IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol. 9, Issue.2, February 2021 E-ISSN: 2347-2693

Fishers Filtered Gravitational Rule Selection and Weighted Rank Based Genetic Algorithm for Association Rule Hiding to Preserve Privacy in Transactional Database

Gayathiri P. [1*], B. Poorna[2]

[1]Department of Computer Applications Rajiv Gandhi Arts and Science College, Puducherry, India
[2]A.M. Jain College, Chennai, TamilNadu, India

[*]*Corresponding Author: gayathirimourali@gmail.com Tel.:+91-9385972065*

DOI: https://doi.org/10.26438/ijcse/v9i2.2938 | Available online at: www.ijcseonline.org

Received: 31/Jan/2021, Accepted: 04/Feb/2021, Published: 28/Feb/2021

Abstract — Several methods had been investigated in the literature for rule hiding involving sensitive items. Some methods use co-operative models for mining functional association rules and some use distortion-based rule hiding technique. The present paper focuses on fast mining of rules using rank-based sensitive rule hiding framework called, Fisher's Filtered Gravitational Search and Rank-based Gene (FFGS-RG) for hiding sensitive association rules. To start with Fisher's Filtered is applied to filter the association rule and speeding up the mining process among the generated rule with Gravitational Search technique to select the sensitive rules from the transactional database. Once the sensitive rules are selected, the gene property of hidden and exposed items is mapped to the vector data item of sensitive rules for minimum distortion based on weighted ranking. The new gene data item population is generated using genetic algorithm operations to minimize the distortion via ranking. With distorted minimized offspring gene data item population, new sensitive rules are generated using Fisher's test that speeds up the rule selection process and provided to the transactional users. The distorted minimized offspring generated new rules are obtained then tested for side effects. This process is continued till the final sensitive rule hiding has minimal distortion on the gene populated data item rules and higher data item utility to the transactional users using weighted rank. A benchmark dataset is used to evaluate the FFGS-RG framework and the results show more efficient in improving the rule hiding accuracy with minimal rule selection time and also optimizing the sensitive rules hiding process.

Keywords – Gravitational Search, Gene Pattern, Rule Hiding, Sensitive rule, Fisher's test.

I. INTRODUCTION

Different data mining techniques can be pre-owned to identify useful knowledge from huge collections of data. In specific, the identification of association rules from huge database furnishes valuable information, such as customer buying habits or behaviours in departmental stores or fraud detection in ATM centres and so on. The extraction of data mining results in the improvement of accuracy. Though accuracy was said to be ensured, results in the disclosure of sensitive knowledge to third parties. Therefore, a tradeoff should be maintained between legitimate mining needs and the securitization of confidential information when data is held or shared.

A heuristic for Confidence and Support Reduction based on Intersection Lattice (HCSRIL) algorithm [1] was designed to hide the group of sensitive association rules in a given transactional database. To reduce the side effects, the HCSRIL algorithm observes the victim item and minimum transactions such that the modification of this item provides lower impact on itemsets in generating a set of frequent itemsets. Besides, HCSRIL algorithm was used in the risk avoidance of a retailer when the retailer's data are distributed in retail supply chain collaboration. However, rule hiding process was not effectively carried out due to the side effects in the modified association data rules.

A new distortion-based rule-hiding method was investigated in [2] that hide sensitive rules by removing certain items in a database to minimize the support or confidence of sensitive rules below certain threshold being specified for data sharing. To minimize the side effects on knowledge, the non-sensitive itemsets present in each transaction was used to sort the supporting transactions following they are relevant with non-sensitive itemsets. To reduce the distortion degree on data, a minimum number of transactions that need to be modified to conceal a sensitive rule was derived. However, distortion-based rule hiding with minimum side effects or minimum distortion to reduce the computational complexity involved in hiding sensitive rules has to be investigated.

To overcome the above limitations, we developed Fisher's Filtered Gravitational Search and Rank-based Gene (FFGS-RG), a novel framework for association rule hiding to preserve privacy on the transactional database. FFGS-RG framework is based on the distortion-based rule hiding method [2]. Association rules are extracted in transaction databases by identifying and mining frequent item sets, thus we formulated the problem of finding a set of candidate business variants correlated to the business classes, as the finding of Frequent Sets of candidate business variants and then as the extraction of association rules from those frequent sets.

The rest of the paper is structured as follows. In Section 2 some background knowledge related to association rule mining is presented. The proposed sensitive rule hiding framework is described in Section 3. Section 4 gives the experiment results with relevant discussions provided in Section 5. We conclude this paper in Section 5.

II. RELATED WORKS

Due to the recent development in technology, dealing with and extracting essential information is considered to be the most challenging tasks. In [4], Fuzzy Frequent Pattern Ubiquitous Streams was investigated to mine fuzzy association rules. However, the frequency of the extracted rules remained unaddressed.

An associative classification (AC) algorithm was designed in [5] that discovered association rule from single label data and therefore minimizing the number of extracted rules. An integrated Classification and Association Rule mining algorithm for discrimination of pandemic and non-pandemic were analyzed in [6] to reduce the time taken to extract rules. On the other hand, closed item set mining was performed in [7] to speed up the mining process.

With the recent trend oriented towards optimality, existing rule acquisition theory should be enriched. In [8], the relationship between three types of rules was analyzed to improve the compactness of association rules. Another hybrid, artificial immune network and particle swarm optimization for supplier selection with the least cost was presented in [9]. A selective analysis of microarray data with the aid of association rule mining was performed in [10]. Another optimization method using Cuckoo for association rule hiding was presented in [11] to protect private information and avoid information leakage.

With the increasing demand in big data analysis, Map Reduced-based Consistent and Inconsistent rule detection algorithm was designed in [12]. This algorithm was found to be flexing and efficient for mining a huge amount of data. To perform secure mining, Fast Distributed Mining (FDM) algorithm was investigated in [13] and was proved to be efficient in terms of communication and computational cost.

Two-view datasets are datasets providing two alternatives, each providing a different view. A score based on Minimum Description Length (MDL) was presented in [14] with the aid of heuristics that resulted in a better trade-off between run time and compression. An effective association rule hiding using gene patterned was presented in [21]. Discovering fuzzy exception and anomalous rules was performed in [15] with the aid of Boolean algebra that in turn resulted in the improvement of comprehensible results.

