association rules which is deliver by association administer mining. Association rule hiding is the philosophy of adjusting the first databases in such how that specific secret association rules without influencing the information and the non-delicate rules [3-4].

To conceal secret association rules several protection saving strategies is utilized, Association rules hiding is one of them. Conventional association rules hiding calculation mean to enhance the first database with the end goal that no delicate association lead is obtain from it [5]. Association run hiding technique fuse diminishing the help or certainty of guidelines and diminishing the help of incessant item sets which contain delicate principles. The help of P in exchanges which isn't supporting Q will be expanded by diminishing certainty of rules and diminishing the help of Q in exchanges supporting both P and Q. [6] The issue can be expressed as takes after: Given a value-based database DB, a set RL of standards mined from database DB, least certainty and least help. RL has a subset RLH, it is an arrangement of secret association rules which are to be covered up. The object is to change DB into a database DB such that all non-secret standards in RL could in any case be mined from DB however no association administer in RLH will be mined [7]. Figure 1 demonstrates the general system for association control stowing away.

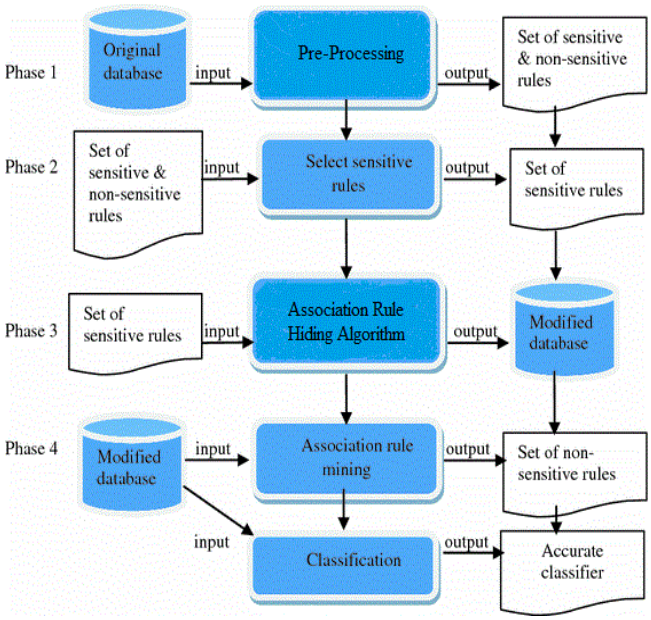

Figure 1. General structure for ARH

Association rule hiding systems go for cleaning the first database with a specific end goal to accomplish the

accompanying goal [8].

(1) The method not decides that is considered as delicate from the proprietor's point of view and can be mined from the first database at pre-indicated limits of certainty and support, can be additionally uncovered from the purified database, when this database is mined at the same or at higher edges [9].

(2) All the non delicate tenets that show up when mining the first database at pre-specified limits of certainty and support can be effectively mined from the purified database at similar edges or higher [10].

(3) No decision that was not gotten from the first database when the database was mined at pre-determined edges of certainty and support, can be gotten from its purified partner when it is mined at the same or at higher edges [11].

2. LITERATURE SURVEY

Abedelaziz Mohaisen et al. [1] utilized information contortion methods to adjust the classified information esteems so that the surmised unique information dissemination could be gotten from the adjusted adaptation of the database. The mined principles likewise were worked out of the first guidelines. Agrawal and Srikanth [2], utilized desire amplification with bending for recreating the first information dissemination. This reproduced dispersion is utilized to build a characterization display. Measurements utilized as a part of these calculations were effectiveness and symptoms. These calculations were first of their kind sequestered from every association rule. Reactions of these calculations were additionally high.

Bertino E et al. [4] goes for adjusting security and revelation of data by endeavoring to limit the effect on cleaned exchanges and furthermore to limit the coincidentally covered up and phantom guidelines. The utility in this work is estimated as the quantity of non-delicate guidelines that were shrouded in view of the reactions of the information alteration process. A disinfection procedure is displayed by the creators to hinder forward induction assault and in reverse derivation assault to stow away secret standards. The work portrayed by Bikramjit Saikia et al. [5] broadens the purification of secret extensive item sets to the sterilization of delicate standards.

