

# A Review on Methodologies of Hiding Association Rules for Privacy Preserving

**Bhushan Mahajan**

Assistant Professor

Department of Computer Engineering

VPCOE, Baramati

*Abstract*— Data mining is the powerful technology to extract pattern or knowledge from huge dataset. Nevertheless, misuse of such technology may lead to the exposé of important information. Privacy preserving with data mining (PPDM) is new explore direction for disclosure of sensitive knowledge. There are various techniques used in Privacy preserving with data mining to hide association rules generated by association rule generation algorithms like Apriori. Main objective of privacy preserving data mining is to discover association rules and to conceal sensitive association rules. Association rule hiding is the process of modifying original database in such way that sensitive rules are moved out. In this paper, a survey of the variety of approaches of association rule hiding has been described.

**Key words:** Data Mining, Privacy Preserving Data Mining, Association Rules, Apriori

## I. INTRODUCTION

Information mining strategies empower individuals to discover helpful data from the expansive database. There are different systems of information digging that are valuable for removing data from expansive database. Affiliation guideline is a standout amongst the most famous information mining procedures [1] being used. Numerous associations offers there database for shared advantages, this has expanded the revelation dangers when the information is discharged to outside gatherings. For instance, let a material store that buy shirts from two organizations,

Levi's and United Colors of Benetton, Levi's applies information mining strategies and mines affiliation rules identified with United shades of Benetton, by applying information mining methods to touchy affiliation standards of

Joined shades of Benetton, Levi's had found that United shades of Benetton is putting forth 40% of markdown furthermore client's purchase pants with shirt, so Levi's offers half rebate on shirts furthermore 20% of markdown on pants, This is the way clients of United Colors of Benetton will now move to Levi's. This Scenario prompts the exploration of touchy learning (or lead) stowing away in database.

To overcome abuse of information mining PPDM (Privacy Preserving Data Mining) was initially presented by Agrawal and Srikant in 1993. Aim of PPDM is to protect security delicate learning from revelation. A large number of the specialists have endeavored to safeguard protection for touchy affiliation rules.

Whatever remains of the paper is sorted out as takes after; in segment 2 Association rules mining, Section 3, Association standard concealing, Section 4 Association guideline concealing Approaches, Section 5 Literature Review, Section 6 Conclusion, Area 7 References.

## II. ASSOCIATION RULE MINING

Affiliation principle mining has been utilized as a part of numerous application spaces of information mining. A few applications are, discovering designs in medicinal database, business examination, market investigation, extraction of data or learning from programming metric, and so on. It was initially presented by Agrawal et al in 1993[1]. Let  $I = \{I_1, I_2, \dots, I_m\}$  is an arrangement of things,  $T = \{T_1, T_2, \dots, T_n\}$  is an arrangement of exchanges, Each of which contains things of the itemset  $I$ . Every exchange  $t_i$  is an arrangement of things such that  $t_i \subseteq I$ . An affiliation guideline is a ramifications of the structure:

$X \rightarrow Y$ , where  $X \subset I$ ,  $Y \subset I$  and  $X \cap Y = \emptyset$ .  $X$  (or  $Y$ ) is an arrangement of things, called itemset. In the standard  $X \rightarrow Y$ ,  $X$  is known as the precursor,  $Y$  is the resulting. Clearly the estimation of the precursor suggests the estimation of the resulting. The forerunner, likewise called the —left hand sidel of a tenet, can comprise both of a solitary thing or of an entire arrangement of things. This applies for the resulting, called the —right hand sidel, too. By and large, just those thing sets that satisfy a specific bolster necessity are mulled over. Backing and certainty [2] are the two most critical quality measures for assessing the interestingness of a principle. The backing of the tenet  $X \rightarrow Y$  is the rate of exchanges in  $T$  that contain  $X \rightarrow Y$ . It decides how visit the standard is pertinent to the exchange set

T. The support of a rule is represented by the formula,

$$\text{Support}(X \rightarrow Y) = \frac{|XUY|}{|D|} \quad (1)$$

Where  $|D|$  is complete number of exchanges in database  $D$ . Assume the backing of a thing is 0.2%, it implies just 0.2 percent of the exchange contain acquiring of this thing.

The certainty of a tenet is characterized as rate of the quantity of exchange that contains  $XUY$  to the aggregate number of records that contains  $X$ . The certainty of the principle is spoken to by recipe,

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{Support } |XUY|}{\text{Support } |X|} \quad (2)$$

Where  $|X|$  signifies the quantity of exchange in database  $D$  that contains itemset  $X$ . Certainty is the measure of quality of affiliation tenet. Assume the certainty of the affiliation guideline ( $X \rightarrow Y$ ) is 90%, it implies that 90% of the exchange that contain  $X$  additionally contains  $Y$  together.

