International Journal of Innovative Research in Engineering Science and Technology

APRIL 2018 ISSN 2320 –981X

A SURVEY ON PRESERVING DATA IN ASSOCIATION RULE HIDING

MS.K.E.Preethiga[1], MS.G.Bhavani[2], Mrs.S.Sivakumari[3]

1.Department Of Computer Science And Engineering

Avinashilingam colleage of engineering for women

Coimbatore

preethigaeswar@gmail.com

2. Department Of Computer Science And Engineering

Avinashilingam colleage of engineering for women

Coimbatore

bhavanigb16@gmail.com

3. Department Of Computer Science And Engineering

Avinashilingam colleage of engineering for women

Coimbatore

hodcseu@gmail.com

ABSTRACT: Privacy preserving data mining is a research field that aims to protect the private information and avoid the leakage of this information during the data mining process. One of the techniques is the Privacy Preserving Association Rule Mining which aims to hide sensitive association rules. An efficient cuckoo optimization algorithm is used for the sensitive association rules hiding (COA4ARH). In this algorithm the act of hiding is performed using the distortion technique. Fitness functions are defined in this method which makes it possible to achieve a solution with the fewest side effects. Introducing an efficient immigration function it improves its ability to escape from any local optimum.

1.INTRODUCTION: The data mining process to extract constructive patterns from large data increases the concern regarding the disclosure of private information. In order to conserve the privacy of sensitive information Privacy Preserving Data Mining (PPDM) has been emerged. PPDM can be applied to different data mining techniques like prediction, classification, clustering and association rule mining. Algorithms that are developed in this domain avoid the disclosure of private information and preserve the value of non-sensitive information in the database. But there exists the following issues in preserving the privacy of sensitive information.

- *Hiding failure:* It means a part of sensitive information could be processed when the data sanitization process is incomplete.
- *Lost rule:* During sanitization process of hiding sensitive information some segments of the non-sensitive data also gets affected.
- *Ghost rule:* As a result of the process of sanitization, new and unrealistic patterns may arise in database that does not occur earlier.

The chief concern and purpose of this review is to make use of cuckoo optimization algorithm for proposing a new algorithm in privacy preserving association rule mining called hiding with the following characteristics.

- Reduced Hiding failure
- Diminished ghost rules
- Decreased lost rule.

The following section explains the various works that exist in this domain.

2.LITERATURE SURVEY

Stanley et.al., proposed a unified framework which combines techniques for efficiently hiding sensitive rules: a set of algorithms to protect sensitive knowledge in transactional databases; retrieval facilities to speed up the process of protecting sensitive knowledge; and a set of metrics to evaluate the effectiveness of the proposed algorithms in terms of information loss and to quantify how much private information has been disclosed.

Dehkordiet.al., proposed a new multi-objective method for hiding sensitive association rules based on the idea of genetic algorithms. The major purpose of this method is fully supporting security of database and keeping the utility and certainty of mined rules at highest level.

Azam Khan et.al., proposed improved genetic algorithm architecture with a new fitness function for hiding sensitive rules by reducing loss of information, lost rules and generation of ghost rules.

Akbarzadeh et.al., proposed that one of the mathematical models of production planning is survey and the problem solved by cuckoo algorithm. Cuckoo Algorithm is efficient method to solve continues nonlinear problem.

TolgaAyavet.al., proposed a solution may not be optimum where a full-exact method, without any need for heuristics. Extensive tests are conducted on 10 real datasets to analyze distance and information loss performances of the algorithm in comparison to a former similar algorithm.

Dhyanendra Jain et.al., proposed that Association rule mining is an important data-mining technique that finds interesting association among a large set of data items. The proposed approach uses the data distortion technique where the position of the sensitive items is altered but its support is never changed. The size of the database remains the same. It uses the idea of representative rules to snip the rules first and then hides the sensitive rules. This method hides maximum number of rules.

Chun-Wei Lin et.al., proposed that a novel hiding-missing-artificial utility (HMAU) algorithm to hide sensitive item sets through transaction deletion. The transaction with the maximal ratio of sensitive to non-sensitive one is thus selected to be entirely deleted. The side effects of hiding failures, missing item sets, and artificial item sets are considered to evaluate whether the transactions are required to be deleted for hiding sensitive item sets. The weights are also assigned as the importance to three factors, which can be set according to the requirement of users.

Ramin Rajabioun proposed a optimization algorithm is inspired by the life of a bird family, called Cuckoo. Special lifestyle of these birds and their characteristics in egg laying and breeding has been the basic motivation for development of this new evolutionary optimization algorithm. The cuckoo population, in different societies, is in two types: mature cuckoos and eggs. The effort to survive among cuckoos constitutes the basis of Cuckoo Optimization Algorithm.

