Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

Mathematical Biosciences and Engineering, 16(3): 1718–1728.
DOI: 10.3934/mbe.2019082
Received: 14 December 2018
Accepted: 31 January 2019
Published: 27 February 2019

Research article

The density-based clustering method for privacy-preserving data mining

Jimmy Ming-Tai Wu[1], Jerry Chun-Wei Lin[2,3,*], Philippe Fournier-Viger[4], Youcef Djenouri[5], Chun-Hao Chen[6] and Zhongcui Li[1]

[1] College of Computer Science and Engineering, Shandong University of Science and Technology, Qindao, Shandong, China

[2] School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, China

[3] Department of Computing, Mathematics, and Physcis, Western Norway University of Applied Sciences, Bergen, Norway

[4] School of Humanities and Social Sciences, Harbin Institute of Technology (Shenzhen), Shenzhen, China

[5] Department of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway

[6] Department of Computer Science and Information Engineering, Tamkang University, New Taipei City, Taiwan

* **Correspondence:** Email: jerrylin@ieee.org.

Abstract: Privacy-preserving data mining has become an interesting and emerging issue in recent years since it can, not only hide the sensitive information but still mine the meaningful knowledge at the same time. Since privacy-preserving data mining is a non-trivial task, which is also concerned as a NP-hard problem, several evolutionary algorithms were presented to find the optimized solutions but most of them focus on considering a single-objective function with the pre-defined weight values of three side effects (*hiding failure*, *missing cost*, and *artificial cost*). In this paper, we aim at designing a multiple objective particle swarm optimization method for hiding the sensitive information based on the density clustering approach (named CMPSO). The presented CMPSO is more flexible to select the most appropriate solutions for hiding the sensitive information based on user's preference. Extensive experiments are carried on two datasets to show that the designed CMPSO algorithm has good performance than the traditional single-objective evolutionary approaches in terms of three side effects.

Keywords: density clustering; Pareto solutions; optimization; PPDM; deletion

1. Introduction

Data mining, also, refers as knowledge discovery in database (KDD) [2, 8, 16, 17], is a commonly way to find the useful and meaningful knowledge from databases, which could be utilized in several real-life applications. Although data mining techniques can discover the useful information and knowledge, it reveals, however, the confidential or private message from the discovered information, which may cause the security threats [11, 12, 39]. Privacy-preserving data mining [3, 5, 35, 37, 31, 32] has thus become an important issue in recent decades since it can not only hide the private information but also discover the required information by varied data mining techniques. Data sanitization is one of the major research of PPDM that perturbs the database to hide the sensitive information. During the sanitization progress, it usually causes three side effects, for instance, *hiding failure*, *missing cost*, and *artificial cost*. Existing methods are used to minimize those three side effects during the sanitization progress, which can be regarded as the NP-hard problem [3, 37]. Several works related to PPDM were respectively discussed and most of them are based on the deletion procedure to hide the sensitive information [13, 14, 21, 27, 38].

Conventional algorithms in PPDM mostly focus on hiding the sensitive information as much as possible; the PPDM is, however considered as the NP-hard problem since it has the trade-off relationships among the derived three side effects. Lin et al. proposed the GA-based algorithms to hide the sensitive itemsets and utilize transaction deletion procedure for sanitization [28, 29]. The optimal transactions for deletion are then obtained and the results showed that less side effects can be achieved compared to the greedy-based method. Lin et al. [30] then designed the PSO-based model to handle the sanitization progress in PPDM. However, the above algorithms rely on the pre-defined weights to design the importances of three side effects, which requires the a-priori knowledge to set up the weight values. The pre-defined weights may seriously affect the results to select the transactions for deletion. To handle this problem, Cheng et al. [10] developed the EMO-based algorithm to consider "data distortion" and "knowledge distortion" in the sanitization progress through item deletion. This approach is efficient but it may lead the incomplete knowledge in the mining procedure. For example, the deleted information in the hospital diagnosis may mislead the wrong diagnosis and treatment.

Multi-objective particle swarm optimization (MOPSO) algorithm [7] is extended from conventional particle swarm optimization algorithm [24], that handles the multi-objective problems to find a set of Pareto solutions. Since the MOPSO-based framework uses the dominance relationship to find better solutions, the traditional method used in the PSO-based framework cannot be used to find better solutions. In this paper, we aim at designing a multiple optimization particle swarm optimization algorithm by adapting the density-based (CMPSO) method for finding better solutions rather than the single objective algorithms. Major contributions are listed below.

