International Journal of Innovation and Scientific Research
ISSN 2351-8014 Vol. 24 No. 2 Jun. 2016, pp. 388-396
© 2015 Innovative Space of Scientific Research Journals
http://www.ijisr.issr-journals.org/

# Differential Evolution Algorithm for Hiding Fuzzy Association Rules Using Mutual Information

***K. Sathiyapriya*[1] *and G. Sudha Sadasivam*[2]**

[1]Research Scholar, Department of Computer Science and Engineering,
PSG College of Technology,
Coimbatore, Tamilnadu, India

[2]Professor, Department of Computer Science and Engineering,
PSG College of Technology,
Coimbatore, Tamilnadu, India

**ABSTRACT:** Data mining is the process of extracting the useful information from the large amount of available data. Association rule mining is a popular tool for discovering useful associations from large amount of data. Once private data is released for mining, it is very difficult to prevent its misuse. Useful associations with hidden information or knowledge that are sensitive to the database owner could be easily exposed using this kind of tool. Therefore it is necessary to hide all the sensitive information that can be mined from the data in the form of association rules before releasing the data. Most of the methods proposed in literature for association rule hiding deals with binary database and few methods for quantitative database suffer with side effects. This paper proposes an approach for hiding sensitive association rules using differential evolution and mutual information. The proposed algorithm hides the rule by decreasing Association Measure of the rule below threshold. Side effects are reduced by choosing the items with higher mutual information. Experimental results on real datasets demonstrat*e* that the proposed method can effectively sanitize the sensitive data with fewer side effects to the non-sensitive data.

**KEYWORDS:** Sensitive rules, Mutual Information, lost rules, Ghost ules, Fuzzy, Rule hiding, Differential evolution.

## 1 INTRODUCTION

Governments and production organizations use data for making decisions that provide social benefits. The key value of large databases today is research. It can be scientific, economic or market oriented. For example, research in medical field helps in finding new medicines for chronic disease; even competing businesses can share data with mutual interests for extracting hidden useful information. The main problem that arises in sharing of large amount of data is confidentiality violation.

If one organisation draws the sensitive information in the form of association rules, of its competitor then it is easy for that firm to get better of it. So each firm before releasing its own data tends to hide all the sensitive information in it. For example: patterns contained in the shared data may contain information about shelf space allocation, catalogue design, discount sales etc. Thus the need for privacy is motivated by business interests and may be due to law (e.g., for medical databases). In order to address this issue, the database owner transforms the original database in such a way that sensitive rules cannot be mined from it. That is the transactions in the original database are slightly modified. However this transformation may lead to the side effects of lost rule, ghost rule and distorted entries. The lost rules are the non sensitive rules that are lost and ghost rules are the new rules that are generated as a result of this modification. Since the database is changed, even when the same mining algorithm with the same parameters is applied on the modified database, obtained rules could be different. A trade off should be found between the amount of modification (knowledge mined) and the data privacy. So Privacy Preserving Data Mining (PPDM) becomes a hot area of research and many algorithms were proposed to hide the confidential and private information before the data is shared. In PPDM, most of the approaches proposed hide sensitive information by perturbing the actual data. It is possible to perturb the data in different ways resulting in numerous solutions in the solution space. Each solution provides different degree of compromise between the knowledge mined and the security of private information. We need to choose an optimal solution that maximizes the knowledge mined without compromising security. Evolutionary algorithms are good in finding optimal solution for a problem with numerous solutions. Using Differential Evolution(DE), it is possible to explore the high value areas of the solution space and represent the real values of database as such. So DE is applied for effective hiding of sensitive information. This paper is organised as follows: section 2 reviews the literature for hiding sensitive association rules and section 3 provides the necessary background for the proposed algorithm. The proposed algorithm is described in section 4. The performance analysis of the proposed algorithm is provided in section 5 and the conclusion and future work is provided in section 6.

