A Deep Q-Learning Sanitization Approach for Privacy Preserving Data Mining

Usman Ahmed
Dept. of Computer Science, Electrical Engineering and Mathematical Sciences
Western Norway University of Applied Sciences
Bergen, Norway
usman.ahmed@hvl.no

Jerry Chun-Wei Lin
Dept. of Computer Science, Electrical Engineering and Mathematical Sciences
Western Norway University of Applied Sciences
Bergen, Norway
jerrylin@ieee.org

Gautam Srivastava
Department of Mathematics and Computer Science
Brandon University
Brandon, Canada
srivastavag@brandonu.ca

Youcef Djenouri
SINTEF
Oslo, Norway
Youcef.Djenouri@sintef.no

ABSTRACT

With the establishment of the 5G network, a number of data-intensive applications will be developed. Privacy of information over the network is increasingly relevant, and require protection. The privacy of information while utilizing data is a trade-off that needs to be addressed. In this paper, we propose data privacy of 5G connected devices over heterogeneous networks (5G-Hetnets). A deep Q learning (DQL) based technique is applied to sensitize sensitive information from a given database while keeping the balance between privacy protection and knowledge discovery during the sanitization process. It takes transaction states as input and results in state and action pair. The DQL discovers the transactions dynamically, then the sanitization operation hide the sensitive information by minimizing side effects. The proposed approach shows significant improvement of performance compared to greedy and meta-heuristics and heuristics approaches.

CCS CONCEPTS

• **Security and privacy** → **Data anonymization and sanitization**; **Domain-specific security and privacy architectures**; **Web application security**; **Social network security and privacy**.

ACM Reference Format:
Usman Ahmed, Jerry Chun-Wei Lin, Gautam Srivastava, and Youcef Djenouri. 2021. A Deep Q-Learning Sanitization Approach for Privacy Preserving Data Mining. In *Adjunct Proceedings of the 2021 International Conference on Distributed Computing and Networking (ICDCN '21 Adjunct), January 5–8, 2021, Nara, Japan.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3427477.3429990

1 INTRODUCTION

In Internet of Things (IoT) environment, Fifth-generation mobile networks (5G) are currently being deployed and likely to replace 4G in advanced countries. The higher bit rates enable the use of software-defined networking (SDN) techniques while providing faster data transfer rates. The more capacity and very low latency allow applications designed for Internet of Things (IoT) networks to connect with data centers, that has lead to a fully mobile and connected society which has been the goal all along. It is obvious that a high level of security would also be essential for secure mobile data transmission from devices globally [15]. The mobile nodes with personal information can be tracked down and become vulnerable for the attackers like eavesdropping, the man in the middle, denial of service, replay and repudiation attack. The higher speed required a high level of QoS in terms of execution time, while a higher volume of data is transferred. The deployment of data-intensive devices to 5G heterogeneous networks requires addressing data privacy and security in a more serious manner. Sensitive data is associated with several applications. For example, data concerning customer purchases and their location. In this scenario, the edge computing network is required to carry out some sensitive tasks without transmitting some data to the central node, i.e. cloud server. However, instead deciding to hide sensitive information locally.

Due to these prevalent issues, privacy-preserving data mining, known as PPDM, has become a critical issue in recent times. The main goal of PPDM is to sensitize sensitive information while permitting required knowledge discovery from the databases. The most common way of sanitizing sensitive information is through addition or deletion operations performed directly on the database. Several heuristic-based methods have been proposed using the approach as mentioned above. Since the root problem in NP-hard, only heuristics have been proposed thus far that can work well in practice. However, since these approaches are "heuristic", they tend to work well in practice but can introduce side effects. Some common side effects may include introducing new artificial patterns and/or hiding non-sensitive patterns. Thereby, selecting what can be considered as an adequate set of sanitization operations for perturbation of the original database to protect sensitive information

ICDCN '21 Adjunct, January 5–8, 2021, Nara, Japan

ACM ISBN 978-1-4503-8184-0…$15.00
https://doi.org/10.1145/3427477.3429990

while simultaneously handling the task of side effect minimization is an *NP*-hard problem.