- We design a multiple optimization particle swarm optimization (named CMPSO) algorithm for hiding the sensitive information, which achieves better performance than the single objective algorithms in terms of three side effects.
- The density-based clustering method is adapted here to obtain better diversity of the derived Pareto solutions.
- Extensive experiments are conducted to show the performance of the designed algorithm in terms of three side effects, and results outperform the single-objective algorithms.

The rest of this paper is organized as follows. Literature review is shown in Section 2. Preliminary and problem statement are discussed in Section 3. The designed algorithm is stated in Section 4. Several experiments are carried and conducted in Section 5. Conclusion and future work are discussed in Section 6.

2. Literature review

The related works of privacy-preserving data mining (PPDM) and the evolutionary computation are described as follows.

2.1. Privacy-preserving data mining

In recent two decades, data mining techniques [17, 16, 33] are efficient to mine and find the potential information from a very large database, especially the relationships between the products may not be easily discovered and visualized. Since the important data can be revealed from the databases, thus the confidential and secure information can also be discovered during the mining procedure, that causes the privacy and security threats to users. Privacy-preserving data mining (PPDM) has become a critical issue in recent years since it can, not only find the useful information but also hide the confidential data after the sanitization progress. Agrawal and Srikant [4] developed a novel reconstruction approach that can accurately estimate the distribution of original data. The classifiers can also be constructed to compare the accuracy between original data and the sanitized one. Verykios et al. [37] developed the hierarchical classification techniques used in PPDM. Dasseni et al. [13] then designed an approach based on the hamming-distance mechanism to decrease the support or confidence of the sensitive information (i.e., association rules) for sanitization. Oliveira and Zaiane [35] provides several sanitization methods, that can hide the frequent itemsets through the heuristic framework. Islam and Brankovic [22] presented a framework using the noise addition method to protect and hide the individual privacy and still maintained high data quality. Hong et al. provided the SIF-IDF method [21] to assign the weight to each transaction, and the sanitization progress is then performed from the transaction with highest score to the lowest one for hiding the sensitive information. Lin et al. [31] then presented a SPMF software to provide several algorithms for handling the sanitization problem in PPDM.

The PPDM is, however, considered as a NP-hard problem [3, 37], it is thus better to provide the meta-heuristic approaches to find the optimal solutions. Han and Ng developed a secure protocol to find a better set of rules without disclosing their own private data based on GAs [18]. Lin et al. then presented several GA-based approaches, such as sGA2DT [29], pGA2DT [29] and cpGA2DT [28] to hide the sensitive information by removing the victim transactions. The encoded chromosome is considered as a set of solutions, and the transaction of the gene within chromosome is concerned as the victim for later deletion. A fitness function is also developed to consider three side effects for evaluation with the pre-defined weights to show the goodness of the chromosome. Several extensions of the evolutionary computation used in PPDM were also developed [20, 30]. Although the above algorithms are efficient to find the optimal transactions for deletion, they still require, however, the pre-defined weights of the side effects; this mechanism can seriously affect the final results of the designed approaches. Cheng et al. [10] then developed a EMO-RH algorithm to hide the sensitive itemsets by removing the itemset. This approach is based on multi-objective method, but the incomplete transactions can thus be produced; the mislead decision could thus be made especially in treatment of the hospital diagnoses. Lin

et al. also applies the multi-objective model [32] for the deletion problem based on the grid method. An NSGA-II model [34] is also utilized to efficiently sanitize the database for transaction deletion.