An Evolutionary Multi-Objective (EMO) method was designed in [16] to minimize the side effects with fewer damages to the non-sensitive rules. Another data mining approach to categorize the road accident locations was presented in [17] with the aid of the k-means algorithm. Optimizing rule hiding using heuristic approaches were presented in [18] and [20]. A novel feature subset algorithm based on association rule mining was presented in [19] with the aid of Partial Least Square Regression (PLSR) based threshold prediction method resulting in the improvement of classification accuracy. However, rule hiding was not concentrated. An efficient and effective in mining high utility sequential pattern with time intervals was presented in [22].

In summary, FFGS-RG framework extracts relevant knowledge by efficiently computing frequent item sets, as well as hiding the sensitive association rules that link business's class variants in more than one probe with the conditions of individuals or organizations. The main contributions of the presented work are:

- The identification of frequent itemset by Fisher's filter test and the generation of sensitive rule via Gravitational Search.
- An efficient implementation of the association rule mining technique that uses optimized search space adapted to taxi service trajectory dataset as opposed to general-purpose pattern representation.
- Performing sensitive rule hiding through weighted rank gene pattern-based association rule hiding and can hide the sensitive rules with minimum distortions (i.e., lesser number of non-sensitive rules hidden, lesser number of new rules generated and all the sensitive rules can hide in the sanitize database).

III. FISHER'S FILTERED GRAVITATIONAL SEARCH AND RANK-BASED GENE FRAMEWORK

This section describes the core algorithm of Fisher's Filtered Gravitational Search and Rank-based Gene (FFGS-RG) framework and its optimizations used to minimize the space search i.e. used to optimize the number of frequent itemsets examined to extract sensitive rules. The framework starts with the problem formulation followed by which basic notations and definitions are provided and finally the framework explained with.

A. Problem formulation

Let a transaction database 'TD', a minimum support threshold 'S_{min}', and a minimum confidence threshold 'δ' be given. Let us further presume that 'R' represents a set of association rules mined from 'TD', whose support and confidence are not less than 'S_{min}'and δ', respectively. Suppose that a set of certain association rules in 'R'regarded as being sensitive, denoted by 'R_{sen}'. The problem is now to transform 'TD' into an exposed database

'TD'' in such a way that all sensitive association rules in 'R_{sen}' are hidden. On the other hand, non-sensitive association rules to be further mined from 'TD' with minimum distortion.

B. Basic notations and definitions

Let '$I = \{i_1, i_2, \dots, i_n\}$' denote set of '$n$' literals. Each member of 'I' is represented as an item and 'P' is an itemset if '$P \subseteq I$'. On the other hand, a transaction 'T' is defined by a set of items, comprising of '$T = \{i_m \mid i_m \subseteq I\}$'. Let us also consider a precise transaction database, namely, '$TD = \{T_1, T_2, \dots, T_n\}$'. An itemset '$P \subseteq$ I' is supported by a transaction '$T \subseteq TD$' if '$P \subseteq TD$'. The frequency of an itemset 'P' in a database is '$\sup of\ P$', represented by '$S\ (P) = |P\ (T)|$', and is mathematically formulated as given below.

$$S\ (P) = |P\ (T)|, where\ P(T) = T \subseteq TD \quad (1)$$

An itemset 'P' is called a frequent itemset if '$S\ (P) \geq S_{min}$', where 'S_{min}' denotes the minimum pre-defined support threshold. The support of a rule '$P \rightarrow Q$' is defined to be the support of itemset '$P \cup Q$' and is as given below.

$$S\ (P \rightarrow Q) = S\ (P \cup Q) \quad (2)$$

The confidence of a rule '$P \rightarrow Q$'is defined as given below.

$$C\ (P \rightarrow Q) = \frac{S\ (P \cup Q)}{S(P)} \quad (3)$$

The block diagram representation of FFGS-RG framework in which hiding sensitive rule is presented in Figure 1. The steps involved in the proposed framework are described as follows. Initially, sample dataset (i.e. Taxi Service Trajectory and Abalone) is considered as an input of the proposed framework. Here, the Taxi Service Trajectory and Abalone dataset is represented in the form of vector format. Followed by which, conversion of Taxi Service Trajectory and Abalone dataset into transactional database takes place. Then, the transactional database is read and passed through Fisher's filter. With the resultant filtered value, sensitive rules are mined based on the Gravitational Search.

Once the sensitive rules are generated from Fisher's Filtered test, sensitive rules have to be mined. Followed by which the database is dynamically updating for every time interval. To handle dynamic nature, sensitive rules are updated but the full scan with the whole database is not followed. Here, the confidence count and weight are measured. With these two measured values, sensitive rule hiding is performed in ascending order. Finally, the sensitive rules are updated (i.e. perform hiding) based on the ranking approach.

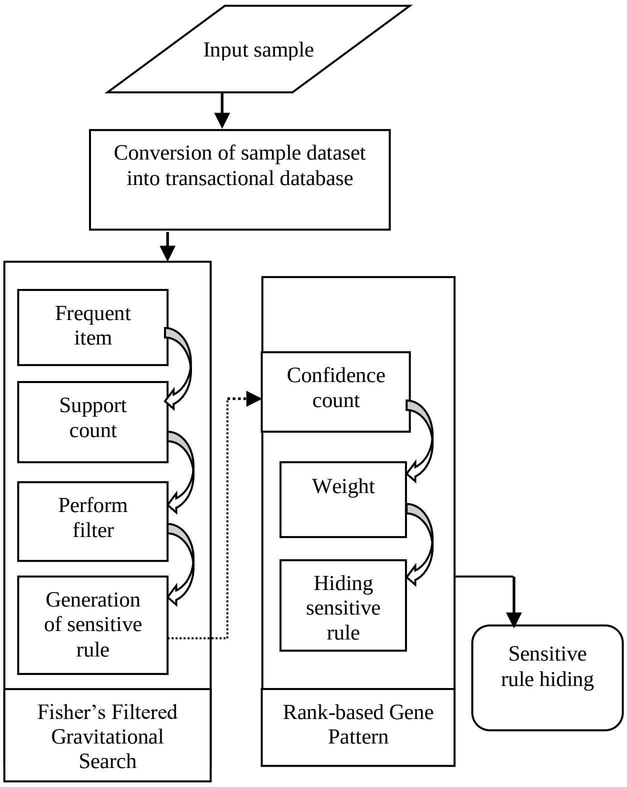

Figure 1. Block diagram of sensitive rule hiding framework

C. Fisher's Filtered Gravitational Search for sensitive rule selection

In this section, Fisher's Filtered Gravitation Search (CGS) technique for selection of sensitive rules is investigated. The CGS technique uses a Fisher's exact test along with the Gravitational Search to discover the association between itemsets and conditional probability. The application of Fisher's exact test along with the GS model removes the potential ambiguous association rules. Figure 2 shows the block diagram of the CGS technique.

