ACSIJ Advances in Computer Science: an International Journal, Vol. 4, Issue 2, No.14 , March 2015
ISSN : 2322-5157
www.ACSIJ.org

ACSIJ
WWW.ACSIJ.ORG

A survey on privacy preserving association rule mining

Narges Jamshidian Ghalehsefidi[1], Mohammad Naderi Dehkordi[2]

[1,2] Department of Computer Engineering, Najafabad Branch, Islamic Azad University, Najafabad, Iran
[1]jamshidian_n@sco.iaun.ac.ir, [2]naderi@iaun.ac.ir

Abstract

By developing information technology and production methods and collecting data, a great amount of data is daily being collected in commercial, medical databases. Some of this information is important with respect to competition concept in organizations and individual misuses. Nowadays in order to mine knowledge among a great amount of data, data mining tools are used. In order to protect information, fast processing and preventing from revealing private data to keep privacy is presented in data mining. In this article, some techniques in preserving privacy of association rule mining are introduced and some hiding algorithms of association rules are evaluated.

Keywords: Association rule, Data mining, Data privacy, Privacy preserving.

1. Introduction

By developing data mining and discovering association rules, two controversial topics are presented. On the one hand data mining is able to analyze a great deal of data in the minimum time, and on the other hand, extreme processing of intelligent algorithms is originated from secret and confidential data in databases. Discovered knowledge by applying various techniques may contain user's or company's private data. Revealing each kind of private information is threatening for that company or the user's security. For instance, by sharing a medical database, the patient's personal information such as names, zip codes … are being threatened and must be preserved. The purpose of privacy preserving data mining is to develop the algorithms that alter the basic data in a way that private data and private knowledge will not be discovered after processing data mining. Nowadays, individuals and companies show their interests to share information accompanied by privacy preserving. Sharing information may be useful for companies and individuals but if confidential information is extracted from shared data, it will decrease company's benefits and threaten it.

To get rid of this danger, private data must be hidden prior to sharing or distributing of databases, to keep them safe from any unauthorized access. Even by privacy preserving data algorithms, sharing data is not completely secure as by using some non-sensitive data, sensitive ones can be inferred. In addition to discovering sensitive information by inferring, there are side effects created by hiding sensitive information for the shared database. Side effects such as missing non sensitive information, changing the size of database, and making new information for sensitive databases such as medical database, are dangerous. Side effect is harmful from one point of view and useful from another. The first approach is to preserve the security of database. If non sensitive information that is related to sensitive information is missed due to hiding, inferring sensitive information will be very difficult and impossible and by making new information the capacity of mined data increases and inferring sensitive information would be difficult. Another approach is the authenticity of database. If any side effect occurs in database, the authenticity of the data will decline and it is very essential for the user. There must be a balance between these two approaches.

1.1 Definition of data mining

Increasing the amount of data has made new opportunities to work upon engineering and trading. Data mining and discovering knowledge have been emerged in engineering and computer science as a new scientific major. Various definitions have been presented in different sources but the most common definition that has been mentioned in most resources is mining information and knowledge and discovering hidden patterns from giant and complicated databases [1]. The steps to discover knowledge in data mining are as follow:

1-Data cleaning: in this step inconsistent data is removed. It takes 60% of data mining time.

2-Data integration: data are usually collected and combined from different sources. They must be in a way to perform better data mining.

3-The purpose of data selection: where data relevant to the analysis task are retrieved from the database.

4-Data transformation: data are transformed in a way that is appropriate for data mining operation.

5-Selecting data mining operation (Classification, clustering and so on), and data mining method (decision tree, neurotic networks and so on).

6-Data mining: a process by which intended patterns are extracted from data.

7-Item pattern evaluation: analyzing obtained patterns and eliminating inappropriate ones.

8-Knowledge presentation: presenting mined knowledge to users.

There are different kinds of data mining methods that all of them concerned with surveying raw data and putting them in a special pattern. Data mining techniques consist

of prediction (classification, regression, time series analysis) and description (clustering, association rules, and sequential patterns). In this article we discuss about privacy preserving in association rule mining. Data mining work range is very extended and used in most of real environments. Some of data mining applications in real environments are banking in predicting swindling ways by credit cards, determining fixed customers, and medical environments in assessing success for medical treatments for chronic diseases and the kind of behavior with patients.