Despite the great success of these new technologies, they have never been used to deal with the PPDM issue. In this paper, a novel approach based on the Deep Q-Learning framework is used to deal with the PPDM issue. A dynamic approach that hides sensitive information by making use of the deletion operation and keeps a balance between privacy protection and knowledge discovery. Deep Q-Learning is a blend of deep learning, especially deep convolution neural networks and reinforcement learning in particular Q-learning [25]. This blend of two methodologies enables Q-learning for suitability for large data as well as practicality for real-world use cases [26]. In this, we use a deletion operation based perturbation strategy to hide sensitive information dynamically choosing the number of transactions to perturb instead of pre-defined statically. The key contributions of the proposed algorithm are mentioned here as follows:

- The develop Deep Q-Learning approach optimize the sanitize sensitive information.
- The propose approach discovers the transactions dynamically to be perturbed for information hiding instead of pre-defined.

2 RELATED WORK

With the extensive growth of data available from 5G networks, coupled with progressively developed data mining technologies used for discovering implicit information, we should see a large amount of growth of extracted knowledge from data that can be used by managers or decision-makers. However, this excess data also opens another front in the privacy protection domain. As data mining technologies mine implicit information from data which may reveal sensitive information as well, this leads to the potential security risks involved in sharing data between different business stakeholders. Thereby, PPDM has slowly risen as being identified as a issue of critical importance when the sanitization sensitive information is at stake keeping in mind the desire to minimize side effects. Furthermore, there is also a desire to hide sensitive information while keeping the database as undisturbed as possible. Agrawal *et al.* presented a quantitative metric that can be used as an evaluation tool for utility inside PPDM methods [1]. Verykios *et al.* presented a classification system that was hierarchical of PPDM techniques [27]. Liu et *al.* then considered the multi-party scenario to secure the private information in the distributed environment [24].

Clifton invented a tool-kit that was accompanied by a group of techniques for specific instances of databases requiring PPDM [6]. Dehkordi *et al.* designed three multi-objective GAs based methods that hide sensitive mining rules by using a removal technique that partially removes database items and considers the modified database for evaluation purposes [8]. However, missing rules and artificial data arises as a result of the sanitization process.

Data mining techniques [10, 11, 22] are designed and developed to extract potential knowledge. The mining methods help to find more information than possible using conventional information retrieval techniques. The methods also mine the sensitive and confidential information on databases. The sensitive information includes private, and personal information can also be mine by using data mining methods. PPDM [20, 28] has been in an issue in recent years in a variety of a domain such as health and medicine, intelligent homes, smart transportation networks, and adaptive city infrastructures. PPDM can mine the relevant information while preserving the anonymity of the sensitive information. Agrawal and Srikant evaluate the data distribution by using the reconstruction method. The method extracts the information using the classification method and then use that information to sanitized it [2]. Verykios *et al.* proposed a hierarchical classification that mines the information by considering the privacy constraints [27]. Another hamming distance approach was developed by Dasseni *et al.* [7]. The privacy of sensitive information is achieved by decrease the support and confidence for sensitive information. Islam and Brankovic [14] proposed a model to hide individual privacy. The method synthesizes the data by adding distinct noise into it. However, the addition does not affect the quality of data. Hong *et al.* [13] proposed SIF-IDF method to sanitized the database. The method computes the weight for each transaction and then sort to achieve sanitization. Lin *et al.* proposed different algorithms for PPDM and privacy-preserving utility mining (PPUM) [17, 18, 21]. However, PPDM is known as NP-hard problem [17, 19]; thus, the learning methods should find the nearly optimized solutions. The privacy preserve mining protocol is proposed by Han and Ng [12]. The GA based method [12] extract better rules by hiding sensitive information. Lin *et al.* proposed the different algorithms i.e., sGA2DT, pGA2DT and cpGA2DT [17, 18, 21]. To test the side effects of data sanitization, Lin *et al.* designed three distinct functions that balance the trade-off between data utility and hiding of sensitive information [17, 19]. Novel approaches are developed that first extract the relevant transaction and then the optimal select set of the transaction to achieve sanitation [12]. EMO algorithm [5] is also developed to delete the set of the transaction to achieve the sanitization. Cheng *et al.* [5] formulated non-sensitive rules, ghost rules and data loss as optimization objectives. Then produce multiple hiding solutions in a single run. The method able to give opportunity for a user to choose the discovered rule freely. Another NSGAII-based approach [23] efficiently delete records to sanitize the database. The PSO based multi-objective models [20, 28] were also studied to optimize the sanitization for record deletions. Lin *et al.* then presented a PPSF library [16] to collect the state-of-the-art algorithms for hiding the sensitive information in PPDM.