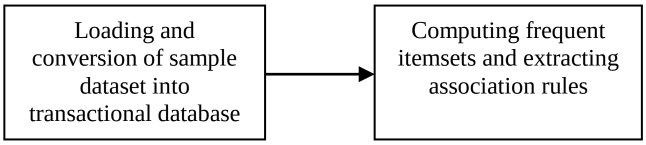

Figure 2. Block diagram of CGS technique

As shown in Figure 2, the input data table produced by Taxi Service Trajectory and Abalone dataset is initially stacked and rearranged obtaining a '$n * m$' matrix named 'T', where 'n' is the number of samples and 'm' is the number of iterations. In this way, each row of the 'T' represents a transaction, where all number of association rules generated using different datasets, on the various iterations, are the items of the transaction.

With 'T' representing the transaction, a transaction over 'I' is represented by a pair '$T = (t\_ID, I)$', with 't_ID' representing the transaction identifier and 'I' denoting the item. The robustness of frequent item sets extraction depends on the probability to identify correlations hidden in the data sets with reduced search space. This feature is based on Support count and is obtained using Fisher's exact test using table 1.

Before we proceed with the Fisher test, certain notations are introduced. The cells are represented by letters 'p', 'q', 'r' and 's'.The totals across rows are denoted by '$p+q$' and '$r+s$', whereas the columns marginal totals are denoted by '$p+r$' and '$q+s$' respectively and the total by 'n'. So, the table with certain attributes in Taxi Service Trajectory dataset like Origin_Call, Origin_Stand, Taxi_ID, Trip_ID is as given in table1.

Table 1. Taxi Service Trajectory Dataset Representation

	Taxi_ID	Trip_ID	Row total
Origin_Call	p	q	$p+q$
Origin_Stand	r	s	$r+s$
Column total	$p+r$	$q+s$	$p+q+r+s=(n)$

Now the support count '$S_c(.)$' for an item 'P' with the probability of obtaining any such set of values is as given below.

$$Prob\ (P) = S_c(P) = \frac{\binom{p+q}{p}\binom{r+s}{r}}{\binom{n}{p+r}} \tag{4}$$

$$= \frac{(p+q)!(r+s)!\,(p+r)!(q+s)!}{p\,!q\,!r\,!s\,!n\,!} \tag{5}$$

From (4) and (5), '$\binom{n}{k}$' denotes the binomial coefficient with '!' representing the factorial operation. Another aspect to be taken into account while designing gravitational search algorithm based on the law of gravity and mass interactions along with Fisher's exact test is to improve the execution time with which the sensitive rules are selected. To have fast access, a two-step scanning is included in the design of Gravitational Search Fisher's Exact Sensitive Rule Generation algorithm 1.

Input: Transaction Database 'TD', Transaction 'T', Itemset '$I = \{i_1, i_2, \ldots, i_n\}$', Minimum pre-defined support threshold 'S_{min}'
Output: Computational-efficient Sensitive Rule Generation
1: Begin
2: For each Transaction Database 'TD' and Transaction 'T'
3: For each Itemset 'I' and Minimum pre-defined support threshold 'S_{min}'
4: Measure support of itemset using eq. (2)
5: Measure confidence value using eq. (3)
6: Measure support count using eq. (4) and eq. (5)
7: If '$S_c(P) \geq S_{min}$'
8: Identify recurrence of each item
9: Fill frequent items list with all items and chosen as sensitive rule 'R_s' (i.e. in the form of vector element)
10: End if
11: If '$S_c(P) < S_{min}$'
12: Perform deletion of all itemsets and chosen as non-sensitive rule 'R_{ns}'
13: End if
14: End for
15: End for
16: End

Algorithm 1 – Gravitational Search Fisher's Exact Sensitive Rule Generation

The first pass is required to identify the recurrence of each item into the transaction database 'TD', to fill the frequent items list with all items for which '$S_c(P) \geq S_{min}$'. Then, the items are arranged based on their decreasing recurrence and it is chosen as the sensitive rule.

The second pass is related to the deletion of all itemsets for which their support is '$S_c(P) < S_{min}$'. The remaining itemsets are organized following the decreasing recurrence in the transactions and it is chosen as the non-sensitive rule. In this way, the mining process speeds up with reduced computational cost as the search space minimizes for a high value of minimum support.

D. Weighted Rank-based Gene Pattern Rule Hiding

Once, the sensitive rules are selected, the next step is to hide the selected sensitive rules. In this section, a Weighted Rank-based Gene (WRG) rule hiding technique is designed. In this section, we specifically introduce the Rank-based rule hiding using distortion technique applied in association rule hiding that was presented in [2].

The proposed WRG technique evaluates the database for associative rules to be hidden and to be exposed from a constructed vector element. Firstly, the characteristics of the distortion technique of frequent itemsets are analyzed. Then, the heuristics for minimizing the distortion of the association rule hiding process is improved. Finally, an efficient algorithm for hiding a specific set of sensitive rules generated from Fisher's Filtered Gravitational Search Exact test is designed.