2. Association rules mining

Association rules mining determine the kind of relation within input data. These rules are determined by supporting and confidential factors. In this section some of basic concepts in privacy preserving of data mining are concerned. Association rule mining was presented by [2]. Imagine that I={i1,i2,…,im} consists of a group of elements and database D={T1,T2,...,Tn} consists of a group of transactions. Every transaction of T∈ D consists of I sub category. The whole template of association rule is x→y If x and y are a sub category of I and x ∩ y = ∅ . So x is called antecedent or LHS (Left Hand Side) of rule and y is called consequent or RHS (Right Hand Side).
Supporting a rule for x→y is defined by proportion of simultaneous repetition of x and y over total transactions. Equation 1 shows the way of rule support calculation.

$$\text{support}(X \rightarrow Y) = \frac{|X \cup Y|}{|D|} \quad (1)$$

Confidence of a rule for x→y is also defined by proportion of simultaneous repetition of x and y over the number of repetition for x. Equation 2 shows the way of rule for confidence calculation.

$$\text{confidence}\ (X \rightarrow Y) = \frac{|X \cup Y|}{|X|} \quad (2)$$

There are different criteria to evaluate the degree of authenticity and the value of presented rules that beneficial and effective rules can be selected based on them among a large group of rules. The most applicable ones are minimum support threshold (MST) and minimum confidence threshold (MCT). If support (x→y)>=MST and confidence (x→y)>=MCT them x→y is valuable and mined from database during data mining.
By data mining repeated patterns can be extracted among raw data. Repeated patterns are a combination of data that are repeatedly observed in transactions. To find these patterns some standard algorithms such as Apriori algorithm, Eclat algorithm and others can be used. These algorithms mine association rules in two levels. In the first step, repeating patterns that their support is larger or equal to MST and then according to extracted patterns mines the rules that their confidence is larger or equal to determined MCT.

3. Techniques for privacy preserving association rule mining

The reason for hiding association rule methods is to sanitize the original database to achieve the following purposes:
1-A rule that is considered sensitive from the owner of database view and can be extracted from an original database must not be extracted from sanitized database. In fact, it must not be equal to or larger than MST or MCT. It is remarkable to note that if the degree of support from one rule declines under MST degree, there is no need to calculate the rule confidence and it can be said that the rule is hidden otherwise the confidence degree of rule must be calculated and if it declines under MCT degree, that rule is hidden. It can be said that most of hidden algorithms of association rules may be failed under particular circumstances unless a kind of solution such as adding new transaction to original database for security reasons is considered.
2-All non-sensitive rules that are mined from original database are also mined from sanitized database. In fact, no non sensitive rule is lost.
3-Except for non-sensitive rules, if a new type of rule is not mined from original database, it cannot be mined from sanitized database either. In fact, no ghost rule is produced.
Those solutions that consist of three purposes are called exact. An accurate hiding cause the least change in original database is ideal. Those inaccurate solutions that make hiding are called approximate.
The factors to evaluate algorithms of privacy preserving association rules mining are: 1- Hiding failure. 2- Degree of dissimilarity between original database and sanitized database. 3- Degree of lost data. 4- Algorithm efficiency for large databases. 5- Authenticity of information in database. 6- Time of performance (the time that is required for algorithm to hide). 7- The degree of inaccurate rules of hiding association rules are surveyed in two major aspects.
The first kind of hiding is based on selective rules as a kind of sensitive rule and the second kind is for frequent hiding patterns that are sensitive from the owner of database point of view.
Typical methods that are used for hiding sensitive rules are: 1- Heuristic approach 2- Border based approach 3- Exact approach. Heuristic approach makes security for a group of determined transactions. Border based approach with two positive and negative elements tries to hide sensitive rules by eliminating the elements in negative group and by preserving elements in positive group. Exact approach is a non-heuristic algorithm that makes hiding possible by integer programming or linear programming. Imagine that

D is an original database R is a group of mined rules from D, and RH consists of a group of sensitive rules that exists in R. privacy preserving association rule mining of database change D to D' in a way that all rules in R except RH are mined from D' database.
Considering studies conducted in hiding association rules privacy preserving association rules of algorithms can be divided into three major parts:

3.1 Techniques based on heuristic

Some developed techniques for data mining techniques such as Classification, clustering, association rules, by this hypothesis that sanitization is NP-hard for them, therefor for sophisticated issues, heuristic approach can be used. In these techniques a series of transactions are selected for making a secure area. In this method the time of performance and memory can be decreased by preprocessing. This technique involves the two following ways:
Association rules based on perturbation: In this method a value is replaced by a new value (such as changing 1 to 0 or adding noise). Therefore sensitive rule support declines so that the utility of changed database is greatly preserved. The utility level is measured by those non sensitive rules that are hidden through side effects of hiding process. In this method it's probable to lose non sensitive association rules and also make new rules if 1 is changed to 0 and vice versa. Due to side effects that are made by this method, it is not appropriate for sensitive programs such as medical programs.
Association rules based on blocking: in this method, for hiding sensitive rules, the value is replaced by a question mark or a right value. So, it declines the degree of support and the level of confidence for sensitive rules. This method is ideal for some specific programs such as medical programs. Finding support and confidence of a rule is difficult in this method. Supporting A is between minimum support and maximum support. The minimum support for A involves transaction percentages that contain an item and the maximum support for A involves transaction percentages that an item contains 1 or a question mark. The confidence level of rules for A→B is between maximum level of confidence rule maxsup (AB)*100/ minsup (A) and the minimum level of confidence rule minsup (AB)*100/ maxsup (A) [3].

3.2 Reconstruction based association rule

Some presented techniques in privacy preserving based on perturbing of data and reconstructing the distributions in aggregate level are made for mining purposes. In these algorithms, at the beginning changes are made in data and then database is reconstructed. This technique contains different methods such as numerical data, binary data and categorical data.

3.3 Cryptography based techniques

In this method data cryptography is used. Methods that are based on cryptography use secure computations. In multiple secure computation parties, parties like to compute some computations on their private inputs; nobody likes to reveal his output to anybody else. Everyone is only aware of his input and results. In [4] 4 secure computations are presented. These methods contain the secure sum, the secure union set, the secure size of intersection set and the scalar product. SMC (secure multi party computation) is mostly used in distributed environments and its aim is to guarantee the authenticity of computations and to protect participated parties in computations against revealing input and output data. In distributed environments, data are distributed in databases by horizontal and vertical distributions [5].

3.3.1 Horizontally partitioned distributed

To find multiple algorithms to guarantee that there is no leakage among inter site information. In this kind of partition transactions, within n database that each one belongs to a partner, are distributed. Generally the support degree of an item-set is obtained locally in each site by the sum of support degree of that item-set

3.3.2 Vertically partitioned Distributed

It computes the sum of support degree of every sub item-sets that have been securely distributed among different sites by the idea of "secure sum". An item-set will be selected as a global frequent item-set that its support degree is larger than MST.

4. Reviewing studies

Hiding sensitive association rules was presented for the first time by [2]. They hid sensitive rules through suggesting a lattice like graph by declining supporting degree of frequent item-sets.
Desseni et al. broadened hiding issue to combine sensitive hiding rules and a group of sensitive elements. They presented their strategies for hiding that was based on declining support or confidence rules and each time one rule was hidden. 1) The first strategy was based on increasing supporting elements on the left side of the rule (LHS). 2) The second strategy was based on declining the frequency level of element on the right side of the rule (RHS). 3) The third strategy was based on the total decline in frequent sensitive rules [6].
Zaïane and Oliveira were the first people who presented simultaneous hiding for several rules. They presented minFIA and maxFIA algorithms, that hid the rules regardless of the number of sensitive rules with two kinds

of scanning of the database. During the first scanning, the sensitive transactions were diagnosed and an index was pointed for them. In the second transaction security was done with the minimum number by deleting selection of single elements (MINFIA by selecting on element with the least supporting degree and MAXFIA by selecting an element with the most supporting degree) [7].

Verikios et al. presented two strategies for hiding sensitive rule by reducing support and confidence with 5 algorithms of 1.a, 1.b, 2.a, 2.b and 2.c. the first three algorithms were proposed with the aim of hiding sensitive rule by reducing support or confidence. The last two algorithms were proposed with the aim of hiding sensitive itemset by reducing their support [8].

Oliveira et al. presented an algorithm called DSA to protect sensitive knowledge before sharing. The aim of this algorithm is to block inference channels during censorship of data, and the effect of censorship is determined by side effect and recovery factor. This algorithm performs better in real databases and then it scans them. By using this algorithm, the database owner just starts sharing the patterns [9].