## 2 RELATED WORK

Agrawal and Srikant defined a quantitative measure to evaluate the usefulness of PPDM methods they proposed[1]. Then heuristic algorithms for identifying optimal data for sanitization based on support and confidence framework were developed [2-5]. Verikios et al. [4] proposed WSDA algorithm that hides the rules by reducing the confidence of the rule. Amiri [7] proposed a set of algorithms that hides sensitive item sets by removing transactions or items. The candidate transactions for removal are identified from the number of sensitive and non-sensitive item sets they support. Hong et al. devised a lattice-based algorithm that uses item deletion for hiding the sensitive information. The proposed lattice structure helped to speed up the sanitization process [8]. Then Hong et al. [9] used the TF – IDF concept in text mining to assign a SIF-IDF value to each transaction. This value ranks the transaction based on how much a transaction supports sensitive item set. Sun [10] proposed a border-based approach for preserving the border of non-sensitive frequent item sets. The importance was given to reduce the number of lost rules. Wu et al. [12] proposed a template based algorithm to eliminate the side effects of rule hiding but it still had the side effect of hiding failure. All above algorithms are distortion-based [13, 14] which perform the hiding task by removing or adding items in a dataset. Blocking based techniques introduced by Saygin et al. [5] performs hiding by replacing some original values of a dataset with unknowns. The blocking-based method does not add any fake information to the actual dataset which helps in avoiding discovering false knowledge in some applications like medical database analysis[4,11]. The optimal sanitization of databases is regarded to be an NP-hard problem [6].

Han and Ng proposed GA based secure protocols for discovering rules among the private data owned by two parties[15]. It uses the true positive and true negative to evaluate the goodness of each decision rule. Genetic algorithms (GAs) were usually used to find optimal solutions in the least amount of time [20]. Dehkordi et al [16] designed fitness function in such a way to reduce the side effects in the proposed multi objective genetic algorithm method. Hong et al. introduced pre large item set that are not large item set but has the potential to become large item set in future due to item insertion or deletion. This helps in updating the original data in single scan[19]. To solve the limitations of traditional GA-based algorithms with high requirements of memory and computations at each evolutionary process, the compact GA (cGA) mechanism [17] and the prelarge concept [18] are adopted in the proposed cpGA2DT algorithm. Shah et al. applied genetic algorithm for hiding sensitive rules[25] . The proposed Fitness function has two parts: i)transaction sensitivity and ii) Transaction priority. Transactions that contain the maximum number of sensitive items and minimum number of data items are chosen for modification. Hameed et al. integrated fuzzy correlation analysis and Apriori algorithm to mine fuzzy association rules[26]. For sanitization, a modification technique where maximum value of fuzzy items that occurs most frequently is substituted with zero.

Considering the side effects and the diversity of database, designing of various approaches for hiding sensitive association rules is still in progress in order to find good solution. The contribution of this work is:

- Hides Fuzzy Association Rules(FAR). Unlike binary association rule hiding that involves insertion and deletion of items, FAR hiding involves modifying the quantity of items. Most of the work in literature is for binary association rule hiding
- Uses Differential Evolution(DE) for FAR hiding. In literature DE was applied only for mining and not for hiding fuzzy association rules.
- Proposes a new measure called Association Measure(AM) which considers both the occurrence frequency and the information one item provides about the other.

## 3 BACKGROUND

An association rule is defined as an inference X→Y, where both X and Y are sets of attributes called items. Here X is called as the body of the rule and Y is called as the head of the rule. It is interpreted as: "for a specified fraction of the existing transactions, a particular value of an item set X determines the value of item set Y with a certain confidence". Support and confidence are the two measures that determine the usefulness of the association rules. Support is the percentage of transactions that contain both X and Y, while confidence is the ratio of the support of X UY to the support of X.

The association rules for quantitative data are different from that of the rules mined from binary dataset. The quantitative rules are specified as intervals for each of the items in the rule. Example, status(X, married), age(X, 40 -50), salary(X, 30,000 - 40,000) → owns(X, Cars) support = 60%, confidence = 40% . In the above example, items age and salary are given as intervals. This type of rules have sharp boundary problem. That is, those with salary 29,500 or age 39 have less probability of owning cars. In order to avoid this problem the quantitative data is fuzzified and fuzzy association rules are mined from this data.