The supervised machine learning techniques are often employed to discover patterns [3]; however, the sensitive information also extracted; as a result, it violates privacy [3]. Bost *et al.* [4] proposed a framework to efficient privacy-preserving protocols for common classifiers, i.e. hyperplane decision, Naive Bayes and decision tree.

3 METHODOLOGY

The designed framework is easily incorporated into an environment where data collaboration and sharing can become a crucial problem.The data input can come from the warehouse or IoT environment and then it is prepossessed to structure, clean, and storing it. The source data can contain the application of transaction data, social media, smart healthcare environments, intelligent homes, smart transportation networks, and adaptive city infrastructures. After analysis, the pattern analysis is done by using the deep Q learning environment mentioned in Figure 1 and section 3.2. Then after sanitization objective archives, the sanitized data is shared with the business collaborator. The raw data of reports findings both can be shared with any collaborator. After that collaborator uses it for further understanding. The details of the algorithm are discuss section below.

3.1 Problem statement

In this paper, the sanitation process is modeled as a Markov Decision Process (MDP), as shown in Figure 1. The details are specified as follows:

- **State:** $s = [p, h, b]$: is defined as a set that is known to include information of transactions (candidates to be delete) $p \in \mathbb{R}_+^D$, the cost to delete the set of transaction $h \in \mathbb{Z}_+^D$, and the remaining transaction after sanitization process $b \in \mathbb{R}_+$, where D is the number of transaction in the projected datasets and $\mathbb{Z}_+$ denotes non-negative integer numbers.
- **Action:** a set of action on s. If the action is deletion then it leads to the union of transactions in $s_{t+1}$and s_t. If action is not delete then, the union operation will not be performed. The action will result in increasing and decreasing of the fitness values in equation 1, where α is the ratio of hiding of sensitive itemset before and after sanitization, β is defined as the # of FIs (frequent itemsets) before as well as after sanitization; and δ is defined as the # of FIs that are in the sanitized database D' that were previously infrequent in the original database D, w_1, w_2 and w_3 are relative importance of each side effect, which can be set at runtime by the user. In our experiment, we use $w_1 = 0.5, w_2 = 0.25$ and $w_3 = 0.25$

$$\text{fitness } (s) = w_1 \times a + w_2 \times \beta + w_3 \times \delta \quad (1)$$

- **Policy:** $\pi(s)$: the method to delete or not delete state s. The probability distribution of a at state s is essentially the policy.
- **Action-value function:** $Q_\pi(s, a)$: the reward achieved by following policy π of action a at state s.
- **Action:** The subscript t is used to denote the time and available action are as follows:
  - **Deletion:** $k = 1$, the set of action in current State that will be deleted. The next state and previous state transaction will be union to make new state the combination of previous and current state.
  - **Not Deletion:** $k = 0$ and no change will be made.
- **Reward:** $r(s, a, s')$: this can be defined as the change in value (fitness) that occurs when action a is taken at state s and arriving at the new state s'. In our policy, if action is delete then fitness value is calculated. If the fitness value decreases, the reward will be 10 else -10.

If we follow the Bellman Equation that was originally given in [25], the reward of action a_t is by expectation $r(s_t, a_t, s_{t+1})$, and the reward of the next state s_{t+1} [25]. The discounted factor γ is return based on assumption shown in Equation 2.

$$\begin{aligned} Q_\pi(s_t, a_t) = \mathbb{E}_{s_{t+1}} [r(s_t, a_t, s_{t+1})] \\ + \gamma \mathbb{E}_{a_{t+1} \sim \pi(s_{t+1})} [Q_\pi(s_{t+1}, a_{t+1})] \end{aligned} \quad (2)$$

3.2 Minimizing fitness as target goal

The goal of the method is to minimize the fitness value at a target time t_f. The Markov property of the model, optimizing policy minimizing the function $Q_\pi(s_t, a_t)$. The policy is learnt by optimizing the fitness value and interacting with environment. We employ the q table based reinforcement learning (RL) method and deep reinforcement learning (DRL) approach to solve this problem.