With 'T' representing the set of transactions in a transactional database 'TD' where 'I_T', represents the number of items of transaction 'T', then, the left itemset and right itemset are denoted as 'I_L' and 'I_R' respectively. The transactional database items are represented as vector elements. With the aid of the Gene Selection performed over the vector elements, the distortion technique identifies how many transactions this rule has greater than 'GC_{min}' minimum pre-defined confidence threshold 'C_{min}'. It is mathematically formulated as given below.

$$GC_{min}(R) = \frac{[(R) - (C_{min})]}{I_L} \tag{6}$$

Similarly, the distortion technique identifies how many transactions this rule has lesser than 'LC_{min}' minimum pre-defined confidence threshold 'C_{min}'. It is obtained as given below.

$$LC_{min}(R) = \frac{[(C_{min}) - (R)]}{I_L} \tag{7}$$

In this study, we further assume that the user minimizes the confidence of a sensitive rule by a Secured Boundary threshold 'SB' below the minimum pre-defined confidence threshold 'C_{min}'.Finally, let '$|T_R|_{prior}$' represents the number of transactions that support 'R' before conducting the hiding process '$prior$' and '$|T_R|_{posterior}$' represents the number of transactions that support 'R' after the hiding process '$posterior$'. Then, confidence count for 'n' number of transactions is as given below.

$$C(R_i) = \frac{|T_{Ri}| - n}{|T_{IL}|} \leq (C_{min} - SB) \quad (8)$$

$$n = [|T_{Ri}| - (C_{min} - SB) * |T_{IL}|] \quad (9)$$

From (8) and (9), the confidence count 'C' for 'n' number of transactions that supports rule 'R_i' is obtained by applying the Secured Boundary threshold 'SB' and left itemset 'T_{IL}' respectively. Then, the 'n' number of transactions are applied to the Gene pattern Min-Max Fitness function. This is mathematically evaluated as given below.

$$\text{F} = \sum \text{Min}\,(R_s \cap R_{ns}) + \text{Max}\,(R_s \cap R_{ns}) \quad (10)$$

With the obtained fitness function 'F', the weight '$Weight$' for each transaction 'T' is mathematically evaluated as given below.

$$eight(T) = \frac{|R_S| * (SB)}{|R_{ns}|} \quad (11)$$

The pseudo-code representation of sensitive rule hiding using Weighted Rank-based Gene is as given below
Input: Transaction 'T', Transactional Database 'TD', left itemset 'I_L', right itemset 'I_R', Itemset '$I = \{i_1, i_2, \ldots, i_n\}$', Secured Boundary threshold 'SB', sensitive rule 'R_s', non-sensitive rule 'R_{ns}'

Output: Minimum-distorted Sensitive Rule Hiding
1. Begin
2. For each Transaction 'T' with Transactional Database 'TD'
3. For each Itemset 'I'
4. Measure confidence count using eq. (8) and eq. (9)
5. Evaluate Gene pattern Min-Max Fitness function using eq. (10)
6. End for
7. End for
8. End

Algorithm 2 – Weighted Rank-based Gene

As illustrated in the algorithm 2, the Weighted Rank-based Gene (WRG) algorithm, initially measures the confidence count for each transaction. Then, the WRG algorithm executes three sub-steps for each sensitive association rule 'R'. With the generated sensitive and non-sensitive rules, the algorithm computes the fitness function using eq (10).
Next, the weight for each transaction is evaluated. Finally, the algorithm sorts the transactions in ascending order according to the weight of each transaction. In this way, the transactions with the lowest weight '$Weight$' will come on top, to hide the sensitive rule, but the modification causes the least impact on 'R_{ns}'.

IV. EXPERIMENTAL SETTINGS

Experiments are conducted to show the performance of the proposed framework FFGS-RG which was run on Windows 7 operating system with Pentium Core i5 and 4 GB of RAM. Heuristic for Confidence and Support Reduction based on Intersection Lattice (HCSRIL) algorithm [1] and Distortion-based Rule Hiding (DRH) [2] method is also designed as a benchmark to be compared with the proposed algorithm.

Two datasets, Abalone and Taxi Service Trajectory are used to evaluate the performance of the proposed framework FFGS-RG in terms of sensitive rule selection time, sensitive rule hiding time, several sensitive rules generated for the number of transactions and number of modified associative rules. The experiments were executed using the Taxi Service Trajectory and Abalone dataset. The FFGS-RG framework is implemented in MATLAB.

The Taxi Service Trajectory dataset contains full reviews of trajectory data of 3000 taxis on holiday and weekend in Wuhan city in China using MATLAB platform that helps in analyzing the distribution and dynamics of the hot spots concerning holiday, weekday and weekend. The University of California, Irvine Machine Learning Repository [3] provides a data set consisting of 4177 samples of physical characteristics of abalones and their age.

V. DISCUSSION

The proposed work is compared against the existing Heuristic for Confidence and Support Reduction based on Intersection Lattice (HCSRIL) algorithm [1] and Distortion-based Rule Hiding method (DRH) [2]. The experiment is conducted on factors such as sensitive rule selection time, sensitive rule hiding time, number of sensitive rules hidden, number of non-sensitive rules hidden and number of new rules generated concerning the number of modified associative rules.

A. Impact of Sensitive rule selection time

Sensitive rule selection time is one of the most important standard metrics used to measure the performance of privacy-preserving on transactional database. Sensitive rule selection time measures the time taken to select sensitive rules. It is measured in terms of Seconds (Sec) and is mathematically formulated as given below.

$$T_{SRS} = \sum_{i=1}^{n} R_{si} * Time\,(R_{si}) \quad (12)$$

From (12), the sensitive rule selection time 'T_{SRS}', is arrived at using the number of sensitive rules 'R_{si}' and the time to obtain the sensitive rules '$Time\,(R_{si})$' respectively. Lower the sensitive rule selection time, more efficient the method is said to be. Table 2 shows the sensitive rule selection time by the proposed FFGS-RG framework and comparison made with the existing methods HCSRIL [1] and DRH [2] on the Abalone and Taxi Service Trajectory dataset.