Let I={$i_1$, $i_2$, $i_3$, ....$i_m$} be the complete item set where each $i_j$ (1≤j ≤m) is a quantitative item. Given a database D={$t_1$, $t_2$,…., $t_n$} where each $t_j$ is a transaction with items I. Let X ={$x_1$, $x_2$,…,$x_p$} and Y = {$y_1$, $y_2$,….., $y_q$} be two large disjoint itemsets, that are subset of I. Then, the fuzzy association rule is given as A→B where A={ $s_1$,$s_2$,…$s_p$} and B ={$p_1$,$p_2$,…..$p_q$} and $S_i$ ∈ {the fuzzy regions related to item $x_i$}, $P_j$ ∈ {the fuzzy regions related to item $y_j$}[23].The membership functions for fuzzy set take values in the interval [0,1] which is known as membership grade or degree of membership.

If the fuzzy set A with elements $z_1$ to $z_n$ have their membership grade as $\mu_1$ to $\mu_n$ respectively in A, then A is represented as $\mathrm{A} = \frac{\mu_1}{z_1} + \frac{\mu_2}{z_2} + \ldots\ldots \frac{\mu_n}{z_n}$ . The support of the fuzzy set A is given by the scalar cardinality of a fuzzy set A which is the summation of the membership grades of all the elements of X in A. Thus

$$|A| = \sum_{i=1}^{n} \mu_A(x_i) \tag{1}$$

The combined support of two fuzzy sets X and Y is denoted as Supp(X∩Y), and it can be found by fuzzy intersection as given below

$$\mu_{x\cap y}(P) = \min\{\mu_x(i), \mu_y(i)\} \tag{2}$$

Where i denotes the region of membership in the corresponding fuzzy set. Find the representative membership region for each item $I_j$ by finding the region with maximum scalar cardinality, countmax.

$$count_{jk} = \sum_{i=1}^{n} f_{ijk} \tag{3}$$

where i is the number of items, j is the number of transactions; k is number of fuzzy membership region. The fuzzy region with maximum count, $\mathrm{count}_j^{\max}$ for item $I_j$ would represent this item in mining process thereafter. With this background, the problem can be formulated as: Let D be the transactional database and QR be the set of interesting fuzzy association rules mined from it under given threshold association measure. Let QRs be the set of sensitive rules to be hided and $QR_n$ is the set of non-sensitive rules mined from D then $QR_s$ ∪ $QR_n$ = QR. Q` is the set of rules mined from sanitized database D`. The objective is to transform D into D` such that $QR_n$ = Q` while minimizing side effects like lost rule $\left(q \in QR_n \middle| q \notin Q`\right)$, Ghost rules $\left(q \in Q` \middle| q \notin QR_n\right)$, hiding failure $\left(q \in QR_s \middle| q \in Q`\right)$ and data distortion.

## 4 PROPOSED METHODOLOGY

Like genetic algorithm, differential evolution allows each fitter individual to evolve to the successive generation. The advantage of differential evolution over genetic algorithm is that it can be applied to real-valued problems over a continuous space easily. Differential evolution begins with a population chosen with equal probability from the problem space. DE generates new parameter vectors by adding the weighted difference vector between two population members to a third member. If the resulting vector yields a lower objective function value than a predetermined population member, the newly generated vector replaces the original vector with which it was compared. In order to keep track of the improvement, the best parameter vector is calculated for every generation.

The strategy adapted is DE/rand/1/bin as perturbation is performed on a randomly chosen vector with single vector difference where the weighted difference between two vectors is used to replace the third vector. Binomial crossover is applied in each of the variables until the random number chosen is less than crossover rate CR.