Numerous administer hiding methodology is first proposed by the creators. In this work creators propose techniques and an arrangement of calculations for hiding secret information from information by insignificantly annoying their qualities. These calculations are effective and require two sweeps of the database regardless of the quantity of delicate things to stow away. The hiding methodologies proposed depend on decreasing the help and certainty of principles that indicate how noteworthy they are.

The imperative on the calculations proposed is that the adjustments in the database presented by the hiding procedure ought to be constrained, such that the data brought about by the procedure is insignificant. Wang et al. [7] likewise proposed a way to deal Forward-Inference Attacks, in the cleaned database created by the cleansing procedure.

Systems like WSDA, PDA [8] and Border-Based [9] enhanced the underlying heuristic calculations to voracious calculations which discovers neighborhood ideal alteration. These methodologies endeavored to insatiably choose the alterations with insignificant symptoms on information utility and exactness.

Association rules are spoken to by if/at that point explanations. These announcements help to discover connections between apparently random information in a social database or other data store. There are different strategies or calculations are accessible to create association leads and to conceal delicate association rules. These calculations are as per the following.

Apriori calculation is utilized to produce the association rules [12]. In association control hiding strategy the single predecessor and subsequent are chosen. Association rules are valuable in advertise bushel database. This principle demonstrates client conduct in advertise crate database [13].

3. METHOD FOR ASSOCIATION RULE HIDING

The goal of association rules hiding calculation is to shroud certain secret information with the goal that they can't be found through information mining procedures. In this exploration work, it is expected that exclusive secret things are given and propose one calculation to change information in database with the goal that delicate things can't be concluded through association rules mining calculations [14].

All the more particularly, given an exchange database D, a base help, a base certainty and an arrangement of things T to be shrouded, the goal is to alter the database DB with the end goal that no association rules containing T on the correct hand side or left hand side will be found [15].

The proposed association rule hiding calculation depends on two calculations to be specific Increase Support of Left hand side (ISL) and DSR. (Decrease Support of Right hand side) to conceal valuable association control from exchanges information with parallel properties [16].

In ISL technique, certainty of an administer is diminished by expanding the help estimation of Left Hand Side. (LHS) of the run the show.

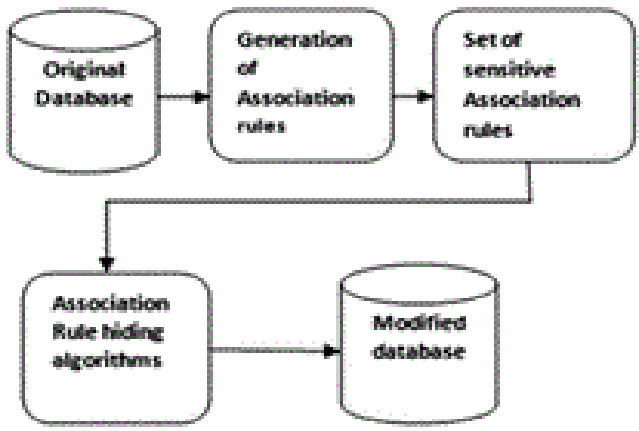

Figure 2. Flow of proposed method

For this reason, just the things from LHS of a manage are decided for adjustment. In DSR technique, certainty of an administer is diminished by the help estimation of Right Hand Side. (R.H.S.) of a run the show [17]. For this reason, just the things from R.H.S. of a control are decided for adjustment [17].

Keeping in mind the end goal to shroud an association lead, $P \rightarrow Q$, either diminishes its help or its certainty to be littler than client indicated. To diminish the certainty of an administer, either expands the help of P, the LHS of the manage, yet not support of P U Q, or abatement the help of the thing set P U Q. For the second case, diminish the help of Y,

the correct hand side of the run, it would decrease the certainty quicker than essentially lessening the help of P U R [18].

To diminish support of a thing, the framework will alter one thing at any given moment by changing from 1 to 0 or from 0 to 1 of a chose exchange. In light of these two ideas, another association lead hiding calculation for hiding secret things in association rules has been proposed [19].

