ISSN No. 0976-5697

Volume 2, No. 5, Sept-Oct 2011

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Data Perturbation Method for Privacy Preseving Data Mining Based On Z-Score Normalization

Santosh Kumar Bhandare
M.Tech., Computer Science & Engineering
Samrat Ashok Technological Institute
Vidisha (M. P.) 464001 India
santosh.mits@gmail.com

Abstract - Data mining system consist of large amount of private and sensitive data such as healthcare, financial and criminal records. Some or all information of database may be confidential. This confidential information of database cannot be share to every one, so privacy protection of confidential information is required in data mining system for avoiding privacy leakage of data. Data perturbation is one of the best methods for privacy preserving data mining. In this paper we used data perturbation method for preserving privacy as well as accuracy. In data perturbation method individual data value are distorted before data mining application. In this paper we present Z- Score normalization transformation based data perturbation. The privacy parameters are used for measurement of privacy protection and the utility measure shows the performance of data mining technique after data distortion. We conducted experiment on real life dataset and the result show that Z-Score normalization transformation based data perturbation method is effective to protect confidential information and also maintain the performance of data mining technique after data distortion.

Keywords: Privacy preserving, normalization, data perturbation, classification, data mining

I. INTRODUCTION

Data mining [1] is the method of finding pattern from large amount of data using tool such as classification. The problem of privacy preserving is very important concern in data mining system. There are a lot of data mining application deal with privacy and security concern. Data mining system consist of large amount of private and secure data i.e. financial, criminal and healthcare records. These confidential records cannot be share to everyone so privacy protection of data is required for avoiding privacy leakage. There are a lot of research has been done in privacy preserving data mining based on randomization, secure multiparty computations, perturbation and Anonymity including K-anonymity and l-diversity.

In this paper we discuss the Z-Score normalization based data perturbation technique in which some or all confidential numerical attributes are distorted for privacy protection in classification analysis. Data perturbation is one of the best techniques for privacy preserving data mining system. In data perturbation the individual data values are distorted before data mining application.

The privacy parameters [2] are used for measurement of the degree of privacy protection. These parameters also show the capability of this technique to concealing the original data. The data utility measures show the performance of data mining technique after data distortion. In this paper we proposed Z-Score normalization transformation based data distortion method. We conducted experiment on real life dataset and the experimental results show that the proposed method is very effective to concealing the confidential information and also preserve the performance of data mining technique after data distortion.

II. RELATED WORK

The literatures on privacy preserving data mining can divide into two categories. In the first category, methods modify the data mining algorithms so that without knowing the exact values of data, they allow data mining operations on distributed dataset. In the second category, methods are modifying the values of the datasets to protect privacy of data values. In this category there are several research has been done in data distortion or data perturbation are as follow:

In the year 1982, T. Dalenius et al., [9] they firstly proposed the idea of Data Swapping. In this technique the database is to transform by switching a subset of attributes between selected pair of records. Therefore the lower order frequency counts are preserved in such a way that the data confidentiality is not compromised. The data swapping is a great data perturbation technique for privacy protection of data values.

In the year 1985, Liew et al., they proposed data distortion method based on probability distribution [17]. This method involves three steps: (i) identification of the underlying density function, (ii) generation of a distorted series from the density function, and (iii) mapping of the distorted series onto the original series.

In the year 2000, Agrawal R. et al., [18] proposed an additive data perturbation method for building decision tree classifiers. Every data element is randomized by adding some noise. These random noise chosen independently by a known distribution like Gaussian distribution. The data miner rebuilds the distribution of the original data from its distorted version. They consider the concrete case of building a Decision-tree classifier from training data in which the values of individual records have been perturbed. These perturbed data records look very different from the original records and the distribution of data values is also very different from the original distribution. Agrawal R. et al. proposed a reconstruction method to exactly approximation the distribution of data values. We build classifiers whose accuracy is comparable to the accuracy of classifiers built with the original data values by using these reconstructed distribution.