The objective is to reduce the association measure of a rule which is the weighted sum of the confidence and the Mutual Information(MI) between the items in the rule. Mutual information, measures the information that one attribute tells about another. Since the mutual information can provide the natural co-occurrence relationships between the attributes, it is used for finding frequent item sets and hence quantitative association rules. When mutual information is applied to quantitative association rule mining large number of irrelevant informative relationships between the items can be eliminated. Let x and y be two items and Let Qx and Qy be the fuzzy quantitative values in the dom(x) and dom(y) respectively [22]. Then the

$$\text{MI} = \sum_{Q_x \in dom(x)} \sum_{Q_y \in dom(y)} P(Q_x, Q_y) \log \frac{p(Q_x, Q_y)}{p(Q_x)\,p(Q_y)} \tag{4}$$

The strength of the relationship between two attributes occurring in the same quantitative association rule is given by Mutual Information(MI). But MI threshold µ does not reveal the amount of information one item tells about the other. If the Minimum confidence is set to 0.9 then it means the rules satisfying the threshold are of high quality. But in case of MI, we do not know how much information 0.9 provides. The MI between two items X and Y is normalized to uniform scale using maximum value of MI between X and Y as

$$\text{I}(x; y) = \frac{\text{I}(x;y)}{\text{I}(x;x)} = \frac{\sum_{Q_X \in dom(x)} \sum_{Q_y \in dom(y)} P(Q_x, Q_y) log \frac{P(Q_x, Q_y)}{P(Q_x) P(Q_y)}}{-\sum_{Q_X \in dom(x)} P(x) log P(x)} \tag{5}$$

Normalizing MI helps to get rid of the localness and make the normalized mutual information a global measure. The association measure is the sum of normalized mutual information and the support of the rule. So the objective function is to decrease the association measure.

F(x) = minimize(CEIL(0.7*(max(Mutual Information, confidence))+0.3*(min(Mutual Information, confidence)))) (6)

### 4.1 Steps of the algorithm

#### Algorithm HidesenRule

**Input:**

Dataset D = { $T_1$, $T_2$, $T_3$,...$T_n$}

T = {$I_1$, $I_2$, .. $I_m$} m - number of items in each transaction,

min_fitness - minimum fitness.

NoG – No. of generations

N is the number of transactions.

Number of parameters, K= Number of distinct items in the sensitive rule set.

Population Size = N (No. of transactions)

Parameter vector $A_{i,G}$ = [$A_{1,i,G}$, $A_{2,i,G}$, ...$A_{K,i,G}$], where G is the number of generation

**Output:** Transformed database D' so that sensitive association rules are hidden, hence cannot be mined.

**Step 1:** Cleaning. Database is pre-processed to remove inconsistency and redundancy.

**Step 2:** Fuzzify the dataset D → F.

**Step 3:** Generate fuzzy association rules. $R_S$ = { $r_1$, $r_2$, $r_3$, ... $r_x$} where x is the number of quantitative rules

**Step 4:** Obtain sensitive rule set $R_h$ from a set of interesting rule obtained from the previous step. $R_h$ = $R_S$ - $NS_y$ where $NS_y$ is the set of y non sensitive association rules.

**Step 5:** Apply differential evolution. CurrentGen = 1.

**Step 6:** While CurrentGen< NoG

**Step 7:** For i = 1 to N repeat steps 8 to 11

**Step** 8: Mutation :

- Expands search space. Randomly choose three vectors $A_{m,G}$, $A_{n,G}$ and $A_{o,G}$ for each parameter vector $A_{i,G}$ , where m ≠n ≠o ≠i .
- Add the weighted difference of two vectors to the third.

  $T_{i,G+1} = A_{m,G} + F(A_{n,G} – A_{o,G})$ where F is the mutation factor from [0.5, 1] and $T_{i,G+1}$ is the donor vector.

**Step 9:** For j = 1 to K repeat steps 10 and 11.

**Step 10:** Crossover: It is performed to pass better individuals to next generation.

$$u_{i,j,G+1} = \begin{cases} T_{i,j,G+1}, \text{if } \mathrm{rand}_j \le CR \text{ or } j = I_{rand} \\ A_{i,G}, \text{if } \mathrm{rand}_j > CR \text{ or } J \ne I_{rand} \end{cases} \tag{7}$$

Here i = 1 to N and J = 1 to K. $\mathrm{rand}_j \sim [0,1]$, $I_{rand}$ is a random integer between [1, K]. The integer random variable makes sure that the target vector and the donor vector are not same.