2.2. Evolutionary computation

Evolutionary computation is inspired by biological evolution that attempts to find the global optimization as the solutions. An iteratively progress of the candidates is then executed to gradually find the optimized solutions. Holland [18] first applies Darwin theory of natural selection and survival of the fittest to designed the genetic algorithms (GAs), that is widely used in computational intelligence. The GAs is the fundamental method, which is the population-based method to solve the NP-hard problems. It consists of mutation, crossover, and selection operators to find the possible solutions in the evolution progress. Particle swarm optimization (PSO) is a swarm intelligence algorithm proposed by Kennedy and Eberhart that is inspired by bird flocking activities [24]. Each particle in the PSO is concerned as a potential solution. In PSO, each bird has its own velocity, which is used to represent the direction to the other solutions, and each particle is then updated with its own best value (represents as *pbest*) and the global best value (represented as *gbest*) according to the designed fitness function during each iteration. The updating equations are listed as follows [24]:

$$v_i(t+1) = w \times v_i(t) + c_1 \times r_1 \times (pbest_i - x_i(t)) + c_2 \times r_2 \times (gbest - x_i(t)) \tag{2.1}$$

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{2.2}$$

From the above equations, w is considered as a factor that is used to balance the influence of global search and local search; v_i represents the velocity of the i-th particle in a population where t presents the t-th iteration. The c_1 and c_2 are the constant values. Both r_1 and r_2 are random numbers generated by a uniform distribution in the range of [0, 1]. The velocity of a particle is updated by equation (2.1), and its position is updated by equation (2.2).

In real-world situations, optimization should consider the multiple objectives [23, 36] since we may focus on varied targets to find the solutions. Several algorithms were presented to solve the multi-objective problems for optimization, such as the Non-dominated Sorting Genetic Algorithm (NSGA) [36], the strength pareto evolutionary algorithm (SPEA) [40], and the pareto archived evolution strategy (PAES) [25]. Coello et al. [7] applies the adaptive grid method to maintain the external archive. Several variants of MOPSO were respectively presented to improve the performance, and the development is still in progress.

3. Preliminary and problem statement

Let $I = \{i_1, i_2, \ldots, i_m\}$ be a finite set of r distinct items occurring in the database D, and D is a set of transactions where $D = \{T_1, T_2, \ldots, T_n\}$, $T_q \in D$. Each T_q is a subset of I and has a unique identifier q, called *TID*. The set of sensitive itemsets is denoted as $SI = \{s_1, s_2, \ldots, s_k\}$, which are given by user. Note that each s_i of SI is a subset of T_q. The μ is the minimum support threshold in D. If $sup(i_j) \geq \mu \times |D|$, the itemset (i_j) is defined as the frequent (large) itemset.

Definition 1. *For each $s_i \in SI$, the number of transactions for deletion of s_i is denoted as N_{s_i} and*

defined as:

$$N_{s_i} = \frac{sup(s_i) - \mu \times |D|}{1 - \mu}. \tag{3.1}$$

Definition 2. *The **M**aximum number of **D**eleted **T**ransactions among all sensitive itemsets in SI is denoted as MDL and defined as:*

$$MDT = max\{N_{s_1}, N_{s_2}, \ldots, N_{s_k}\}. \tag{3.2}$$

Thus, the MDT is treated as the size of the particle in the designed CMPSO algorithm. In the PPDM, three side effects can be considered as "*hiding failure*", "*missing cost*", and "*artificial cost*", which should be minimized in the PPDM. We can obtain the definitions for three side effects as follows.

Definition 3. *Let α be the number of sensitive itemsets that fails to be hidden, in which $\alpha = |SI \cup L'|$.*

Definition 4. *Let β be the number of missing itemsets (cost), which is the non-sensitive itemset and the large itemset in the original database but will be hidden after the sanitization progress as: $\beta = L - SI - L'$.*

Definition 5. *Let γ be the number of artificial itemsets (cost), which was not the frequent itemset but will be arisen as the frequent one after the sanitization progress as: $\gamma = L' - L$.*

Problem Statement: The problem of PPDM with transaction deletion based on the multi-objective particle swarm optimization is to minimize three side effects by considering the sanitized results in three objective functions, which can be defined as:

$$f = min[f_1, f_2, f_3], \tag{3.3}$$

where f_1 equals to the number of α, f_2 equals to the number of β, and f_3 equals to the number of the γ.