Table 2. Comparative Results of Sensitive Rule Selection Time

Number of modified associative rules	Sensitive Rule Selection time (Sec) – using Taxi Service Trajectory dataset		
	FFGS-RG	HCSRIL	DRH
100	212	250	265
200	242	265	287
300	317	380	462
400	365	440	477
500	432	473	557
600	540	589	627
700	615	652	714
800	689	702	739
900	735	755	785
1000	814	834	864

To assess the performance of FFGS-RG framework and compare it to other methods, namely, HCSRIL [1] and DRH [2] all three methods were implemented using MATLAB tool and tabulation are shown in table 1. The results on FFGS-RG framework are investigated with the small stage information which is obtained from experimental work. To support transient performance, in Table 2, a Gravitational Search Fisher's Exact Sensitive Rule Generation algorithm is applied and a comparison made with two other existing methods HCSRIL [1] and DRH [2].

Figure 3 show that the FFGS-RG framework provides minimum sensitive rule selection time when compared to HCSRIL [1] and DRH [2] using Taxi Service Trajectory dataset. The sensitive rule selection time is reduced with the application of Gravitational Search Fisher's Exact Sensitive Rule Generation algorithm.

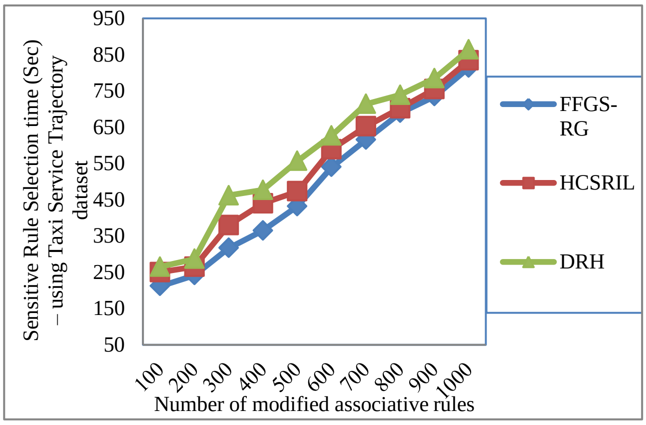

Figure 3. Sensitive rule selection time using taxi service trajectory dataset

Figure 3 it is obvious to see that the Fisher's exact test has the best performance in sensitive rule selection time since it obtains the sensitive rule only after filtering the testing dataset. The proposed framework greatly reduces the sensitive rule selection time compared to the state-of-the-art methods since for FFGS-RG framework it is unnecessary to but scan the entire database for evaluating the filter value at each iteration.

With the application of Gravitational Search Fisher's Exact Sensitive Rule Generation algorithm, for each test sample, the number of sensitive rules is obtained, according to the support count for each transaction. With the convergence of probability of minimum pre-defined support threshold and support count, recurrence of each item is measured and obtained. This convergence of recurrence at an early stage, in turn, reduces the sensitive rule selection time in an efficient manner using fisher's exact test. This in turn extracts the presence of sensitive rules in a transactional database for each transaction. This, in turn, helps in improving the sensitive rule generation and therefore the sensitive rule selection time for each transaction is reduced using FFGS-RG by 3% and 12% compared to HCSRIL and DRH when applied with Abalone dataset. Similarly, the sensitive rule selection time for each transaction is reduced using FFGS-RG by 9% and 16% compared to HCSRIL and DRH when applied with Taxi Service Trajectory dataset.

B. *Impact of Sensitive rule hiding time*

Experiments were also conducted to evaluate the impact of sensitive rule hiding time of the proposed framework and comparison made with two other methods. Sensitive rule hiding time measures the time taken to hide sensitive rules. It is measured in terms of Seconds (Sec) and is mathematically formulated as given below.

$$T_{SRH} = \sum_{i=1}^{n} |T_{Ri}|_{posterior} * Time\ \left||T_{Ri}|_{posterior}\right| \qquad (13)$$

From (13), the sensitive rule hiding time 'T_{SRH}' is obtained using the number of transactions that support 'R' after the hiding process '$posterior$'. Lower the sensitive rule hiding time, more efficient the method is said to be.

Table 3. Comparative results of sensitive rule hiding time

Number of modified associative rules	Sensitive rule hiding time (Sec) – using Taxi Service Trajectory dataset		
	FFGS-RG	HCSRIL	DRH
100	622	850	905
200	652	882	927
300	727	989	1088
400	775	1050	1114
500	842	1139	1197
600	950	1213	1267
700	1025	1280	1364
800	1099	1310	1379
900	1145	1358	1425
1000	1224	1439	1582

Table 3 shows the utility measure obtained by running the algorithm RGPRH to achieve sensitive rule hiding time on the Abalone and Taxi Service Trajectory dataset respectively. In table 3, the results are reported for different values of modified associative rules.

To better understand the effectiveness of the proposed FFGS-RG framework, extensive experimental results are reported in table 3. MATLAB tool is used to measure and experiment the factors by analyzing the percentage of result with the help of table and graph values. Results are presented for a different number of modified associative rules. The results reported here confirm that with the increase in the number of modified associative rules, the sensitive rule hiding time also increases.

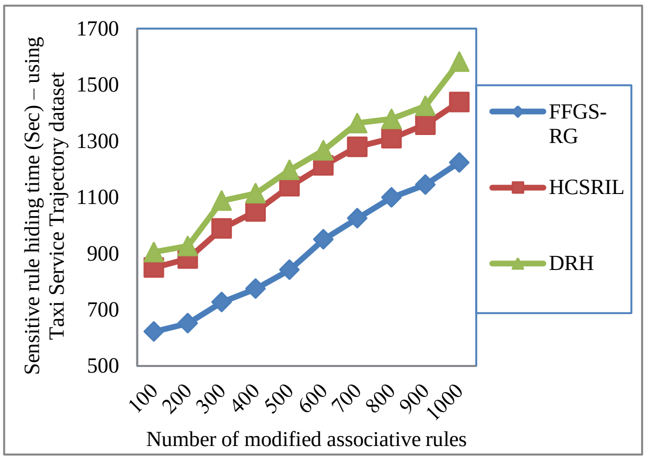

Figure 4. Sensitive rule hiding time using taxi service trajectory dataset

From Figure 4, it is obvious to see that when the number of modified associative rules were at 20 in taxi service trajectory dataset, the proposed Rank-based Gene Pattern Rule Hiding algorithm produced more transactions to be deleted for hiding sensitive rules. Since the proposed Rank-based Gene Pattern Rule Hiding algorithm considers three sub-steps for each sensitive association rule, the selected transactions for deletion may consist of fewer large transactions rather than many sensitive association rules.
The sensitive rule hiding time is reduced in the FFGS-RG framework as it uses the lowest weight for each generated sensitive rule with the aid of the fitness function. By applying the Min-Max fitness function in FFGS-RG framework, the algorithm sorts the transaction in the ascending order according to the weight. This results in the reduced sensitive rule hiding time while assessing the selected sensitive rules. This in turn reduces the sensitive rule hiding time by 22% compared to HCSRIL [1] and 27% compared to DRH [2].