A principle  $X \rightarrow Y$  is solid if bolster  $(X \rightarrow Y) \geq \text{min\_support}$  and certainty  $(X \rightarrow Y) \geq \text{min\_confidence}$ , where  $\text{min\_support}$  and  $\text{min\_confidence}$  [2] are two given least

edges, min\_support and min\_confidence are two client characterized values.

### III. ASSOCIATION RULE HIDING TECHNIQUES

AIS Algorithm

Apriori Algorithm FP-Tree Algorithm

Dynamic Item Set Counting Algorithm Pincer Search Algorithm Fapriori Algorithm

#### A. AIS Algorithm

AIS (Agrawal, Imielinski, and Swami) calculation [1] was first calculation for mining affiliation rules in [Agrawal et al]. There are two stages in the calculation, first stage is to create visit thing sets and second stage to produce certain and continuous affiliation standard. This calculation basically created to produce extensive thing sets in an exchange database. In this calculation one and only thing subsequent affiliation tenets are produced, which implies that the ensuing of those principles just contain one thing, for instance we just create rules like  $A \cap B = C$  however not those standards as  $A = B \cap C$ .

#### B. Apriori Algorithm

Apriori calculation [3] created by Agrawal and Srikant is most broadly utilized calculation as a part of Association standard mining. There are two stages in this calculation; first stage is to create the regular thing sets by producing the bolster number and least backing of thing sets. Second stage is to produce the guideline by client characterized parameter called least certainty. Apriori is the base up pursuit calculation, moving upward level shrewd in the grid. There are two downsides of apriori calculation, one is the mind boggling competitor set era so more often than not space and memory is squandered in applicant set era, second is different output of database.

#### C. FP-Tree Algorithm

FP-tree [4] was produced by Han et al in 2000, FP-tree calculation works like partition and vanquish way, there are two output in the database, in the main sweep rundown of successive thing sets created by sorting recurrence in dropping request. In second output the database is packed into a FP-tree. At that point FP-development begins to dig the FP-tree for every thing whose backing is bigger than least backing by recursively constructing its contingent FP-tree. The calculation performs mining recursively on FP-tree. The issue of discovering incessant thing sets is changed over to looking and building trees recursively.

It is appropriate to incremental database, since when new datasets arrives then reiteration of entire procedure is finished.

#### D. Dynamic Item Set Counting Algorithm

Dynamic Item Set Counting was proposed by canister et al in 1997, basis of element set numbering calculation is that it works like train pruning over information, with stops at interims M between exchange document, it has made one disregard the information, and it starts from the very beginning again from the earliest starting point for the following pass[5].

#### E. Pincer Search Algorithm

The pincer-search figuring calculation begins from the littlest arrangement of continuous itemsets and moves upward till it achieves the biggest incessant itemsets. The pincer-see calculation depends on the rule of discovering successive thing sets in a base up way, in any case, at the same time, it keeps up a rundown of maximal continuous thing sets [5].

In this calculation, notwithstanding numbering the backings of the hopeful in the base up bearing, it additionally tallies the backing of the itemsets of some utilizing as a part of top up methodology.

#### F. Apriori Algorithm

Mihir et al [23], presented FApriori calculation, is a changed Apriori calculation in light of checkpoint. They proposed a strategy which can be blend of Apriori calculation and diminished stockpiling required to store hopeful and execution time by decreasing CPU time.

They presented checkpoint idea in view of bolster worth to decrease execution time and storage room required to store hopeful produced amid filtering of datasets [23].

### IV. ASSOCIATION RULE HIDING APPROACHES

Information mining methods has more extensive applications now days with some of information security issues. Affiliation Rule covering up is utilized as a part of PPDM to manage information security issues. Affiliation guideline stowing away is the procedure to change over the first database into cleaned database, so that touchy affiliation standard will be covered up.