4. Proposed CMPSO algorithm

In this section, we develop a CMPSO algorithm to hide the sensitive information by adapting the density-based clustering method to obtain higher diversity of the derived Pareto solutions, and the effects can be shown in terms of three side effects. To obtain better transactions for deletion in PPDM, the database is first processed to project the transactions with any of the sensitive information within the set of *SI*, and the projected database is named as D^*. Thus, only the transactions with the sensitive information are projected for later deletion, and those transactions in D^* are considered as the candidate of the particle in the evolution process. After that, the CMPSO algorithm is then iteratively performed to evaluate three side effects in the evolutionary progress. In CMPSO algorithm, each particle can be represented as a possible solution with MDT vectors, and each vector is the transaction ID, which shows the potential transaction for deletion. Note that a vector in a particle can be a ***null*** value. In the updating progress of the CMPSO in PPDM, the formulas are thus defined as follows [30]:

$$v_i(t+1) = (pbest - x_i(t)) \cup (gbest - x_i(t)). \tag{4.1}$$

$$x_i(t+1) = rand(x_i(t), null) + v_i(t+1). \tag{4.2}$$

Algorithm 1: Designed CMPSO Algorithm

Input: D^*, the projected database; L, the set of large itemsets for evolution; SI, the set of sesntive information to be hidden.

Output: D', the sanitized database; $Pset$, the set of Pareto solutions.

```
1  initial N Particles with MDT size;
2  for each particle p in N do
3      evaluate f(p) := [f1(p), f2(p), f3(p)];
4      obtain the non-dominated solutions Pset;
5  while termination criteria is not achieved do
6      Gbest_update(Pset);
7      update pbest;
8      for each particle p' in N(t + 1) do
9          evaluate f(p') := [f1(p'), f2(p'), f3(p')];
10         update the non-dominated solutions Pset;
```

Thus, the particles in the designed CMPSO is updated by the above equations for the velocity and its position. The designed CMPSO algorithm is then described in Algorithm 1 as follows.

From Algorithm 1, the designed CMPSO algorithm generates N particles with MDT size (Line 1). Each particle is then evaluated by the fitness functions with three side effects (Lines 2 to 3) to find the non-dominated solutions (Line 4). In the updating progress, the *pbest* and *gbest* are then respectively updated (Lines 6 to 7). The *gbest* is then updated by the sub-function to obtain a better particle for evolution. After the updating progress (Line 8), the updated particles are then evaluated again to update the non-dominated solutions. The algorithm is recursively performed until the termination criteria is achieved (Lines 5 to 10). After that, the final results will be returned with a set of non-dominated solutions (Pareto solutions).

To obtain better Pareto solutions with higher diversity, the density-based clustering method is adapted here to select the appropriate *gbest* in the evolutionary progress, which was described at Line 6 in Algorithm 1. Details are then stated in Algorithm 2.

For each Pareto solution in Algorithm 2, it finds a cluster center (Line 4) with satisfied solutions (number of Pareto solutions is no less than *minpts*) within a radius r (Line 3). After that, each solution within a cluster is then assigned with a probability (Lines 6 to 7), and a global solution (*gbest*) is then derived randomly by the assigned probability (Line 8). This progress increases the diversity of the discovered Pareto solutions in the evolution progress. After that, the *gbest* is then derived and used in the Algorithm 1 for iteratively evolution.

5. Experimental results

Substantial experiments were carried to compare the effectiveness and efficiency of the proposed CMPSO algorithm to the state-of-the-art single-objective cpGA2DT [28] and PSO2DT [30] approaches. The algorithms in the experiments are implemented in Java language, performing on a PC with an Intel Core i7-6700 quad-core processor and 8 GB of RAM under the 64-bit Microsoft

Algorithm 2: Gbest_update(*Paretos*)

Input: *Paretos*, a set of Pareto solutions; *minpts*, the minimum number of solutions; *r*, the radius of a cluster.

Output: *gbest*, a global best particle for the updating progress.

```
set i := 1;
for each p in Paretos do
    if sizeof(p, r) ≥ minpts then
        c_i ← p;
        i++;
for each c_t, t := 1 to i do
    prob(p ∈ c_t) := (1/i) × 1/sizeof(c_t);
gbest := rand(prob(p ∈ Paretos));
```

Windows 10 operating system. Two real-world datasets [15] called chess and foodmart are used in the experiments. Results in terms of three side effects are respectively discussed and analyzed as follows.