**Step 11:** Selection: Selection helps to choose the better individual that minimizes the objective function value.

$$A_{i,G+1} = \begin{cases} u_{i,G+1} \; \textit{if} \; f(u_{i,G+1}) \le f(A_{i,G}) \\ A_{i,G} \quad \text{otherwise} \end{cases} \quad i = 1,2,...N. \tag{8}$$

## 5 Performance Evaluations

The proposed approach was implemented in JAVA on the Windows platform and executed on an Intel CPU with four 2.67GHz processors and 8 GB of RAM. Extensive experiments were carried out on benchmarked datasets. The experiment results were measured according to side effects on knowledge and data as hiding failure, number of lost rule, Number of ghost rules and data modification or distortion. The objective is to reduce the values of these side effect metrics. Hiding failure refers to the proportion of sensitive rules which fail to be hidden. Data distortion denotes the count of sanitized transactions. The population size was equal to the number of transactions. The maximum number of generation for evolution was 100. The crossover rate was set to 0.5 and the weighing factor was set to 0.8. For each test case, the proposed algorithm was run for five times to get the average result. The threshold association measure was set as 30% and Minimum Support was set as 40%. For the problem of association rule hiding, solutions which do not reveal any sensitive rules but miss some non-sensitive ones or generate some ghost ones, are more preferred than solutions which reveal a few sensitive rules but produce no or fewer missing non-sensitive ones and ghost rules. The proposed algorithm produces no ghost rules and there is no hiding failure.

### 5.1 Datasets

The proposed algorithm was tested using two real datasets, breast cancer dataset and wine quality dataset from Wisconsin datasets [27]. These datasets exhibit varying characteristics with respect to the number of transactions and items they contain. The first dataset is breast cancer dataset which consists of one ID attribute and nine quantitative attributes with 699 instances. The ID attribute was ignored. The second dataset is white wine quality dataset. It contains 12 attributes and 4899 instances. The quality attribute was ignored.

### 5.2 Results And Analysis

The proposed method is compared with the genetic algorithm based [23] and PSO based [24] quantitative association rule hiding approaches. The GA based approach performs cross over in such a way that it reduces the support of the items in the right hand side of the rule. It maintains two factors namely difference factor and modification factor to reduce the side effects of knowledge and data distortion.

Two different runs of experiments were conducted to explore the performance of the proposed algorithm. In the first run, the consequence of different values of association measure on hiding was explored on the two real datasets. The number of rules mined for varying Association Measure (AM) was presented in table 1. We tried to hide three strong rules. And the number of lost rules generated while hiding the sensitive rules for the breast cancer and wine quality dataset were presented in figure 1 and 2 respectively. In the second run the effect of hiding rules in terms of data distortion on varying number of transactions was studied.

*Table 1.   Number rules for varying values of Association Measure (AM)*

| No. of Transactions | AM in % | No. of rules before hiding | | No. of rules after hiding | |
|---|---|---|---|---|---|
| | | Breast Cancer | Wine Quality | Breast Cancer | Wine Quality |
| **100** | 10 | 27 | 65 | 18 | 50 |
| **200** | 30 | 23 | 57 | 16 | 59 |
| **300** | 40 | 20 | 47 | 15 | 39 |
| **400** | 60 | 19 | 42 | 13 | 35 |
| **500** | 70 | 13 | 38 | 16 | 31 |

From the results it is evident that DE approach has higher side effect in terms of lost rule than the GA and PSO based approach. This is because to hide the rules with high confidence more number of transactions has to be distorted. As a result non sensitive rule with the same item on the antecedent or the consequent of the rule also gets hided. Since DE considers the quantitative value of an item in all the transactions as vector and this vector as a whole is manipulated, the amount of data distortion is also higher as shown in table 2 and table 3 for breast cancer and wine quality dataset respectively.