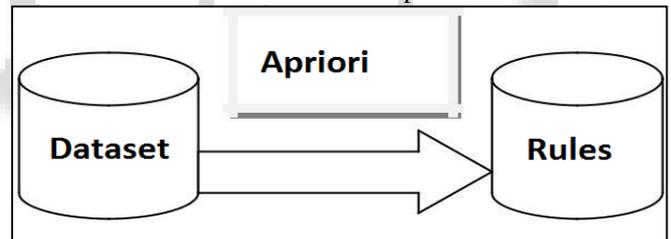


Fig. 1:

Many approaches have been proposed for Hiding Association Rules are classified into five categories:

#### A. Heuristic Based Approach

These approaches further divided into two groups:

##### 1) Data Distortion Techniques

This method forever erases some itemsets from the database, information twisting procedure attempt to shroud affiliation rules by diminishing or expanding backing to reduction or expansion the bolster quality is changed by new esteem. For instance, it changes "1" to "\_0" or "\_0" to "\_1". By diminishing or expanding bolster things in delicate tenet, or by changing estimation of certainty of things in touchy principle.

ThiM.Attallah et al. [7] proposed heuristic calculations; they additionally give verification of NP-hardness of ideal cleansing [7]. Verykios et al. [8] proposed five suspicions to shroud touchy information in database. Y-H Wu et al. [9] proposed technique to lessen the symptoms in purified database, which are delivered by different methodologies [8]. K.Duraiswamy et al. [10] proposed a bunching based way to deal with decrease the time many-sided quality of the concealing procedure.

## 2) Data Blocking Techniques

This system replaces '\_' and '0' with '?' in those exchange of database, so sections in the database are not altered. In this way, it turns out to be extremely troublesome for those who know the estimation of '?'. This system is proficient than information mutilation procedure since it doesn't erase things for all time from database.

## 3) Border Based Approach

In this methodology, cross section of the regular and rare thing sets is altered to shroud touchy affiliation rules in unique database. The itemsets amongst successive and occasional makes the outskirts, fringe separate regular and rare itemsets [6]. Sun Yu [11] were the first to propose the fringe correction process. The creators in [12] proposed more effective calculations than proposed by Sun Yu [11].

## 4) Reconstruction Based Approach

These methodologies are proficient than the Heuristic based methodologies and create less symptoms than heuristic based methodologies. In this methodology first continuous thing set is extricated from non visit thing set and security ensured information is discharged. The new discharged information is then reproduced from the sterilized learning base. This methodology, first perform information bothering and after that recreate the database. Fundamentally this methodology remakes the database in a way that all delicate data has been covered up. This technique can't promise to locate a steady one inside a polynomial time [6]. Y. Guo [13] proposed a FP tree based calculation which reproduce the first database by utilizing non normal for database and proficiently produces number of secure databases.

## 5) Exact Based Approach

This methodology was proposed by Gkoulalas and Verykios [14] for finding ideal answer for tenet concealing issue. This methodology guideline concealing issue into imperatives fulfillment issue (CSP) and comprehend it by utilizing paired whole number programming (BIP). This methodology is better however experiences high time intricacy to CSP.

## 6) Cryptographic Based Approach

Cryptographic based methodologies generally utilized as a part of multiparty calculation. The idea of secure multiparty calculation was presented in [15]. Secure Multiparty Computation is that a calculation is secure if toward the end of the calculation, no gathering knows anything with the exception of its own information and the outcomes. One approach to view this is to envision a trusted outsider – everybody gives their contribution to the trusted party, who plays out the calculation and sends the outcomes to the members.

Vaidya and Clifton [16] proposed secure methodology for sharing affiliation rules when information are vertically divided. The creators in [17] proposed secure mining of affiliation guidelines over even apportioned information.

## V. CONCLUSION AND FUTURE WORK

In this paper we have considered different affiliation principle mining procedures and methodologies. We have talked about about different real calculation for concealing touchy affiliation rules calculations. Different calculations are appropriate to static database, when new dataset arrive these calculations did not work effectively. There is need to propose new calculation for incremental datasets. Along

these lines, in future the emphasis will be on creating calculation that will apply on incremental datasets. What's more, memory use of such calculation must be diminished in future.

## ACKNOWLEDGMENT

This is to acknowledge and thanks to all individuals who played defining role in shaping this paper. Without their constant support, guidance and assistance this paper would not have been completed. In addition, the authors would like to thank to Prof. S. S. Nandgaonkar, Asst Professor at VPCOE, for her guidance, support, encouragement and advice.