Lee et al. presented a new technique to hide sensitive patterns. In this method a new database that is secured will be obtained by multiplying matrix of sanitization over the main database. In this article three algorithms are presented to create matrix of sanitization and security where no new rule is made in all of them. The first algorithm called Hidden first hides sensitive patterns to -1 by setting matrix entries of matrix. In this algorithm some non-sensitive patterns are lost. The second algorithm called Non-Hidden first makes security in a way that non sensitive pattern is not lost. The third algorithm is the combination of two algorithms called NHF and HF that hides all sensitive patterns with the least amount of effect on non-sensitive patterns and controls these effects with one variable [10].

Menon et al. proposed two blanket and intelligent strategies for the integer programming algorithm which is an exact approach. Both strategies focus on selecting items to be removed. The blanket strategy loses more non-sensitive patterns. This algorithm aims to increase accuracy and reduce runtime [11].

Wang et al. presented two algorithms called DSR and ISL to hide association rules that there is no need for data mining and selecting of sensitive rules and just gets a group of sensitive items as an input and then Hiding will be performed by an algorithm. In both algorithms at first those rules that their sensitive items are located at their left sides are selected for security. The first algorithm called ISL declines the confidence level rule by increasing the degree of support for a group of elements located in the left side of the sensitive rule, and the second algorithm called DSR declines the degree of support for a group of elements located in the right side of the sensitive rule [12][13].

Wang et al. suggested two algorithms called DCIS and DCDS. The first algorithm increases the degree of support for a group of elements located in the left side of the rule and the second one declines the degree of support for a group of elements located in the right side of the rule to decrease the confidence level of the rule [14].

Amiri suggested three algorithms. In the first algorithm called aggregate supporting, sensitive rule is decreased by deleting some transactions. The second algorithm called disaggregate, declines supporting degree of sensitive rules by deleting some sensitive elements. The third algorithm called Hybrid determines the identified transactions through aggregate method and then specifies the required elements for deleting through disaggregate method [15].

Using the two techniques of data blocking and distortion, Verkios et al. hide sensitive rules. Using the data distortion technique, the WSDA algorithm selects and removes the best transaction and the victim item. By replacing a question mark (?), the BA algorithm reduces the confidence of sensitive rule to under the MCT-SM. The SM is specified by the user [16].

Duraiswamy suggested an algorithm called SRH that calculates the required number of transactions to hide sensitive rules by mincounting and selects those transactions that fully support that rule, then arrange transactions in the ascendant, finally deletes the right item from transactions. Due to clustering sensitive rules, time complexity will be decreased and updating will be performed only after hiding all rules. But this algorithm can only hide those rules that consist of antecedent and consequent single items [17].

Chandra et al. suggested an algorithm based on ISL. In this algorithm a Mconf(modified confidence) and a Msup(modified Support) and a hiding counter are used. Mconf and Msup are dependent on hiding counter variable and for each rule this variable exists. Each time an item is added to the selected rule until Mconf of the rule gets larger or equal to the least amount of confidence level. This algorithm does not have any side effect and this is just a theoretical method that is in primitive step [18].

Dehkordi et al. used the genetic algorithm for database sanitization. This algorithm uses three operators for Selection, Crossover and Mutation for creating generations. In each generation, using four evaluation functions presented in the paper, the best populations are selected and are used to create the next generation. Each evaluation function was proposed to reduce a side effect including reducing lost rule, reducing ghost rules, and reducing database changes [19].

Modi et al. introduced an algorithm called DSRRC that tries to hide rules at the same time and by the least amount of changes over the database through clustering rules according to common item RHS. The only disadvantage of

this algorithm is in hiding those rules that have just common RHS item [20].

Chandrakar et al. introduced an algorithm called Hybrid that was made of two strategies, one was used to delete the right item and the other was used to insert the left item. The former decreased the level of support and the latter increased the confidence level of a rule. The main purpose of this algorithm was to preserve to discover sensitive rules, but it did not consider the side effects [21].

Oliveira et al. suggested two algorithms that are performed by two kinds of scanning over database where indexing of transactions is done in the first scan and providing security is done in the second one. This algorithm has been proposed to decline the number of lost rules. In this algorithm revealing threshold is used that determines mining of association rules and makes an agreement between the lost rule and hidden rule. Suggested algorithm names are Round Robin and Random that in the first algorithm, selecting sensitive items is sequential and in the second one, selecting sensitive items is randomly done [22].

Kumar Jain et al. suggested a similar algorithm proposed by chanadrakar. The authors claimed that the number of changes in database and the time for hiding are marginal. Hiding failure in this algorithm is equal to zero [23].