In the year 2002, Sweeney L. et al., [19] in this paper the k-Anonymity model consider the problem that a data owner wants to share a collection of person-specific data without revealing the identity of an individual. This goal is achieve by data generalization and suppression methods are used to protect the confidential information. This paper also examines the re-identification attacks.

In the year 2005, Chen et al., they proposed a rotation based perturbation method [20]. The proposed method maintains zero loss of accuracy for many classifiers. Experimental results show that the rotation perturbation can greatly improve the privacy quality without sacrificing accuracy.

In the year 2006, Weimin Ouyang et at., [16] in this paper they proposed a secure multi-party computation protocols for privacy preserving sequential pattern mining based on homomorphic encryption. When the protocol is applying each party transfer their own data to a trusted third party. Using this proposed method each parties securely keep their data privacy. We can also apply the secure multi-party computation protocol to clustering and classification.

In the year 2006, Wang et al., has used the Non-negative matrix factorization (NNMF) for data mining [8]. In this work, they combined non-negative matrix decomposition with distortion processing. The presented method have two important aspects (i) non-negative matrix factorization (NMF) is used to provide a least square compression version of original datasets and (ii) Using iterative methods to solve the least square optimization problem is provided an attractive flexibility for data administrator. The presented result given that the careful choice of iterative parameter settings, two sparse non-negative factors can solve by some efficient algorithms. Alternating least square using projected gradients in computing NNMF converges faster than multiplicative update methods. Iterative NMF based distortion method provides good solution for data mining problem on the basis of discriminate functions.

In the year 2006, Xu et al., proposed Singular value decomposition (SVD) based data distortion strategy for privacy protection [15]. In this work they propose a sparsified Singular Value Decomposition (SVD) method for data distortion. They conducted experiment on synthetic and real world datasets and the experimental result show that the sparsified SVD method is effective in preserving privacy as well as maintaining the performance of the datasets.

In the year 2007, Saif et al., also used non-negative matrix factorization for data perturbation [10]. They investigated the use of truncated non-negative matrix factorization (NMF) with sparseness constraints. The experimental results show that the Non-negative matrix factorization with sparseness constraints provides an efficient data perturbation tool for privacy preserving data mining. The privacy parameter used in the proposed work provides some indication on the ability of these techniques for concealing the original data values.

In the year 2007, Xu et al., has used the Fast Fourier Transform (FFT) for data perturbation [11]. The dataset is distorted or perturbed by using Fast Fourier Transform (FFT) for privacy protection of data values.

In the year 2008, Liu et al., has used the wavelet transformation for data distortion or data perturbation to preserve the privacy of data [12]. Privacy preserving strategy based on wavelet perturbation; keep the data privacy and data statistical properties and data mining utilities at the same time. The results show that presented method keep the distance before and after data perturbation and it also preserve the basic statistical properties of original data while maximizing the data utilities.

In the year 2009, Yingjie Wu et al., [21] , also proposed k-Anonymity privacy preserving for re-publication of incremental datasets. The k-Anonymity based privacy preserving is very popular approach for privacy protection. In this paper, they presented an effective approach for republication of incremental datasets. They analyze some possible generalizations in the anonymization for incremental updates. In the k-anonymity model, privacy is assured by ensuring that any record in the released data is indistinguishable from at least $k-1$ other records with respect to a set of attributes called quasi-identifier. There are a lot of k-anonymization algorithms have been developed, most of the existing methods assume that the dataset is fixed. In this paper they perform experiment on real life dataset, and the experimental result show that the proposed approach is effective.