3.3 Population selection based on the multi threshold frequent itemsets

Previously in PPDM domain, Apriori based algorithm is used to identified frequent item-sets, and a fixed minimum threshold is set to extract the patterns. However, the pattern length should be in dynamic in nature as different attribute posses different contextual information. The contextual information is highly correlated with the process data instance. In case of a health care environment, in the population of 3000, it not easy to identify the user that occasional flu, i.e. the number of the patient having flu issues. It is not easy to identify patients with a high minimum threshold count for the short pattern. In case of longer pattern, more attributes should be introduced and added in the record, i.e. the district name or street number. It is easier to identify the patient with the same minimum threshold value for the short pattern. In order to solve the problem, the multiple support threshold is used to mine frequent item-sets. The minimum threshold for different length pattern is set with the function defined in the equation. The longer patterns should be set a lower threshold value. The threshold value is defined by the user based on preference or experience. Note that this threshold function can be manually defined by user's preference or experience. The equation 3, assign the threshold value $\delta(n)$ for the varied size (n) of the pattern.

$$\begin{aligned} f(n) &= \frac{\delta(1)}{N(1)} \times N(n) \\ \delta(n) &= \begin{cases} f(n), & f(n) > \delta_m \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (3)$$

The mean value for the distribution function is set as 1, thus $N(n) < N(m)$ if $m < n$ where n is the parameter and $N(n)$ is the value of normal distribution function. The function in equation objective is to make the sure that the longer patterns has a lower threshold value and also ensure that thresholds are larger than δ_m.

3.4 Deep reinforcement learning (DRL)

We are using 4 layer feed forward neural network for deletion or not deletion of the set of transactions. The input layer contains 64 neurons, 32 neurons layer and 8 neuron layer [25, 26]. Next, dense hidden layer contains advanced activation function known as parametric rectified is regular rectifier linear unit. The last hidden layer composed of sigmoid function as details are mentioned in the Table 1. The primary reason for using deep neural network instead of something else is that deep architectures are more efficient and in our case, complex hidden patterns can not be explored by anything

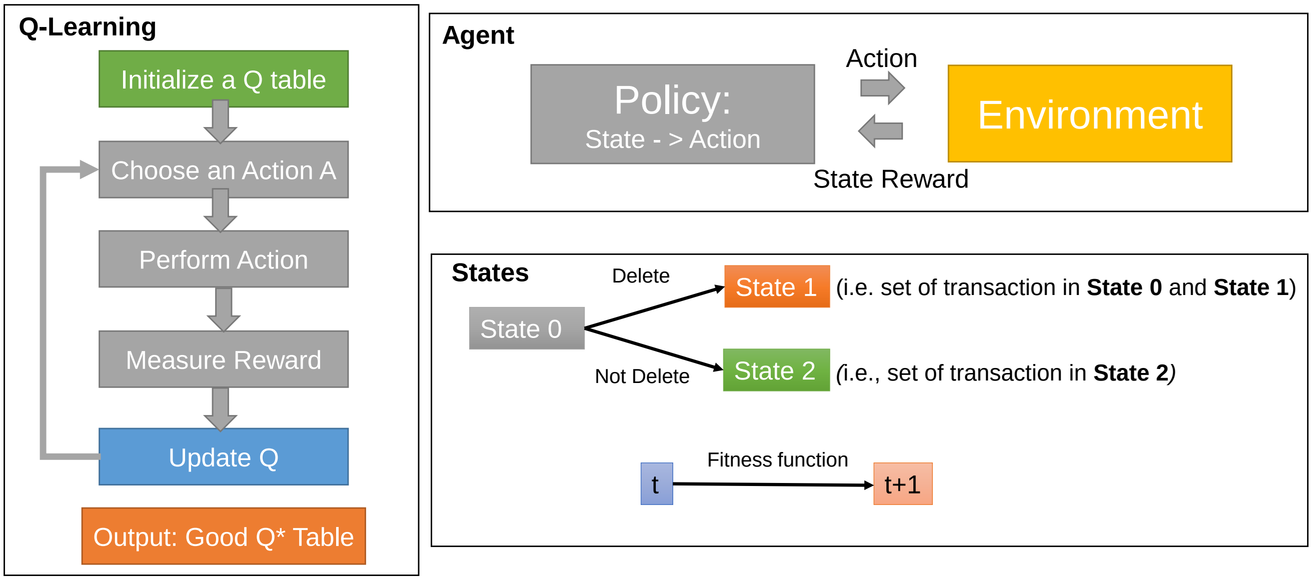

Figure 1: A DQL frame for data sanitization

Table 1: The architecture of neural network and hyper parameters.