## REFERENCES

- [1] R. Agarwal, T. Imielinski, and A. Swami, "Mining associations between sets of items in large databases". SIGMOD93, pages 207-216, Washington, D.C, USA, May 1993 Padam Gulwani "Association Rule Hiding by
- [2] Positions Swapping of Support and Confidence". International journal of Information Technology and Computer Science, 2012.
- [3] Rakesh Agrawal and Ramakrishna Srikant, Fast Algorithms for Mining Association Rules in Large Databases, Proceedings of the Twentieth International Conference on Very Large Databases, pp. 487-499, Santiago, Chile, 1994.
- [4] Han, J. and Pei, J. 2000. Mining frequent patterns by pattern-growth: methodology and implications. ACM SIGKDD Explorations Newsletter.
- [5] Privacy preserving data mining based on association rule-a survey S.vijayrani,dr, Dr tamilrasi, R.seethalakshmi Khyati B. Jadav, Jignesh Vania, Dhiren R.
- [6] Patel "A Survey on Association Rule Hiding Methods" International Journal of Computer Applications, November 2013.
- [7] M. Atallah, A. Elmagarmid, M. Ibrahim, E. Bertino, and V. Verykios, "Disclosure limitation of sensitive rules," in Proceedings of the 1999 Workshop on Knowledge and Data Engineering Exchange, ser. KDEX '99. Washington, DC, USA: IEEE Computer Society, 1999, pp. 45-52.
- [8] V.S. Verykios, A.K. Elmagarmid, E. Bertino, Y. Saygin, and E. Dasseni, "Association rule hiding," IEEE Transactions on Knowledge and Data Engineering, vol. 16(4), pp. 434-447, April 2004.
- [9] Y. H. Wu, C.M. Chiang and A.L.P. Chen, "Hiding Sensitive Association Rules with Limited Side Effects," IEEE Transactions on Knowledge and Data Engineering, vol.19 (1), pp. 29-42, Jan. 2007.
- [10] K. Duraiswamy, and D. Manjula, "Advanced Approach in Sensitive Rule Hiding" Modern Applied Science, vol. 3(2), Feb. 2009.
- [11] X. Sun and P.S. Yu, "A Border-Based Approach for Hiding Sensitive Frequent Itemsets," In Proc. Fifth IEEE Int'l Conf. Data Mining (ICDM'05), pp. 426-433, Nov. 2005.
- [12] Moustakides and V.S. Verykios, "A Max-Min Approach for Hiding Frequent Itemsets," Y. Guo,

- “Reconstruction-Based Association Rule Hiding,” In Proc. Of SIGMOD2007
- [13] Ph.D. Workshop on Innovative Database Research 2007 (IDAR2007), June 2007.
- [14] Gkoulalas-Divanis and V.S. Verykios, “An Integer Programming Approach for Frequent Itemset Hiding,” In Proc. ACM Conf. Information and Knowledge Management (CIKM '06), Nov. 2006.
- [15] Yao, A. C.-chi. (1986). “How to Generate and Exchange Secrets”. Exchange Organizational Behavior Teaching Journal, (1), 162-167
- [16] J. Vaidya and C. Clifton, “Privacy preserving association rule mining in vertically partitioned data,” In proc. Int'l Conf. Knowledge Discovery and Data Mining, pp. 639–644, July 2002.
- [17] Shyue-Liang Wang, Ayat Jafari “Using Unknowns for Hiding Sensitive Predictive Association Rules” IEEE 2005
- [18] Chirag Modi, Udai Pratap Roa, Dhiren Patel “Maintain Privacy and Data Quality in Privacy Preserving Association Rule Mining” IEEE 2010
- [19] Ramesh Chandra Belwal, Jitendra Varshney, Sohel Ahmed Khan, Anand Sharma, Mahua Bhattacharya, “Hiding Sensitive Association Rules Efficiently By Introducing New Variable Hiding counter”, IEEE International conference on Service Operations, Logistics and informatics, Vol.1, Oct. 2008, pp 130-134
- [20] N Domadiya and U. P. Rao, “Hiding Sensitive Association Rules to Maintain Privacy and Data Quality in Database” 2013 3rd IEEE International Advance Computing Conference (IACC), pp. 1306-1310, 2013.
- [21] Chih-Chia Weng, Shan-Tai Chen & Hung-Che Lo “A Novel Algorithm for Completely Hiding Sensitive Association Rules” IEEE Intelligent Systems Design and Applications, 2008.
- [22] Niteen Dhutraj, Siddhart Sasane, Vivek Kshirsagar “Hiding Sensitive Association Rule For Privacy Preservation”, IEEE Transactions on Knowledge and Data Engineering Year 2013.
- [23] Mihir R. Patel, Dipti P. Rana, Rupa G. Mehta, “FApriori: A modified Apriori Algorithm Based on checkpoint” 2013 International Conference on Information System and Computer Network, IEEE 2013.