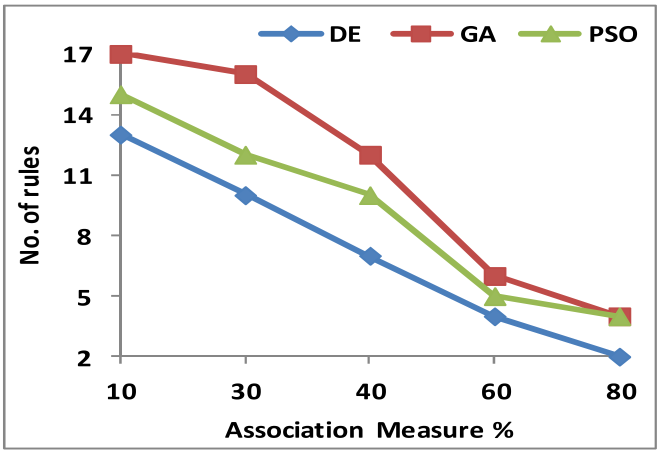

*Fig. 1.   Number of Lost rules in Breast Cancer Dataset*

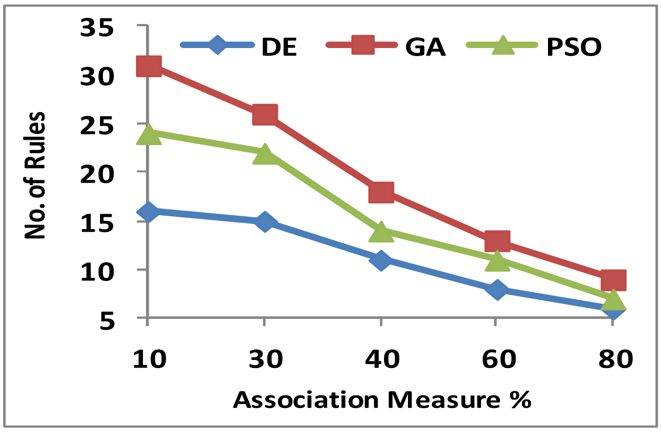

*Table 2. Data Distortion for Breast Cancer Dataset*

| Total Entries | Modified Entries | | |
|---|---|---|---|
| | DE | GA | PSO |
| 1800 | 432 | 412 | 425 |
| 2700 | 398 | 345 | 367 |
| 3600 | 371 | 265 | 324 |
| 4500 | 297 | 215 | 267 |
| 5400 | 260 | 185 | 223 |

*Table 3. Data Distortion for Wine Quality Dataset*

| Total Entries | Modified Entries | | |
|---|---|---|---|
| | DE | GA | PSO |
| 11000 | 6342 | 4810 | 4065 |
| 22000 | 5778 | 3968 | 3675 |
| 33000 | 4311 | 3227 | 2087 |
| 44000 | 2097 | 1856 | 1796 |
| 49500 | 1660 | 1464 | 1538 |

The number of ghost rules produced while hiding three strong rules were given in figure 3 and 4 respectively for the breast cancer and wine quality dataset. Since the objective is a minimization function it reduces the association measure value of all the items in sensitive transactions, no items support is increased. Thus no ghost rules are generated in this approach.