Vijayarani and Prabha proposed a heuristic algorithm called ABC based on honey bee's movement to find the best nutrition source. In this algorithm the best transactions to delete sensitive rules are found through random selection of transactions and computing the probability of selected transactions and no beneficial association rule is lost [24].

Komal Shah et al. modified DSRRC algorithm and called it as ADSRRC. In this article the drawbacks of DSRRC were the dependency of making any change in transaction, and the arrangement of transactions in database. To solve the problem they claimed that transactions are arranged according to their descending level of sensitivity and length. Moreover DSRRC algorithm rearranges transactions according to their sensitivity after each change in transactions but in ADSRRC algorithm, arrangement of transactions is performed only once. In addition, an algorithm called RRLR is proposed in this article that can also hide those rules that contain multi elements on their right side. In this algorithm, to hide a sensitive rule both support and confidence are decreased [25].

Jain et al. proposed an algorithm that used distortion technique. In this algorithm it only alters the position of sensitive item that is located on the left side of the rule in a way that it doesn't change the degree of support for sensitive items and the size of database. The input of algorithm consists of sensitive items that selection of the rules that their left side involves sensitive item is made and their right side will be combined together. In addition to mentioned advantages of this algorithm, maximum hiding of rules with the minimum stages and declining the number of ghost rule are also involved as its advantages [26].

Gante et al. presented a framework that minimizes the side effects and hides sensitive rules after mining association rule and item sets, to select a kind of hierarchy structure for ISL, DSR or Hybrid algorithms. At this project, it was attempted to get MST and MCT automatically on the contrary to ISL and DSR algorithms that users determines their values. According to this framework, an algorithm that has the least side effect is selected and security is made according to it [27].

Gulwani suggested an algorithm that supporting sensitive item remains unchanged and the size of database won't change either. This algorithm needs less number of scanning for hiding and hides more rules compared with 1.a and ISLF algorithms. The input of algorithm is a sensitive item that selects all rules that involve. This item is either at the right side or at the left side .Security is made through deleting the left side item and inserting it in a transaction that supports sensitive items marginally [28].

Dutraj et al. presented an algorithm similar to that of chanadrakar that was based on two concepts: SMC (secure multiparty computation) and hiding association rules. In this article dataset is distributed over the network. In this algorithm trusted third party uses SMC model and is divided into three major parts based on it. The first party collects data security from each part. The second part is gathered to produce association rule, the third part hides association rules by hybrid algorithm that is a combination of ISL and DSR [29].

Radadiya et al. presented an algorithm called ADSRRC to remove DSRRC restrictions. This algorithm hides those rules that consist of several items at their left and right side [30].

Domadiya et al. proposed an algorithm called MDSRRC that doesn't have any restriction in the number of items located at their left or right side of the association rule. This algorithm fills the restriction for DSRRC and selects the best item for deletion based on its repetition on the right side of the rule. This algorithm is similar to algorithm proposed by Radadiya [31].

Using the Border-based approach, Moustakides and Verykios proposed the MaxMin algorithm. The purpose of this algorithm is to reduce lost itemsets. In this algorithm, the victim item is selected in two stages. At the first stage, the least frequent itemset is selected and then among the items, the selected itemset, the item with the highest frequency is selected [32].

Hong et al. proposed the SIF-IDF algorithm for hiding the sensitive itemset. In this algorithm, the best transaction is selected and the sensitive items will be removed. This algorithm focuses on selecting the best transaction, and for selecting the victim item, it considers only its degree of sensitivity. The algorithm execution time is high and the

order of the entry of sensitive itemsets for hiding affects the final result [33].

Lin et al. proposed the HMAU algorithm for hiding the sensitive itemset. In this algorithm, a suitable transaction is selected based on side effects, including hiding failure, lost itemsets and new itemsets for removal. The aim of transaction removal is to reduce the support of sensitive itemset [34].

Using the blocking technique, Saygin et al. proposed the CR, GIH and CR2 algorithms for hiding sensitive rules. The CR and CR2 algorithms were proposed with the aim of increasing the LHS of sensitive rule in order to reduce the rule confidence. In the CR, a question mark (?) replaces LHS and in the CR2, a question mark (?) replaces the lost item of LHS. The GIH algorithm reduces the support of sensitive rule by replacing the RHS item with a question mark (?) [3][35].