Hidden Layer	4
Hidden Neuron Structure	64-32-4
Initial Parameter	
Activation function (Hidden Layer)	Parametric rectified unit
Activation function (Output Layer)	Softmax
Weight initialization	Uniform
Batch size	10
Epochs	100
Learning rate	0.001
Beta 1	0.9
Beta 2	0.999
Epsilon	1.00E-08

else. Our goal was to increase articulateness of the network, while maintaining an amenable size network that can be quickly trained.

4 EXPERIMENT AND RESULTS

In this section, we compared the proposed RL and DRL model with state of the evolutionary algorithm based methods, i.e. sGA2DT, pGA2DT, cpGA2DT, PSO2DT and Greedy sanitization algorithm [16]. The experiments were carried out on a Windows 10 PC with `AMD Ryzen 5 PRO 3500U` processor and 16 GB of RAM. We used two datasets, i.e., chess and T10I4D100K, for the experiments. All these datasets are available on the SPMF data mining library [9]. For experimentation purposes, we used the state size 100, support threshold and sensitive itemset selection of different dataset are also set according to dataset property. The number of episodes is set to 100; however, the increase in the episode could result in more improved results for both proposed algorithms. We compare the methods using the fitness values equation one and hiding factor (how good an algorithm is in hiding sensitive information).

4.1 Hiding factor analysis

In general, DRL performs better than the RL and evolutionary-based models as seen in Figure 2. In the case of chess dataset, the proposed model effectively hides sensitive itemsets by large percentage improvement. However, in case T10I4D100K (dense dataset), the proposed model able to hide the sensitive itemset successfully. It is concluded that the dataset that has more sparsity in case of unique itemsets, then RL methods required more training time.

4.2 Fitness value analysis

The fitness value is mentioned in Equation 1. The model having lower fitness values means it has able to achieve all the objective successfully. In fitness analysis, RL and DRL method performs better for the chess dataset as mentioned in Figure 3. However, not able to perform better for the large datasets. By nature, RL methods required hours and hours interaction with the environment to learn the complex patterns of the environment. In our case, we train the model for limited number states and episodes. The model required more training to optimize all the objective. However, the proposed framework is able to hide the sensitive items but in cost of non-sensitive itemsets too. In future, we explore the method to set the best states and episodes for the RL methods.

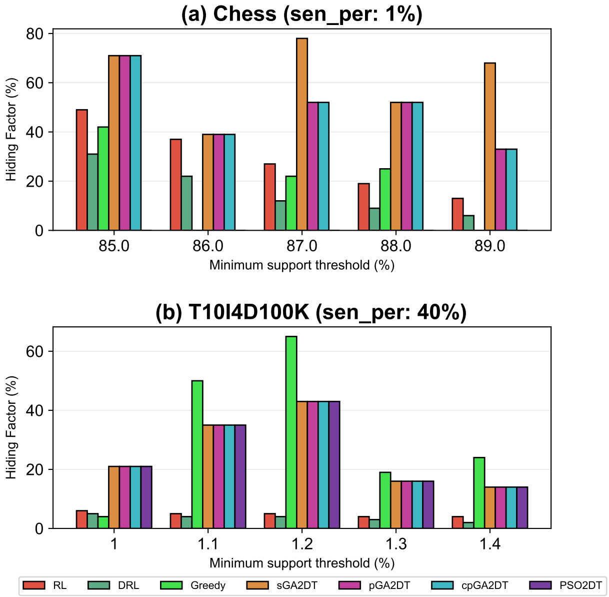

Figure 2: Method comparison in terms of hiding factor obtained for various minimum support threshold values and sens_per: *percentage of selected sensitive itemsets*

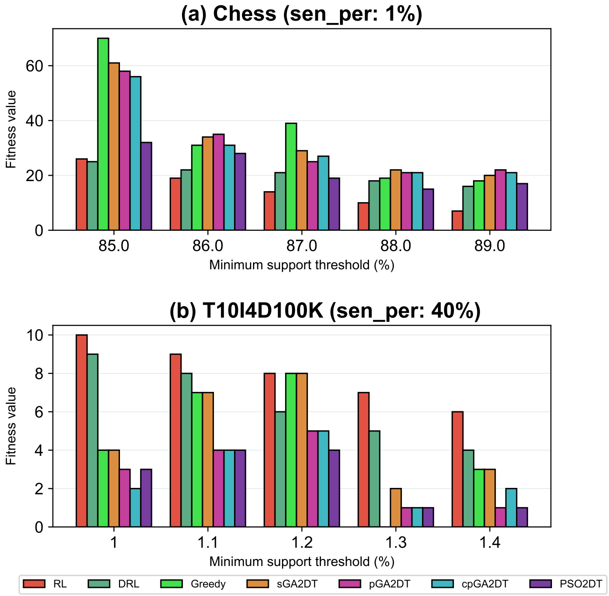

Figure 3: Method comparison in terms of Fitness values obtained for various minimum support threshold values and sens_per: *percentage of selected sensitive itemsets*