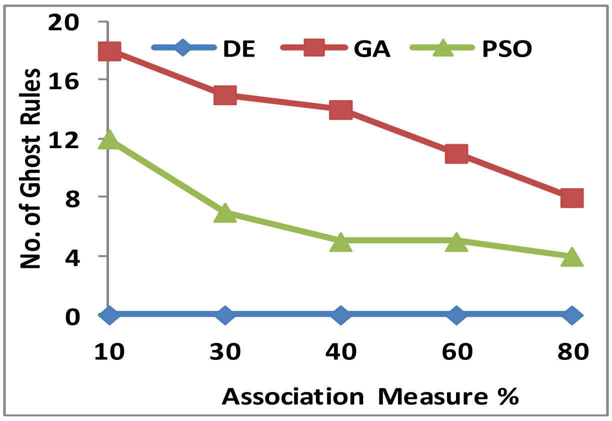

*Fig. 3. Number of Ghost rules in Breast Cancer Dataset*

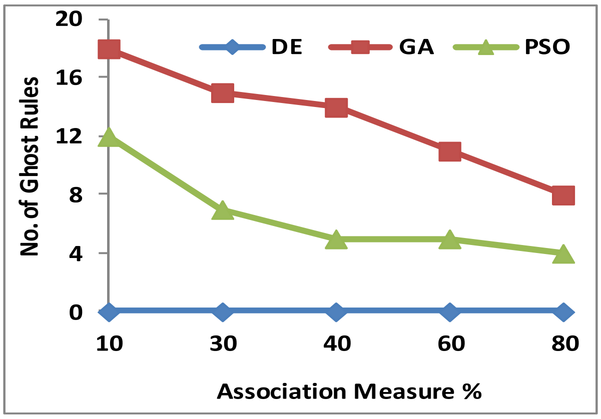

*Fig. 4. Number of Ghost rules in Wine Quality Dataset*

From the above results it is evident that when the threshold association measure increases the number of modification to the data decreases as a result the side effects in terms lost rules, ghost rules and knowledge distortion is also lower. The proposed algorithm hides all the sensitive rules and there is no hiding failure. But the data distortion is higher which in turn increases the knowledge distortion (lost rules). But this knowledge distortion is lesser than when compared with GA and PSO based approach. And the algorithm also performs consistently on both the datasets. While sharing the data, the level of data accuracy can be specified in the agreement so that the data quality and privacy can be maintained

## 6 CONCLUSION

In this paper, we used the differential Evolution algorithm to hide the quantitative association rules. Unlike binary association rule hiding which involves addition and deletion of items, quantitative rule mining requires the quantity of items to be modified without distorting the knowledge gained from it. A new measure called association measure is proposed which not only considers the occurrence frequency of items but also the information one item provides about the other. The objective is to hide the quantitative sensitive association rules while minimizing the side effects. The performance of the algorithm is empirically compared with the GA based and PSO based algorithm. The results show that the proposed algorithm generates no hiding failure, ghost rules and less number of lost rules. In future, methods would be proposed to reduce the data distortion.

## REFERENCES

[1] Agrawal R. and Srikant R., “Privacy-preserving data mining,” *SIGMOD Record*, vol. 29, no. 2, pp. 439–450, 2000.

[2] Dasseni E., Verykios V.S., Elmagarmid A.K. and Bertino E., “Hiding association rules by using confidence and support,” *In Proceedings of the International Workshop on Information Hiding*, pp. 369–383, 2001.

[3] Verykios V.S., Elmagarmid A.K., Bertino E., Saygin Y. and Dasseni E., “Association rule hiding”, *IEEE Transactions on Knowledge and Data Engineering*, Vol.16, pp. 434–447, 2004.

[4] Dasseni E., Verykios V.S., Elmagarmid A.K. and Bertino E., “Hiding association rules by using confidence and support”, *In: Information Hiding. Springer Berlin Heidelberg*, vol. 2137, pp. 369–383,2001

[5] Verykios V.S., Pontikakis E.D., Theodoridis Y. and Chang L., “Efficient algorithms for distortion and blocking techniques in association rule hiding”, *Distributed and Parallel Databases*, Vol. 22, pp. 85–104, 2007.

[6] Saygin Y., Verykios V.S.and Clifton C., “Using unknowns to prevent discovery of association rules”, *ACM SIGMOD Record* , Vol. 30, pp. 45–54, 2001.

[7] Verykios V. S., Bertino E., Fovino I.N., Provenza, L.P., Saygin Y. and Theodoridis Y., “State-of-the-art in privacy preserving data mining,” *SIGMOD Record*, vol. 33, no. 1, pp. 50–57, 2004.

[8] Amiri A.,”Dare to share: Protecting sensitive knowledge with data sanitization”, *Decision Support Systems*, Vol.43, pp.181–191, 2007.

[8] Hong T.P., Lin C., Yang K. and Wang S., “A lattice-based data sanitization approach,” *In Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC ’11)*, pp. 2325–2329, October 2011.