5. Evaluation

Diagnosing proper criteria in evaluation of algorithms and significant privacy preserving tools, and meeting all required criteria for an algorithm is a difficult task. Usually, a kind of balance must be made among required criteria and by considering the user's need some criteria act better. For instance, if we increase the degree of security level, the performance time will be increased. A primary list of criteria is presented [5]:

The performance: the suggested methods and algorithms are effective considering the performance time to secure database.

The data utility: after applying privacy preserving methods on database, the lost information and new information must be minimized.

The level of uncertainty: the hidden sensitive information must not be revealed by inferring non sensitive information.

The resistance: The privacy preserving algorithms are different from data mining techniques. To evaluate presented algorithm, data mining techniques that are different from techniques presented in the algorithm, is required. This parameter is also called transversal endurance.

Scalability: designed algorithms should be able to secure great database as well as their efficiency.

In this part, some algorithms mentioned in the previous part are evaluated. Evaluation of algorithms is usually done based on the number of lost rules, artificial rules, performance time, degree of performance change, and hiding failure. Imagine that Rh (D) is a sensitive rule and R (D) is a mined rule from major database rather than sensitive rule. Rh (D') will be sensitive rule and R (D') will be mined rule from cleaned database rather than sensitive rule.

Lost rule: Non-sensitive rules that will be lost due to hiding sensitive rule and are not available. By applying (3), this criterion is measured [36].

$$\text{lost rule}=\frac{|R(D)|-|R(D')\cap R(D)|}{|R(D)|} \quad (3)$$

Non sensitive rules are usually lost due to deleting an item. It is sometimes lost by inserting an item, as by inserting an item the degree of support increases and decreases the confidence level of those items that have this item on their left side.

Artificial rule: those rules that are not mined by support and confidence levels that are defined by users from the main database, but will be mined from cleaned database after security in made. By applying (4), this criterion is measured [36].

$$\text{Artificial rule}=\frac{|R(D')|-|R(D)\cap R(D')|}{|R(D')|} \quad (4)$$

Artificial rules are also made by deletion or insertion of an item. When sensitive item is deleted, its degree of support will be decreased and increases the degree of confidence level for the rule that has the item on its left side.

Dissimilarity: performed change is between main database and cleaned database. By applying (5), this criterion is measured. In this formula i stands for an item in the main database of D and fD(i) is its frequency in the database. fD'(i) is the frequency of an item in the cleaned database [36].

$$\text{Dissimilarity}=\frac{\sum_{i=1}^{n}|fD(i)-fD'(i)|}{\sum_{i=1}^{n}fD(i)} \quad (5)$$

Hiding failure: the degree of sensitive rules that are mined after applying security on cleaned database is measurable by (6) [36].

$$\text{Hiding Failure}=\frac{|R_h(D')|}{|R_h(D)|} \quad (6)$$

In table 1 side effects of some algorithms mentioned in the previous part are presented.

Table 1:Side effects of algorithms

Algorithm	Side effect		
	Lost rule	Artificial rule	Hiding Failure
[2]	√		
1.a	√	√	√
1.b	√		√
2.a	√		√
2.b	√	√	
2.c	√	√	
MinFIA	√		√
MaxFIA	√		√
Hidden-First	√		
Non-Hidden-First	√		√
HPCME			√
ISL	√	√	√
DSR	√		
DCIS		√	
DCDS	√	√	
WSDA	√	√	
BA	√	√	
SRH	√		
[21]	√	√	
Round Robin	√		√
Random	√		√
DSRRC	√		
ADSRRC	√		
RRLR	√		√
MDSRRC	√		

6. Conclusion

In this article privacy preserving techniques were introduced and discussed. Distortion and blocking techniques have been more concentrated on privacy preserving and have been more emphasized on hiding rules or preventing from making sensitive rules. These methods are simple and have many side effects. Side effects involve losing non sensitive rules, making ghost rules that are dangerous for sensitive database such as medical science and lead to failure in hiding. Another challenge in this issue is about inference sensitive rules by using non sensitive ones. In distributed methods security level mode by cryptography is high and efficiency will be decreased. Horizontal distributed method in a database is simple due to having records but in vertical distributed method, there is a possibility for information leakage from one part to another.

Many algorithms and methods have been recently presented for privacy preserving of data mining. However there is an opportunity for further study, research and development in this issue.

References

[1] J. Han and M. Kamber, "Introduction Data Mining: Concepts and Techniques", 2nd ed., CA: San Francisco, 2006, pp. 5–7.