5 CONCLUSION

In this paper, a deep Q-Learning based DRL approach is proposed to sensitize sensitive information through transaction deletion. The number of transactions to be deleted are dynamically calculated instead of pre-calculated. A neural network-based value optimization function in DRL enables the proposed approach suitable for dealing with a large database. The proposed approach considers minimizing the general side effects of the sanitization process and to keep the balance between privacy protection and knowledge discovery. Experiments are conducted to show that the proposed RL and DRL technique outperforms the state of the art meta-heuristic and greedy techniques considering sanitization process side effects.

REFERENCES

[1] Charu C. Aggarwal, Jian Pei, and Bo Zhang. 2006. On privacy preservation against adversarial data mining. In *Proceedings of the Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 510–516.

[2] Rakesh Agrawal and Ramakrishnan Srikant. 2000. Privacy-Preserving Data Mining. In *Proceedings of the 2000 ACM SIGMOD International Conference onManagement of Data*. ACM, 439–450.

[3] Waleed S. Alnumay, Uttam Ghosh, and Pushpita Chatterjee. 2019. A Trust-Based Predictive Model for Mobile Ad Hoc Network in Internet of Things. *Sensors* 19, 6 (2019), 1467.

[4] Raphael Bost, Raluca Ada Popa, Stephen Tu, and Shafi Goldwasser. 2015. Machine Learning Classification over Encrypted Data. In *22nd Annual Network and Distributed System Security Symposium, NDSS*. The Internet Society.

[5] Peng Cheng, Ivan Lee, Jerry Chun-Wei Lin, and Jeng-Shyang Pan. 2016. Association rule hiding based on evolutionary multi-objective optimization. *Intelligent Data Analysis* 20 (2016), 495–514.

[6] Chris Clifton, Murat Kantarcioglu, Jaideep Vaidya, Xiaodong Lin, and Michael Y Zhu. 2002. Tools for privacy preserving distributed data mining. *ACM Sigkdd Explorations Newsletter* 4, 2 (2002), 28–34.

[7] Elena Dasseni, Vassilios S. Verykios, Ahmed K. Elmagarmid, and Elisa Bertino. 2001. Hiding Association Rules by Using Confidence and Support. In *Information Hiding, International Workshop, IHW (Lecture Notes in Computer Science)*, Vol. 2137. Springer, 369–383.

[8] Mohammad Naderi Dehkordi, Kambiz Badie, and Ahmad Khadem Zadeh. 2009. A Novel Method for Privacy Preserving in Association Rule Mining Based on Genetic Algorithms. *Journal of Software* 4, 6 (2009), 555–562.

[9] Philippe Fournier-Viger, Jerry Chun-Wei Lin, Antonio Gomariz, Ted Gueniche, Azadeh Soltani, Zhihong Deng, and Hoang Thanh Lam. 2016. The SPMF Open-Source Data Mining Library Version 2. In *Machine Learning and Knowledge Discovery in Databases (Lecture Notes in Computer Science)*, Vol. 9853. Springer, 36–40.

[10] Wensheng Gan, Jerry Chun-Wei Lin, Han-Chieh Chao, and Justin Zhan. 2017. Data mining in distributed environment: a survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 7, 6 (2017).

[11] Wensheng Gan, Jerry Chun-Wei Lin, Philippe Fournier-Viger, Han-Chieh Chao, Tzung-Pei Hong, and Hamido Fujita. 2018. A survey of incremental high-utility itemset mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 8, 2 (2018).

[12] Shuguo Han and Wee Keong Ng. 2007. Privacy-Preserving Genetic Algorithms for Rule Discovery. In *International conference on data warehousing and knowledge discovery (Lecture Notes in Computer Science)*, Vol. 4654. Springer, 407–417.

[13] Tzung-Pei Hong, Jerry Chun-Wei Lin, Kuo-Tung Yang, and Shyue-Liang Wang. 2013. Using TF-IDF to hide sensitive itemsets. *Appl. Intell.* 38, 4 (2013), 502–510.