In the proposed calculation, a control P → Q is covered up by diminishing the help estimation of P U Q and expanding the help estimation of P. That can increment and abatement the help of the LHS and RHS thing of the run correspondingly [20].

This calculation first endeavors to shroud the principles in which thing to be concealed i.e, P is in right hand side and afterward attempts to conceal the standards in which P is in left hand side. For this calculation t is an exchange [21], T is an arrangement of exchanges, RL is utilized for control, RHS. (R) is Right Hand Side of run RL, LHS. (R) is the Left Hand Side of the rules R, Confidence. (R) is the certainty of the manage R, an arrangement of things H to be covered up [22].

Enhanced DSR Algorithm

1. Sort the given database according to Relevance count in descending order.
2. After Sorting the database then perform pre-processing on it.
3. Then Calculate DC = C_1 − Cx × MCT + 1.
4. Find T = t in DB | t fully support U;
5. Store all Transactions T in DDB.
6. Choose the first transaction t from T;
7. While (DC > 0)&&(T!=NULL)
   - 7.1 Modify t by putting 0 instead of 1 for RHS item;
   - 7.2 Check for loss of rule if yes then go to step 7.4
   - 7.3 Remove and save the transaction t from T.
   - 7.4 Change the relevance count accordingly and decrease the value of DC by 1.
   - 7.5 Consider Next Transaction.
   - 7.6 end
8. Calculate Confident level of t in T.
9. If (DC==0)then
   - 9.1 Rule Hiding is Done.
   - 9.2 Otherwise Rule Hiding Not Possible.

The above proposed algorithm is the enhancement of DSR algorithm in which Association rule Hiding is done effectively and accurately in less time.

4. RESULTS

In this paper, we take cancer data set information. Dataset is made out of 2654 records. Every patient is portrayed in informational index by 12 characteristics [23]. All characteristics are numerical values. To start with we analyze the quantity of association rules of calculations by changing from 1586 to 2654.

The Proposed technique is compared with the existing methods where the proposed method in very less time performs more rule hiding efficiently.

Table 1. Numbers of rules hide

Dataset	Number of Rule Hide	
	Exsiting	Proposed
2654	18	26
1994	23	32
1100	25	36
608	26	38

Table 2. Execution time require

Dataset	Execution Time (Second)	
	Exsiting	Proposed
2654	32	19
1994	29	17
1100	23	14
608	21	10

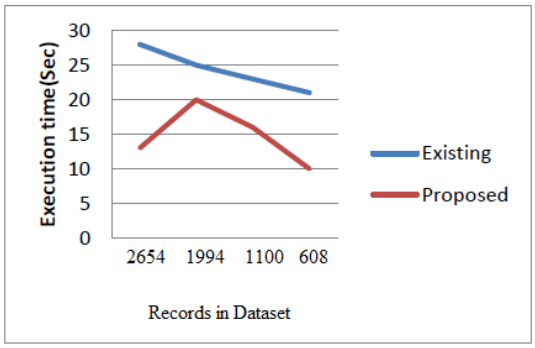

Figure 3. Execution time

5. CONCLUSION

Information mining is a notable examination field to find profitable example from gigantic measure of information. A novel approach for connection control mining using conFigure uration technique is compelling with diminishing in different conditions inspecting of database and less memory space. Enhanced DSR algorithm is proposed which is used for association rule hiding which performs better on the dataset considered and by using this method sensitive information can be covered avoiding unauthorized access.

REFERENCES

[1] Mohaisen A, Hong D. (2008). Privacy preserving association rule mining revisited. Journal of the Computing Research Repository 1-16.

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

[3] Agrawal R, Strikant R. (2004). Fast algorithms for mining association rules. Proceedings of the International Conference on Very Large Data Bases, pp. 487-499.

[4] Bertino E, Nai F, Parasiliti P. (2005). A framework for evaluating privacy preserving data mining algorithms. Journal of Data Mining and Knowledge Discovery 11(2): 121-154. https://doi.org/10.1007/s10618-005-0006-6

[5] Bikramjit S, Debkumar B. (2009). Study of association rule mining and different hiding techniques. PhD thesis, Department of computer Science Engineering, National Institute of Technology 55-63.