[2] M. Atallah, E. Bertino, A. Elmagarmid, M. Ibrahim, V. Verykios, "Disclosure Limitation of Sensitive Rules", Knowledge and Data Engineering Exchange, 1999, pp. 45-52.

[3] Y. l. Saygin, V. S.Verykios, C. Clifton, "Using Unknowns to Prevent Discovery of Association Rules", SIGMOD Rec, Vol. 30, No. 4, 2001, pp. 45–54.

[4] C. Clifton, M. Kantarcioglou, X. Lin, M. Y. Zhu, "Tools for Privacy Preserving Distributed Data Mining", ACM SIGKDD Explorations, Vol. 4, No. 2, 2002, pp. 28–34.

[5] V. S. Verykios, E. Bertino, I. N. Fovino, L. P. Provenza, Y. Saygin, Y. Theodoridis, "State of the Art in Privacy Preserving Data Mining", SIGMOD Rec, Vol. 33, No.1, 2004, pp. 50-57.

[6] E. Dasseni, V. S. Verykios, A. K. Elmagarmid, E. Bertino, "Hiding Association Rules by Using Confidence and Support", IHW '01 Proceedings of the 4th International Workshop on Information Hiding, 2001, pp. 369-383.

[7] S. R. M. Oliveira, O. R. Zaïane, "Privacy Preserving Frequent Itemset Mining", Security and Data Mining, Vol.14, 2002, pp. 1–11.

[8] V. S. Verykios, A. K. Elmagarmid, E. Bertino, Y. Saygin, E. Dasseni, "Association rule hiding", IEEE Transactions on Knowledge and Data Engineering, Vol. 16, No. 4, 2004, pp. 434–347.

[9] S. R. M. Oliveira, O. R. Zaïane, Y. Saygin, "Secure Association Rule Sharing", Advances In Knowledge Discovery And Data Mining, Vol. 3056, 2004, pp. 74-85.

[10] G. Lee, C. Y. Chang, A. L. P. Chen, "Hiding sensitive patterns in association rules mining", Computer Software and Applications Conference, Vol. 1, 2004, pp. 424-429.

[11] S. Menon, S. Sarkar, S. Mukherjee, "Maximizing Accuracy of Shared Databases when Concealing Sensitive Patterns", Information System Research, Vol. 16, No. 3, 2005, pp. 256–570.

[12] Sh-L. Wang, A. Jafari, "Hiding Sensitive Predictive Association Rules", Systems Man and Cybernetics, Vol. 1, 2005, pp. 164 – 169.

[13] Sh-L. Wang, B. Parikh, A. Jafari, 'Hiding informative association rule sets', Expert Systems with Applications, Vol. 33, No. 2, 2007, pp. 316–323.

[14] Sh-L.Wang, D. Patel, A. Jafari, "Hiding Collaborative Recommendation Association Rules", Applied Intelligence, Vol. 27, No. 1, 2007, pp. 67-77.

[15] A. Amiri, "Dare to share: Protecting Sensitive Knowledge with Data Sanitization", Decision Support Systems, Vol. 43, No. 1, 2007, pp. 181–191.

[16] V. S. Verykios, E. D. Pontikakis, Y. Theodoridis, L. Chang, "Efficient algorithms for distortion and blocking techniques in association rule hiding", Distributed and Parallel Databases, Vol. 22, No. 1, 2007, pp. 85–104.

[17] K. Duraiswamy, D. Manjula, N. Maheswari, "A New Approach to Sensitive Rule Hiding", Computer and Information Science, Vol. 1, No. 3, 2008, pp. 107–111.

[18] B. R. Chandra, V. jitendra, A. K. Sohel, S. Anand, B. Mahua, "Hiding Sensitive Association Rules Efficiently By Introducing New Variable Hiding counter", Service