[14] Md Zahidul Islam and Ljiljana Brankovic. 2011. Privacy preserving data mining: A noise addition framework using a novel clustering technique. *Knowledge-Based Systems* 24, 8 (2011), 1214–1223.

[15] Rabia Khan, Pardeep Kumar, Dushantha Nalin K. Jayakody, and Madhusanka Liyanage. 2020. A Survey on Security and Privacy of 5G Technologies: Potential Solutions, Recent Advancements, and Future Directions. *IEEE Commun. Surv. Tutorials* 22, 1 (2020), 196–248.

[16] Jerry Chun-Wei Lin, Philippe Fournier-Viger, Lintai Wu, Wensheng Gan, Youcef Djenouri, and Ji Zhang. 2018. PPSF: An open-source privacy-preserving and security mining framework. In *International Conference on Data Mining Workshops*. IEEE, 1459–1463.

[17] Jerry Chun-Wei Lin, Tzung-Pei Hong, Kuo-Tung Yang, and Shyue-Liang Wang. 2015. The GA-based algorithms for optimizing hiding sensitive itemsets through transaction deletion. *Applied Intelligence* 42, 2 (2015), 210–230.

[18] Jerry Chun-Wei Lin, Tzung Pei Hong, and Hung Chuan Hsu. 2014. Reducing side effects of hiding sensitive itemsets in privacy preserving data mining. *The*

Scientific World Journal 2014 (2014).

[19] Jerry Chun-Wei Lin, Qiankun Liu, Philippe Fournier-Viger, Tzung-Pei Hong, Miroslav Voznák, and Justin Zhan. 2016. A sanitization approach for hiding sensitive itemsets based on particle swarm optimization. *Engineering Applications of Artificial Intelligence* 53 (2016), 1–18.

[20] Jerry Chun-Wei Lin, Jimmy Ming-Tai Wu, Philippe Fournier-Viger, Youcef Djenouri, Chun-Hao Chen, and Yuyu Zhang. 2019. A sanitization approach to secure shared data in an iot environment. *IEEE Access* 7 (2019), 25359–25368.

[21] Jerry Chun-Wei Lin, Tsu-Yang Wu, Philippe Fournier-Viger, Guo Lin, Justin Zhan, and Miroslav Voznák. 2016. Fast algorithms for hiding sensitive high-utility itemsets in privacy-preserving utility mining. *Engineering Applications of Artificial Intelligence* 55 (2016), 269–284.

[22] Jerry Chun-Wei Lin, Lu Yang, Philippe Fournier-Viger, and Tzung-Pei Hong. 2019. Mining of skyline patterns by considering both frequent and utility constraints. *Engineering Applications of Artificial Intelligence* 77 (2019), 229–238.

[23] Jerry Chun-Wei Lin, Yuyu Zhang, Binbin Zhang, Philippe Fournier-Viger, and Youcef Djenouri. 2019. Hiding sensitive itemsets with multiple objective optimization. *Soft Computing* 23, 23 (2019), 12779–12797.

[24] Jun Liu, Yuan Tian, Yu Zhou, Yang Xiao, and Nirwan Ansari. 2020. Privacy preserving distributed data mining based on secure multi-party computation. *Comput. Commun.* 153 (2020), 208–216.

[25] Thanh Thi Nguyen, Ngoc Duy Nguyen, and Saeid Nahavandi. 2020. Deep Reinforcement Learning for Multiagent Systems: A Review of Challenges, Solutions, and Applications. *IEEE Trans. Cybern.* 50, 9 (2020), 3826–3839.

[26] Weiyu Si, Jinke Li, Peng Ding, and Ruonan Rao. 2017. A Multi-objective Deep Reinforcement Learning Approach for Stock Index Future's Intraday Trading. In *10th International Symposium on Computational Intelligence and Design ISCID*. IEEE, 431–436.

[27] Vassilios S. Verykios, Elisa Bertino, Igor Nai Fovino, Loredana Parasiliti Provenza, Yücel Saygin, and Yannis Theodoridis. 2004. State-of-the-art in privacy preserving data mining. *SIGMOD Rec.* 33, 1 (2004), 50–57.

[28] Tsu-Yang Wu, Jerry Chun-Wei Lin Lin, Yuyu Zhang, and Chun-Hao Chen. 2019. A grid-based swarm intelligence algorithm for privacy-preserving data mining. *Applied Sciences* 9, 4 (2019), 774.