C. Impact of the number of sensitive rules hidden

In FFGS-RG framework, the number of sensitive rules hidden is defined as the ratio of the number of hiding the sensitive rules without any side effects to the total number of modified associative rules given. It is also used to measure the rate of distortion, where the objective remains in reducing the distortion degree on data. The number of sensitive rules hidden is mathematically formulated as,

$$\text{No. of sensitive rules hidden} = \frac{\text{Hiding the sensitive rules without any side effects}}{\text{Total number of modified associative rules}} \tag{14}$$

From equation (14), the number of sensitive rules hidden is measured in terms of numbers. When the number of hidden rules is lesser, the method is said to be more efficient. Table 4 provides a comparative analysis of several sensitive rules hidden using Abalone and Taxi service trajectory dataset.

Table 4. Comparative results of the number of sensitive rules hidden

Number of modified associative rules	Number of sensitive rules hidden – using Taxi Service Trajectory dataset		
	FFGS-RG	HCSRIL	DRH
20	0	2	0
40	0	4	0
60	0	5	0
80	0	6	0
100	0	7	0

The targeting results of the number of sensitive rules hidden using proposed FFGS-RG framework with two state-of-the-art methods [1], [2] in table 4 presented for comparison based on the number of modified associative rules. MATLAB tool is used to calculate and experiment the factors through analyzing the percentage of result with the help of table and graph values. The number of sensitive rules hidden using proposed FFGS-RG framework offers comparable values than the state-of-the-art methods.

To explore the influence of the number of sensitive rules hidden on proposed FFGS-RG framework, the experiments were performed by varying the number of modified associative rules as shown in Figure 5. It also shows that the proposed FFGS-RG framework shows competitive results with state-of-the-art methods, namely HCSRIL [1] and DRH [2]. This is because of the application of Rank-based Gene Pattern Rule Hiding algorithm for hiding each sensitive association rule with lesser weight value according to the fitness function by using the proposed FFGS-RG framework.

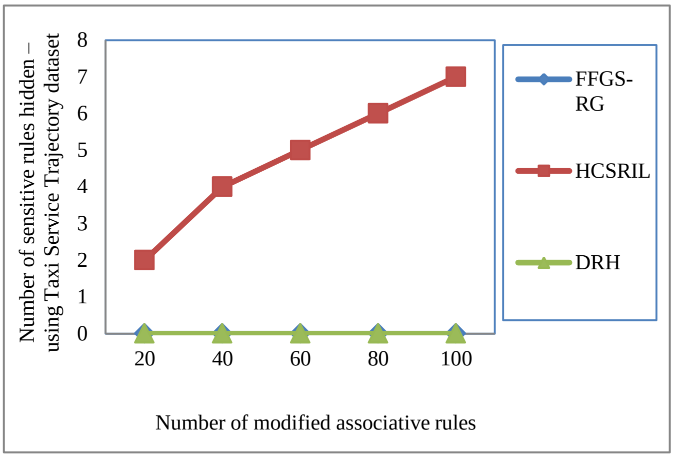

Figure 5. Number of sensitive rules hidden using taxi service trajectory dataset

Based on Min-Max fitness function, the algorithm sorts the transaction within the ascending order with respect to the weight for each generated sensitive rule during the transaction. In this manner, the transactions with the lowest weight will come on top and thus efficiently hide the sensitive rule. Hence, the proposed FFGS-RG framework reduces the number of sensitive rules hidden as compared with existing HCSRIL [1] and DRH [2] using taxi service trajectory dataset.

D. Impact of number of non-sensitive rules hidden

In FFGS-RG framework, the number of non-sensitive rules hidden is defined as the ratio of the number of non-sensitive rules hidden with minimum distortion to the total number of modified associative rules. The number of sensitive rules hidden is mathematically written as follows.

$$\text{No. of non} - \text{sensitive rules hidden} = \frac{\text{hidding the no.of non}-\text{sensitive rules with minimum distortion}}{\text{Total number of modified associative rules}} \quad (15)$$

From the equation (15), the number of non-sensitive rules is hidden according to the total number of modified associative rules given in the database. When the number of non-sensitive rules hidden is lesser, the method is said to be more efficient. Table 4 provides the comparative analysis of number of non-sensitive rules hidden using Abalone and Taxi service trajectory dataset.

Table 5. Comparative results of number of non-sensitive rules hidden

Number of modified associative rules	Number of non-sensitive rules hidden– using Taxi Service Trajectory dataset		
	FFGS-RG	HCSRIL	DRH
20	5	8	10
40	7	10	12
60	9	12	14
80	13	16	18
100	15	18	21

The evaluated the performance of the proposed FFGS-RG framework and is compared with the other methods such as HCSRIL [1] and DRH [2] all three methods were implemented using MATLAB tool and tabulation are shown in table 5. The results on the proposed FFGS-RG framework are investigated the number of non-sensitive rules hidden based on a number of modified associative rules which is obtained from experimental work. Higher the number of modified associative rules, the higher the number of non-sensitive rules hidden. However, the number of non-sensitive rules hidden using FFGS-RG framework offers lesser comparable values with two other existing methods HCSRIL [1] and DRH [2] in the above-given table.

This in turns mine the non-sensitive association rules from 'TD' with lower distortion efficiently. Therefore, the proposed FFGS-RG framework provides a reduced number of non-sensitive rules hidden by 28% compared to HCSRIL [1] and 39% compared to DRH [2] using an abalone dataset.