Operations and logistics and informatics, Vol. 1, 2008, pp. 130-134.
[19] M. N. Dehkordi, K. Badie, A. K. Zadeh, "A Novel Method for Privacy Preserving in Association Rule Mining Based on Genetic Algorithms", Journal Of Software, Vol. 4, No. 6, 2009, pp. 555–562.
[20] C. N. Modi, U. P. Rao, D. R. Patel, "Maintaining privacy and data quality in privacy preserving association rule mining", Computing Communication and Networking Technologies, 2010, pp. 1-6.
[21] I. Chandrakar, Y. U. Rani, M. Manasa, K. Renuka, "Hybrid Algorithm for Privacy Preserving Association Rule Mining", Journal of Computer Science, Vol. 6, No. 12, 2010, pp. 1494-1498.
[22] S. R. M. Oliveira, O. R. Zaïane, "Algorithms for Balancing Privacy and Knowledge Discovery in Association Rule Mining", Database Engineering and Applications Symposium, 2003, pp. 54-63.
[23] Y. K. Jain, V. K. Yadav, G. S. Panday, "An Efficient Association Rule Hiding Algorithm for Privacy Preserving Data Mining", International Journal on Computer Science and Engineering, Vol. 3, No. 7, 2011, pp. 2792-2798.
[24] S. Vijayarani, M.S. Prabha, "Association Rule Hiding using Artificial Bee Colony Algorithm", International Journal of Computer Applications, Vol. 33, No. 2, 2011, pp. 41–47.
[25] K.Shah, A.Thakkar, A.Ganatra, "Association Rule Hiding by Heuristic Approach to Reduce Side Effects & Hide Multiple R.H.S. Items", International Journal of Computer Applications, Vol. 45, No. 1, 2012, pp.1–7.
[26] D. Jain, A. Sinhal, N. Gupta, P. Narwariya, D. Saraswat, A. Pandey, "Hiding Sensitive Association Rules Without Altering the Support of Sensitive Item(S)", International Journal of Artificial Intelligence & Applications, Vol. 3, No. 2, 2012, pp. 75–84.
[27] D. Gatne, M. Jhade, "Privacy Preservation with Limited Side Effect", International Journal of Emerging Technology and Advanced Engineering, Vol. 2, No. 6, 2012, pp. 389–392.
[28] P. Gulwani, "Association Rule Hiding by Positions Swapping of Support and Confidence", International Journal Information Technology and Computer Science, Vol. 4, No. 4, 2012, pp. 54–61.
[29] N. Dhutraj, S. Sasane, V. Kshirsagar, "Hiding Sensitive Association Rule for Privacy Preservation", IEEE Transactions On Knowledge And Data Engineering, 2013, pp. 1-3.
[30] N. R. Radadiya, N. B. Prajapati, K. H. Shah, "Privacy Preserving in Association Rule mining", International Journal of Agriculture Innovations and Research, Vol. 2, No. 4, 2013, pp. 208–213.
[31] N. H. Domadiya, U. P. Rao, "Hiding Sensitive Association Rules to Maintain Privacy and Data Quality in Database", Advance Computing Conference, 2012, pp.1306-1310.
[32] G. V. Moustakides, V. S. Verykios, "A MaxMin Approach for Hiding Frequent Itemsets", Data & Knowledge Engineering, Vol. 65, No. 1, 2008, pp. 75-89.
[33] T-P. Hong, C-W. Lin, K-T. Yang, S-L. Wang, "Using TF-IDF to hide sensitive itemsets", Applied Intelligence, Vol. 38, 2013, pp. 502-510.
[34] C-W. Lin, T-P. Hong, H-C. Hsu, "Reducing Side Effects of Hiding Sensitive Itemsets in Privacy Preserving Data Mining", The Scientific World Journal, Vol. 2014, 2014, 12 pages.
[35] Y. Saygin, V. S. Verykios, A. K. Elmagarmid, "Privacy preserving association rule mining", Proceedings of the 12th International Workshop on Research Issues in Data Engineering: Engineering e-Commerce/ e-Business Systems, 2002, pp. 151-158.
[36] V. Verykios, and A. Gkoulalas-Divanis, "Privacy-preserving data mining models and algorithm", Advances in Database Systems, Eds, Springer US, 2008, pp. 282-283.

Narges Jamshidian Ghalehsefidi was born in Isfahan, Iran in 1988. She received a B.S. degree in Computer Engineering from Najafabad Branch, Islamic Azad University, Najafabad, Iran, and is currently an M.S. student in Computer Engineering (major: Software) at this university. Her field of research is Preserving Privacy in Data Mining.

Mohammad Naderi Dehkordi was born in Isfahan, Iran, in 1977. He has a Bachelor's degree in Computer Engineering from Isfahan University of Technology, Isfahan, Iran and a Master's degree in Computer Engineering from Najafabad Branch, Islamic Azad University. He has received his Ph.D. in Computer Engineering from Science and Research Branch, Islamic Azad University, Tehran, Iran, majoring in Privacy Preserving Data Mining. His main research interests include On-Line Analytical Processing and engineering privacy preserving in Hippocratic database.