5.1. Side effects

In this section, the state-of-the-art single-objective algorithms such as cpGA2DT [28] and PSO2DT [30] are then compared to the designed CMPSO algorithm. The number of population for all evolutionary algorithm is set as 50. Since the multi-objective algorithm would produce a set of Pareto-front solutions, we evaluate the side effects by average of the produced solutions. Thus, in the dense dataset such as chess, the CMPSO sometimes shows worse results in terms of side effects compared to the single objective algorithms, but in the sparse dataset such as foodmart, the CMPSO generally obtains good performance than the others. Results for the chess dataset are shown in Figure 1.

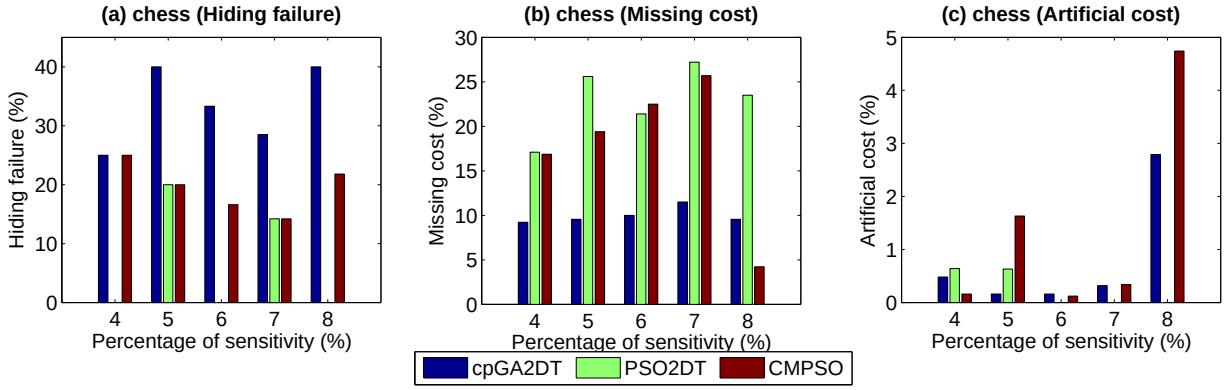

Figure 1. Compared results of the chess dataset.

From Figure 1, it can be seen that the single-objective algorithms such as cpGA2DT and PSO2DT sometimes have better performance than the designed CMPSO approach. The reasons are that the designed CMPSO algorithm evaluates three side effects by the average value of the derived solutions. For the dense dataset such as chess, the diversity of the derived solutions is not high; the discovered Pareto solutions could be converged together due to the characteristics of the dataset. Moreover, The

designed CMPSO algorithm produces a set of solutions, for example in the experiments, the number of Pareto solutions can be more than 20. Thus, the average values of three side effects sometimes could not achieve better results. Experiments under the sparse foodmart dataset are shown in Figure 2.

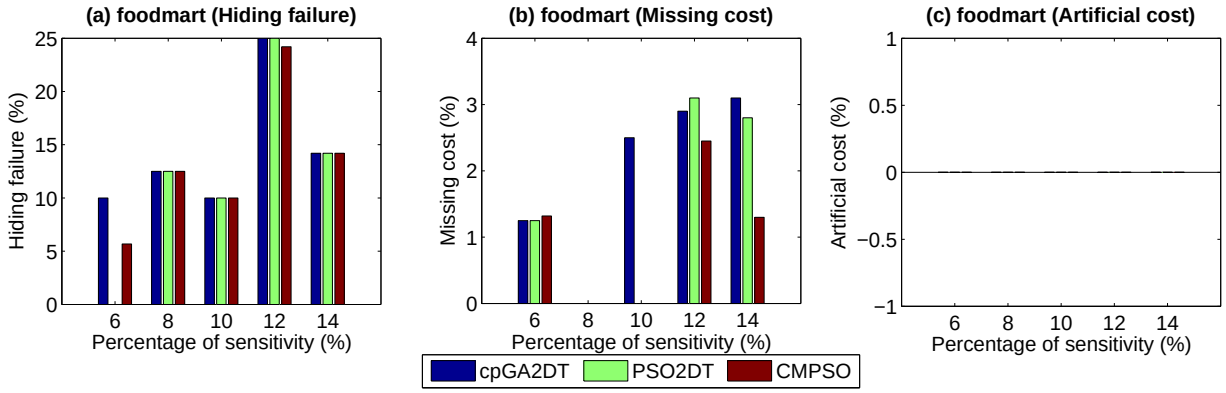

Figure 2. Compared results of the foodmart dataset.