Figure 6 demonstrates the non-sensitive rules hidden on the proposed FFGS-RG framework with a varied number of modified associative rules taken as the input to perform privacy-preserving. It also shows that the FFGS-RG framework shows competitive results with state-of-the-art methods, namely HCSRIL [1] and DRH [2]. This is because of the application of Gravitational Search Fisher's Exact Sensitive Rule Generation algorithm using proposed FFGS-RG framework. In order to fill the frequent items list with all items, the algorithm is necessary to identify the recurrence of each item into the transaction database.

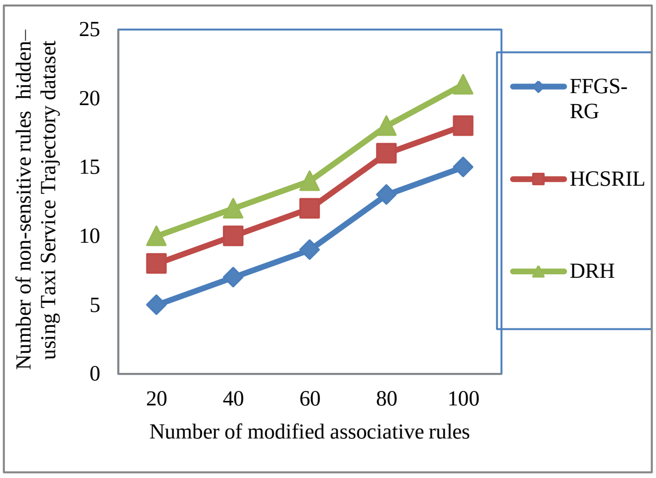

Figure 6. Number of non-sensitive rules hidden using taxi service trajectory dataset

Besides, the items are arranged according to their decreasing recurrence and it is chosen as the sensitive rule. After that, the deletion of all itemsets is performed for their support. The remaining itemsets are organized following the reducing recurrence during the transactions and it is selected as the non-sensitive rule. Like this, the mining process speeds up with reduced computational cost as the search space reduced for a high value of minimum support. This in turn efficiently hide the number of non-sensitive rules is decreased by 26% compared to HCSRIL and 37% compared to DRH using taxi service trajectory dataset.

E. Impact of the number of new rules generated

The number of new rules generated is defined as the rate at which the number of generating the new rules with minimum distortion by applying the proposed FFGS-RG framework. The number of new rules generated is mathematically formulated as given below.

$$\text{No. of new rules generated} = \frac{\text{generate the new rules with minimum distortion}}{\text{Total number of modified associative rules}} \quad (16)$$

From equation (16), the number of new rules generated according to the total number of modified associative rules given in the database. When the number of new rules generated is lesser, the method is said to be more efficient. Table 6 provides a comparative analysis of the number of new rules generated by applying the Abalone and Taxi service trajectory dataset.

Table 6 shows the comparative results of the number of new rules generated using the proposed FFGS-RG framework. The proposed framework is compared with two state-of-the-art methods [1], [2] in table 6 presented based on the number of modified associative rules. While improving the number of modified associative rules, the number of new rules generated is improved. However, from the given table above, the number of new rules generated is seen to be lower comparatively than the HCSRIL [1] and DRH [2] respectively.

Table 6. Comparative results of the number of new rules generated

Number of modified associative rules	Number of new rules generated – using Taxi Service Trajectory dataset		
	FFGS-RG	HCSRIL	DRH
20	5	7	8
40	7	9	10
60	10	12	13
80	12	14	16
100	14	17	18

Experiments are then conducted to show the number of new rules generated for the three methods for various numbers of modified associative rules.

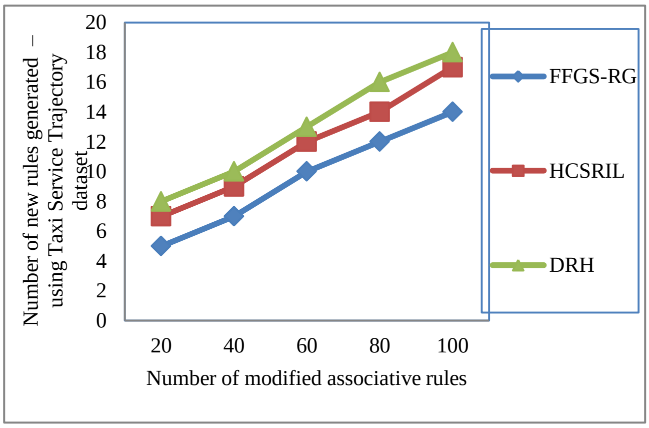

Figure 7. Number of new rules generated using taxi service trajectory dataset

To explore the influence of the number of new rules generated on proposed FFGS-RG framework, the experiments were performed by varying the number of modified associative rules as shown in Figure 7. It also demonstrates that the proposed FFGS-RG framework shows competitive results with state-of-the-art methods, namely HCSRIL [1] and DRH [2]. This is due to the application of weighted rank gene pattern-based association rule generates the new gene data item population once the sensitive rules are selected through the genetic algorithm to reduces the distortion by weighted ranking. Based on distorted minimized offspring gene data item population, new sensitive rules are generated by using Fisher's test which speeds up the rule selection process and presented to the transactional users. The distorted minimized offspring generated new rules are achieved, and then tested for side effects. This in turn minimizes the number of new rules generated by 19% compared to HCSRIL and 28% compared to DRH using taxi service trajectory dataset.

VI. CONCLUSION

In this paper, a compact Fisher's Filtered Gravitational Search and Rank-based Gene (FFGS-RG) framework is thus proposed to hide the sensitive rules through Gravitational Search and Gene-based pattern hiding. A flexible fitness function with a minimum pre-defined confidence threshold is also designed to consider the general distortion effects of hiding failure, transactions that support sensitive rule before hiding process and posterior to the hiding process. A secured boundary threshold is adopted in the proposed algorithm to speed up the mining process and therefore to reduce the cost involved in the computation. Experiments are conducted to show that the proposed FFGS-RG framework outperforms better than the state-of-the-art-works algorithms considering all criteria of distortion effects. In the experiments, two benchmark datasets were used to respectively evaluate the performances of the two proposed algorithms. Experimental results showed that the proposed FFGS-RG framework outperforms the aggregate algorithm in terms of sensitive rule selection time, sensitive rule hiding time, the number of sensitive rules hidden, the number of non-sensitive rules hidden and the number of new rules generated for the number of modified associative rules.