[6] Peng B, Geng XY, Zhang J. (2010). Combined data distortion strategies for privacy-preserving data mining. Proceedings of the 3rd International Conference on Advanced Computer Theory and Engineering 241-253. https://doi.org/10.1109/ICACTE.2010.5578952

[7] Verykios VS, Gkoulalas-Divanis A. (2008). A survey of association rule hiding methods for privacy. Privacy-Preserving Data Mining 267-289. https://doi.org/10.1007/978-0-387-70992-5_11

[8] Ma T, Wang S, Liu Z. (2010). Privacy preserving based on association rule mining. 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE). https://doi.org/10.1109/ICACTE.2010.5578938

[9] Vassilios SV. (2013). Association rule hiding methods. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 3(1): 28-36. http://dx.doi.org/10.1002/widm.1082

[10] Patil SP, Patewar TM. (2012). A novel approach for efficient mining and hiding of sensitive association rule. 2012 Nirma University International Conference on Engineering (NUiCONE). https://doi.org/10.1109/NUICONE.2012.6493184

[11] Modi CN, Rao UP, Patel DR. (2010). Maintaining privacy and data quality in privacy preserving association rule mining. 2010 Second International conference on Computing, Communication and Networking Technologies. https://doi.org/10.1109/ICCCNT.2010.5592589

[12] Wu CM, Huang YF. (2012). Privacy preserving association rules by using branch-and bound algorithm. Advances in Computer Science and Engineering 409-416. https://doi.org/10.1007/978-3-642-27948-5_54

[13] Diaz I, Rodriguez LJ, Troiano L. (2013). On mining sensitive rules to identify privacy threats. International Conference on Hybrid Artificial Intelligence Systems 232-241. https://doi.org/10.1007/978-3-642-40846-5_24

[14] Verykios VS, Gkoulalas-Divanis A. (2008). A survey of association rule hiding methods for privacy. Privacy-Preserving Data Mining: Model and algorithm 34: 267-289. https://doi.org/10.1007/978-0-387-70992-5_11

[15] Dasseni E, Verykios VS, Elmagarmid AK, Bertino E. (2001). Hiding association rules by using confidence and support. Proceedings of the 4th International Workshop on Information Hiding 369-383. https://doi.org/10.1007/3-540-45496-9_27

[16] Verykios VS, Emagarmid AK, Bertino E, Saygin Y, Dasseni E. (2004). Association rule hiding. IEEE Transactions on Knowledge and Data Engineering 16(4): 434-447.

[17] Oliveira SRM, Zaïane OR. (2002). Privacy preserving frequent itemset mining. In Proceedings of the 2002 IEEE International Conference on Privacy, Security and Data Mining (CRPITS), pp. 43-54.

[18] Bikku T, Gopi AP, Prasanna RL. (2019). Swarming the high-dimensional datasets using ensemble classification algorithm. In First International Conference on Artificial Intelligence and Cognitive Computing 583-591.

[19] Lakshman NV, Peda GA, Ashok KN. (2018). Different techniques for hiding the text information using text steganography techniques: A survey. Ingénierie des Systèmes d'Information 23(6): 115-125.

[20] Peda GA, Lakshman NV, Ashok KN. (2018). Dynamic load balancing for client server assignment in distributed system using genetical gorithm. Ingénierie des Systèmes d'Information 23(6): 87-98.

[21] Lakshman NV, Peda GA. (2017). Visual cryptography for gray scale images with enhanced security mechanisms. Traitement du Signal 35(3-4): 197-208. https://doi.org/10.3166/ts.34.197-208

[22] Peda GA, Lakshman NV. (2017). Protected strength approach for image steganography. Traitement du Signal 35(3-4): 175-181. https://doi.org/10.3166/ts.34.175-181.

[23] Gopi A. (2015). Designing an adversarial model against reactive and proactive routing protocols in MANETS: A comparative performance study. International Journal of Electrical & Computer Engineering 2088-8708.