From Figure 2, we can observe that the designed CMPSO algorithm has good performance than the single-objective cpGA2DT and PSO2DT algorithms in most cases. For example, when the sensitive percentages are respectively set as 8%, 10%, 12%, and 14%, the designed CMPSO achieves best performance than the other single-objective algorithms. The reasons are that when the dataset is a sparse dataset, higher diversity of the designed CMPSO algorithm can be obtained. Thus, the selected transactions may contain the minimal side effects for deletion. Moreover, all three algorithms have no artificial cost under varied percentages of sensitivity.

6. Conclusion and future work

In recent years, many algorithms were presented to hide the private/confidential information in privacy-preserving data mining (PPDM) and most of them were designed to sanitize the database for hiding the sensitive information. In this paper, we first introduce a CMPSO algorithm for hiding the sensitive information based on density clustering method for sanitizing the databases. The designed updating algorithm for *gbest* can achieve good performance compared to the single objective algorithms, especially for the sparse dataset, which can be seen in the experiments.

Acknowledgments

This research was partially supported by the Shenzhen Technical Project under KQJSCX20170726103424709 and JCYJ20170307151733005

References

1. R. Agrawal and R. Srikant, Quest synthetic data generator, IBM Almaden Research Center. Available from: `http://www.Almaden.ibm.com/cs/quest/syndata.html`, (1994).
2. R. Agrawal and R. Srikant, Fast algorithms for mining association rules in large databases, *The International Conference on Very Large Data Base*, (1994), 487–499.

3. M. Atallah, E. Bertino, A. Elmagarmid, et al., Disclosure limitation of sensitive rules, *The Workshop on Knowledge and Data Engineering Exchange*, (1999), 45–52.
4. R. Agrawal and R. Srikant, Privacy-preserving data mining, *ACM SIGMOD Record*, **29** (2000), 439–450.
5. C. C. Aggarwal, J. Pei and B. Zhang, On privacy preservation against adversarial data mining, *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2006), 510–516.
6. D. W. Cheung, J. Han, V. T. Ng, et al., Maintenance of discovered association rules in large databases: An incremental updating technique, *The International Conference on Data Engineering*, (1996), 106–114.
7. C. A. Coello and M. S. Lechuga, MOPSO: a proposal for multiple objective particle swarm optimization, *IEEE Congress on Evolutionary Computation*, (2002), 1051–1056.
8. M. S. Chen, J. Han and P. S. Yu, Data mining: An overview from a database perspective, *IEEE T. Knowl. Data En.*, **8** (1996), 866–883.
9. C. Clifton, M. Kantarcioglu, J. Vaidya, et al., Tools for privacy preserving distributed data mining, *ACM SIGKDD Explorations*, **4** (2003), 1–7.
10. P. Cheng, I. Lee, C. W. Lin, et al., Association rule hiding based on evolutionary multi-objective optimization, *Intell. Data Anal.*, **20** (2016), 495–514.
11. C. M. Chen, B. Xiang, Y. Liu, et al., A secure authentication protocol for internet of vehicles, *IEEE Access* (2019), DOI:10.1109/ACCESS.2019.2891105.
12. C. M. Chen, B. Xiang, K. H. Wang, et al., A robust mutual authentication with a key agreement scheme for session initiation protocol, *Appl. Sci.*, **8** (2018).
13. E. Dasseni, V. S. Verykios, A. K. Elmagarmid, et al., Hiding association rules by using confidence and support, *International Workshop on Information Hiding*, (2001), 369–383.
14. A. Evfimievski, R. Srikant, R. Agrawal, et al., Privacy preserving mining of association rules, *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (2002), 217–228.
15. P. Fournier-Viger, J. C. W. Lin, A. Gomariz, et al., The SPMF open-source data mining library version 2, *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, (2016), 36–40.
16. W. Gan, J. C. W. Lin, H. C. Chao, et al., Data mining in distributed environment: a survey *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **7** (2017), 1–19.
17. W. Gan, J. C. W. Lin, P. P. Fournier-Viger, et al., A survey of incremental highutility itemset mining, *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, **8** (2018), 1–23.
18. J. H. Holland, Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control and artificial intelligence, MIT Press, (1992).
19. J. Han, J. Pei, Y. Yin, et al., Mining frequent patterns without candidate generation: A frequent-pattern tree approach, *Data Mining and Knowledge Discovery*, **8** (2004), 53–87.
20. S. Han and W. K. Ng, Privacy-preserving genetic algorithms for rule discovery, *International Conference on Data Warehousing and Knowledge Discovery*, (2007), 407–417.