[9] Hong T.P., Lin C.W., Yang K.T. and Wang S.L., “Using TF-IDF to hide sensitive itemsets”, *Applied Intelligence,* Vol. 38, pp. 502–510, 2013.

[10] Sun X. and Yu P.S., "A border-based approach for hiding sensitive frequent itemsets", *In: Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM),* pp. 426–433, 2005.

[11] Pontikakis E.D., Theodoridis Y., Tsitsonis A.A., Chang L. and Verykios V.S., "A quantitative and qualitative analysis of blocking in association rule hiding", *In: Proceedings of the 2004 ACM workshop on Privacy in the electronic society. ACM,* pp. 29–30, 2004 .

[12] Wu Y.H., Chiang C.M. and Chen A.L., "Hiding sensitive association rules with limited side effects", *IEEE Transactions on Knowledge and Data Engineering*, Vol.19, pp. 29–42, 2007.

[13] Patil S. P. and Patewar T.M., "A novel approach for efficient mining and hiding of sensitive association rule", *In Proceedings of the 2012 Nirma University International Conference on Engineering*, pp. 1–6, 2012.

[14] Hong T.P., Lin C., Chang C. and Wang S., "Hiding sensitive itemsets by inserting dummy transactions", *In Proceedings of the IEEE International Conference on Granular Computing (GrC '11),* pp. 246–249, November 2011.

[15] Han S. and Ng W.K., "Privacy-preserving genetic algorithms for rule discovery," *In Data Warehousing andKnowledge Discovery, LNCS, Springer, Berlin, Germany*, vol. 4654, pp. 407–417, 2007.

[16] Dehkordi M.N., Badie K., and Zadeh A. K., "A novel method for privacy preserving in association rule mining based on genetic algorithms", *Journal of Software*, vol. 4, no. 6, pp. 555–562, 2009.

[17] Harik G. R., Lobo F.G. and Goldberg D.E., "The compact genetic algorithm", *IEEE Transactions on Evolutionary Computation*, vol. 3, no. 4, pp. 287–297, 1999.

[18] Hong T. P. and Wang C. Y., "Maintenance of association rules using pre-large item sets", *In Intelligent Databases: Technologies and Applications*, pp. 44–60, 2007.

[19] Hong T.P., Wang C.Y. and Tao Y. H., "A new incremental data mining algorithm using pre-large itemsets", *Intelligent Data Analysis*, vol. 5, pp. 111–129, 2001.

[20] Hong T. P., Yang I., Lin C. and Wang S., "Evolutionary privacy preserving data mining", *In Proceedings of the World Automation Congress (WAC "10),* pp. 1–7, September 2010.

[21] Hegerty B., Hung C.C. and Kasprak K., ''A comparative study on differential evolution and genetic algorithms for some combinatorial problems'' , *In Proceedings of 8th Mexican International Conference on Artificial Intelligence*, 2009.

[22] Ke Y., Cheng J. and Ng W.," An Information-Theoretic Approach to Quantitative Association Rule Mining", *Knowledge and Information Systems*, Vol. 16 Iss. 2, pp. 213-244, July 2008

[23] Sathiyapriya K., Sadasivam G. S. and Karthikeyan V.B., "A New Method for preserving privacy in Quantitative Association Rules using Genetic Algorithm", *International Journal of Computer Applications,* Vol. 60, Iss.12, pp.12-19, 2012.

[24] SathiyaPriya K., Sadasivam G. S. and Sathiyan S.," Hiding Quantitative Sensitive Association Rules", *Proceedings of PSG-ACM National Conference on Intelligent Computing (NCIC-2013)*, pp.26-27, April 2013.

[25] Sonia Hameed, F. Shahzad and S. Asghar," Sanitizing Sensitive Association Rules using Fuzzy Correlation Scheme", *The Nucleus,* Vol. 50, No. 4, Pp. 359-367, 2013.

[26] R. A. Shah and S. Asghar, "Privacy preserving in association rules using genetic algorithm," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 22, issue 2, pp. 434-450, March 2014.

[27] https://archive.ics.uci.edu/ml/datasets