Xin-She Yanget.al., proposed a new meta heuristic algorithm, called Cuckoo Search (CS), for solving optimization problems. This algorithm is based on the obligate brood parasitic behaviour of some cuckoo species in combination with the L´evy flight behaviour of some birds and fruit flies.

ArisGkoulalas-Divaniset.al., discussed the process of border revision, which plays a significant role towards the identification of exact hiding solutions, and provide efficient algorithms for the computation of the revised borders. On analysing the algorithms that identify exact hiding solutions, the functionality has been extended for one of them to effectively identify exact solutions .

XingzhiSunet.al., proposed study such a problem in the context of hiding sensitive frequent itemsets by judiciously modifying the transactions in the database. The quality of the sanitized database especially on preserving the non-sensitive frequent itemsets are considered. To preserve the non-sensitive frequent item sets, a border-based approach to efficiently evaluate the impact of any modification to the database during the hiding process.

Aris G koulalas et.al., proposed that the problem of securing sensitive knowledge from being exposed in patterns extracted during association rule mining. Instead of hiding the produced rules directly, the sensitive frequent itemsets are hided that may lead to the production of these rules. By trying to minimize the distance between the original database and its sanitize diversion, a novel exact algorithm

for association rule hiding and evaluate it on real world datasets demonstrating its effectiveness towards solving the problem.

Elena Dasseni Large proposed that the confidentiality issues of a broad category of rules, which are called association rules. If the disclosure risk of some of these rules is above a certain privacy threshold, those rules must be characterized as sensitive. Sometimes, sensitive rules should not he disclosed to the public since, among other things, they may be used for in ferencing sensitive data, or they may provide business competitors with an advantage.

Shyue-Liang Wang et.al., proposed that the pattern inversion tree to store the added data set in one database scan. It is then sanitized and merged to the original sanitized database. Various characteristics of the proposed algorithm are analyzed. Numerical experiments and running time analyses show that the proposed approach out performs the direct sanitization approach on original and added data sets, with similar side effects.

Xin-She Yang et.al., proposed a new cuckoo search for multi objective optimization. It validate against a set of multiobjective test functions, and then apply it to solve structural design problems such as beam design and disc brake design. The main characteristics of the algorithm and their implications are also analysed.

Gayathiri et.al., proposed that the Association Rule Hiding technique in data mining performs the generation of sensitive association rules by the way of hiding based on the transactional data items. The property of hiding rules not the data makes the sensitive rule hiding process is a minimal side effects and higher data utility technique.

REFERENCES:

1) Stanley R. M. Oliveira and Osmar R. Za¨ıane.
A unified framework for protecting sensitive association rules in business collaboration.

2) Mohammad Naderi Dehkordi(2009). A Novel Method for Privacy Preserving in Association Rule Mining Based on Genetic Algorithms.

3) Azam Khan, Muhammad Shuaib Qureshi, Ayyaz Hussain(2014). Improved Genetic Algorithm Approach for Sensitive Association Rules Hiding.

4) A. Akbarzadeh1, E. Shadkam2(2015). The study of cuckoo optimization Algorithm for production planning problem.

5) Tolga Ayav and Belgin Ergenc. Full-Exact Approach for Frequent Itemset Hiding.

6) Dhyanendra Jain ,Amit sinhal,Neetesh Gupta, Priusha Narwariya,Deepika Saraswat,Amit Pandey(2012). Hiding Sensitive Association Rules without Altering the Support of Sensitive Item(s).

7)Chun-WeiLin, Tzung-PeiHong and Hung-ChuanHsu3(2014). Reducing Side Effects of Hiding Sensitive Itemsets in Privacy Preserving Data Mining.

8) Ramin Rajabioun(2011). Cuckoo Optimization Algorithm.

9) Xin-She Yang and suash Deb(2009). Cuckoo Search via L´evy Flights.

10) Aris Gkoulalas-Divanis · Vassilios S. Verykios(2008). Hiding sensitive knowledge without side effects.

11) Xingzhi Sun and Philip S. Yu(2007). Hiding Sensitive Frequent Itemsets by a Border-Based Approach.

12) Aris Gkoulalas-Divanis and Vassilios S. Verykios(2006). An Integer Programming Approach for Frequent Itemset Hiding.

13) Elena Dasseni, Vassilios S. Verkios and Ahmed K. Elmagarmid(2000). Hiding Association Rules by Using Confidence and Support.

14) Shyue-Liang Wang. Maintenance of Sanitizing Informative Association Rules.

15)Xin-She Yang and suash Deb(2011). Multiobjective Cuckoo search for design optimization.

16) Gayathiri P and Dr. B Poorna(2015). Association Rule Hiding Techniques for Privacy Preserving Data Mining: A Study.