21. T. P. Hong, C. W. Lin, K. T. Yang, et al., Using TF-IDF to hide sensitive itemsets, *Appl. Intell.*, **38** (2012), 502–510.
22. M. Z. Islam and L. Brankovic, Privacy preserving data mining: A noise addition framework using a novel clustering technique, *Knowl. Based Syst.*, **24** (2011), 1214–1223.
23. S. Jeyadevi, S. Baskar, C. K. Babulal, et al. Solving multiobjective optimal reactive power dispatch using modified NSGA-II, *Int. J. Elec. Power.*, **33** (2011), 219–228.
24. J. Kennedy and R. Eberhart, Particle swarm optimization, *IEEE International Conference on Neural Networks*, (1995), 1942–1948.
25. J. Knowles and D. Corne, The pareto archived evolution strategy: a new baseline algorithm for Pareto multiobjective optimisation, (1999), 98–105.
26. Y. Lindell and B. Pinkas, Privacy preserving data mining, *The Annual International Cryptology Conference on Advances in Cryptology*, (2000), 36–54.
27. C. W. Lin, T. P. Hong, C. C. Chang, et al., A greedy-based approach for hiding sensitive itemsets by transaction insertion, *J. Inform. Hiding Multimed. Signal Proc.*, **4** (2013), 201–227.
28. C. W. Lin, B. Zhang, K. T. Yang, et al., Efficiently hiding sensitive itemsets with transaction deletion based on genetic algorithms, *The Scientific World J.*, (2014).
29. C. W. Lin, T. P. Hong, K. T. Yang, et al., The GA-based algorithms for optimizing hiding sensitive itemsets through transaction deletion, *Appl. Intell.*, **42** (2015), 210–230.
30. J. C. W. Lin, Q. Liu and P. Fournier-Viger, A sanitization approach for hiding sensitive itemsets based on particle swarm optimization, *Eng. Appl. Artif. Intel.*, **53** (2016), 1–18.
31. J. C. W. Lin, P. Fournier-Viger, L. Wu, et al., PPSF: An open-source privacy-preserving and security mining framework, *IEEE International Conference on Data Mining Workshop*, (2018), 1459–1463.
32. J. C. W. Lin, Y. Zhang, C. H. Chen, et al., A multiple objective PSO-based approach for data sanitization, *The 2018 Conference on Technologies and Applications of Artificial Intelligence*, (2018), 148–151.
33. J. C. W. Lin, L. Yang, P. Fournier-Viger, et al., Mining of skyline patterns by considering both frequent and utility constraints, *Eng. Appl. Artif. Intel.*, **77** (2019), 229–238.
34. J. C. W. Lin, Y. Zhang, B. Zhang, et al., Hiding sensitive itemsets with multiple objective optimization, *Soft Comput.*, (2019), 1–19.
35. S. R. M. Oliveira and O. R. Zaane, Privacy preserving frequent itemset mining, *IEEE International Conference on Privacy, Security and Data Mining*, (2002), 43–54.
36. N. Srinivas and K. Deb, Multiobjective optimization using nondominated sorting in genetic algorithms, *Evol. Comput.*, **2** (1994), 221–248.
37. V. S. Verykios, E. Bertino, I. N. Fovino, et al., State-of-the-art in privacy preserving data mining, *ACM SIGMOD Record*, **33** (2004), 50–57.
38. 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*, **19** (2007), 29–42.

39. T. Y. Wu, C. M. Chen, K. H. Wang, et al., A provably secure certificateless public key encryption with keyword search, *J. Chin. Inst. Eng.*, (2019), DOI:10.1080/02533839.2018.1537807.

40. E. Zitzler and L. Thiele, Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach, *IEEE T. Evolut. Comput.*, **3** (1994), 257–271.

AIMS Press