REFERENCES

[1] Hai Quoc Le, Somjit Arch-int, Huy Xuan Nguyen, Ngamnij Arch-int, "Association rule hiding in risk management for retail supply chain collaboration", Computers in Industry, Elsevier, **September 2013.**

[2] Peng Cheng, John F. Roddick, Shu-Chuan Chu, Chun-Wei Lin, "Privacy preservation through a greedy, distortion-based rule-hiding method", Applied Intelligence, Springer, **May** 2015 (relevance sorting approach) (Distortion-based Rule Hiding method).

[3] UCI Machine Learning Repository: Abalone Data Set. http://archive.ics.uci.edu/ml/datasets/Abalone.

[4] AmalMoustafa, BadrAbuelnasr, Mohamed Said Abougabal, "Efficient mining fuzzy association rules from ubiquitous data streams", Alexandria Engineering Journal, Elsevier, Apr 2015.

[5] NedaAbdelhamid, "Multi-label rules for phishingClassification", Applied Computing and Informatics, Saudi Computer Society, King Saud University, Elsevier, **Jul 2014.**

[6] FatemehKargarfard, Ashkan Sami, EsmaeilEbrahimie, "Knowledge discovery and sequence-based prediction of pandemicinfluenza using an integrated classification and association rule mining (CBA) algorithm", Journal of Biomedical Informatics, Elsevier, **Jul 2015.**

[7] H. Vathsala, and Shashidhar G. Koolagudi, "Closed Item-Set Mining for Prediction of Indian Summer MonsoonRainfall a Data Mining Model with Land and Ocean Variablesas Predictors", Eleventh International Multi-Conference on Information Processing-2015 (IMCIP-2015), Elsevier, **Jul 2015.**

[8] JunpingXie, Minhua Yang, Jinhai Li, ZhongZheng, "Rule acquisition and optimal scale selection in multi-scale formal decision contexts and their applications to smart city", Future Generation Computer Systems, Elsevier, **Mar 2017.**

[9] R.J. Kuo, C.M. Pai, R.H. Lin, H.C. Chu, "The integration of association rule mining and artificial immune network for supplier selection and order quantity allocation", Applied Mathematics and Computation, Elsevier, **June 2015.**

[10] Alagukumar, S, Lawrance. R, "A Selective Analysis of Microarray Data using Association Rule Mining", Graph Algorithms, High Performance Implementations and Applications, Elsevier, **Aug 2015.**

[11] MahtabHosseinAfshari, Mohammad NaderiDehkordi, Mehdi Akbari, "Association Rule Hiding using Cuckoo Optimization Algorithm", Expert Systems With Applications, Elsevier, **Aug 2016.**

[12] Dinesh J. Prajapati, Sanjay Garg, N.C. Chauhan, "Interesting association rule mining with consistent and inconsistent rule detection from big sales data in distributed environment", Future Computing and Informatics Journal, Elsevier, **May 2017.**

[13] TamirTassa, "Secure Mining of Association Rules in Horizontally Distributed Databases", IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, **VOL. 26, NO. 4, APRIL 2014.**

[14] Matthijs van Leeuwen and Esther Galbrun, "Association Discovery in Two-View Data", IEEE Transactions on Knowledge and Data Engineering, **Volume: 27, Issue: 12, Dec. 1 2015.**

[15] M. Dolores Ruiz, Daniel Sanchez, Miguel Delgado1, "Discovering Fuzzy Exception and Anomalous Rules", IEEE Transactions on Fuzzy Systems, **volume: 24, Issue: 4, Aug. 1 2016.**

[16] Peng Cheng, Chun-Wei Lin, Jeng-Shyang Pan, "Use HypE to Hide Association Rules by Adding Items", PLOS ONE | DOI: 10.1371/journal.pone.0127834 **June 12, 2015.**

[17] Sachin Kumar, DurgaToshniwal, "A data mining approach to characterize road accident locations", Journal of Modern Transportation, Springer, **May 2016.**

[18] Akbar Telikani, AsadollahShahbahrami, "Optimizing association rule hiding using combination of border and heuristic approaches, Applied Intelligence, **Apr 2017.**

[19] Guangtao Wang and Qinbao Song, "A novel feature subset selection algorithm based on association rule mining", Intelligent Data Analysis, ACM, **Volume 17 Issue 5, September 2013.**

[20] Gayathiri P and B Poorna," Gravitational Search Algorithm for Effective Selection of Sensitive Association Rules", Journal of Theoretical and Applied Information Technology, **Vol.96. No 10, 31st May 2018.**

[21] Gayathiri P and B Poorna," Effective Gene Patterned Association Rule Hiding Algorithm for Privacy Preserving Data Mining on Transactional Database", CYBERNETICS AND INFORMATION TECHNOLOGIES, BULGARIAN ACADEMY OF SCIENCES, **Volume 17, No 3, 2017.**

[22] Tran Huy Duong, Demetrovics Janos and Vu Duc Thi," An Algorithm for Mining High Utility Sequential Patterns with Time Interval" CYBERNETICS AND INFORMATION TECHNOLOGIES, BULGARIAN ACADEMY OF SCIENCES, **Volume 19, No 4, 2019.**

AUTHORS PROFILE

Dr Gayathiri P. pursed Ph.D. Degree from Bharathiar Univercity, Coimbatore, India and currently working as Assistant Professor in Department of Computer Sciences, Rajiv Gandhi Arts and Science College, Pondicherry University , India. She has published research papers in reputed international journals including Thomson Reuters (SCI & Web of Science) and Scoupes and it's also available online. Her main research work focuses on Data Security and Privacy, Big Data Analytics, Data Mining, She has 20 years of teaching experience and 8 years of Research Experience.

Dr B. Poorna. pursed Ph.D. Degree from Madras Univercity, India and currently working in A.M. Jain College, Chennai, India. She has published more than 40 research papers in reputed international journals including Thomson Reuters (SCI & Web of Science) and Scoupes and it's also available online. Her main research work focuses on Data Security and Privacy, Big Data Analytics, Data Mining, She has 20 years of teaching experience and 8 years of Research Experience.