In the year 2009, Chieh-Ming Wu et al., [22] they proposed greedy based technique for hiding the number of sensitive rules. This greedy technique is used to avoid the undesired side effects in the rule hiding process. So in this method the sensitive and confidential rules are hidden by greedy approach. The experimental results show that all the sensitive rules are hidden without generating the fake rules. The proposed technique generate lower number of missing rules than the other techniques

In the year 2009, Lin et al., has presented a method for data perturbation. In this method, the data matrix is vertically partitioned into several sub-metrics and held by different owners [13]. For perturbing their individual data, each data holder can randomly and independently choose a rotation matrix. The presented results show that random rotation based method for data perturbation preserve the data privacy without affecting the accuracy.

In the year 2010, Peng et al., [23] used combine data distortion strategies for privacy preserving data mining. They designed four schemes via attribute partition, with single value decomposition (SVD), non-negative matrix factorization (NMF), discrete wavelet transformation

(DWT) for distortion of sub matrix of the original dataset for privacy preserving. The basic idea of the proposed strategies was to perform distortion on sub matrices of original dataset using different method. Binary classification based on the support vector machine is used for measurement of Data utility. The real life dataset is taken form University of California, Irvine (UCI), Machine Learning Repository [42]. It is the Wisconsin breast cancer original dataset (WBC). The original version is used here. This dataset contain of 699 instances, 10 integer-valued attributes and one class attribute. In this paper they performed experiment on synthetic dataset as well as real life dataset. The experimental results show that proposed method was very efficient in maintaining data privacy as well as data utility in comparison to the individual data distortion techniques such as SVD, NMF and DWT.

In the year 2010, Peng et al., used combine data distortion strategies for privacy preserving data mining [14]. They designed four schemes via attribute partition, with single value decomposition (SVD), non-negative matrix factorization (NMF), discrete wavelet transformation (DWT) for distortion of sub matrix of the original dataset for privacy preserving. The basic idea of the proposed strategies was to perform distortion on sub matrices of original dataset using different method. The results show that proposed method was very efficient in maintaining data privacy as well as data utility in comparison to the individual data distortion techniques such as SVD, NMF and DWT.

III. ASSUMPTIONS

The object-attribute relationship of real life data sets are encode into vector – space format [3]. In this format a 2-dimentional is used to share the dataset. Row of the matrix indicates individual object and each column represent a particular attribute of these objects. In this matrix, we assume that every element is fixed, discrete and numerical. Any missing element is not allowed.

IV. DATA DISTORTION MEASURES

We used the same set of privacy parameters proposed in [2]. The privacy measures depends only on the original matrix M and its distorted matrix $\overline{D}$

A. Value Difference (VD):

After a data matrix is distorted by data distortion method, the value of its elements changes. The value difference of the datasets is defined by the relative value difference in the Frobenius norm. On the other hand VD is the ratio of the Frobenius norm of the difference of D and $\overline{D}$ to the Frobenius norm of D.

$$VD = \frac{\left\| D - \overline{D} \right\|}{\left\| D \right\|} \qquad \ldots\ldots (1)$$

Where || denotes the Frobenius norm of the enclosed argument

B. Position Difference:

The order of the value of the data element changes after data distortion. We use several metrics to measure the position difference of the data element.

a. RP- RP parameter is used to represent the average change of rank for all attributes after data distortion. For a dataset D with v data object and u attributes. Let S_j^i is the rank (in ascending order) of the jth element in attribute i. Similarly $\overline{S_j^i}$ is the rank of the corresponding distorted element. Then the RP parameter is given by:

$$RP = \frac{\sum_{i=1}^{u}\sum_{j=1}^{v}\left| S_j^i - \overline{S_j^i} \right|}{u * v} \qquad \ldots\ldots (2)$$

b. RK- RK parameter represents the percentage of elements that keeps their rank in each column after distortion. The RK parameter is given by:

$$RK = \frac{\sum_{i=1}^{u}\sum_{j=1}^{v} Rk_j^i}{u * v} \qquad \ldots\ldots (3)$$

Where $Rk_j^i = 1$ if $S_j^i = \overline{S_j^i}$ otherwise $Rk_j^i = 0$

c. CP- CP parameter is used to measure how the rank of the average value of each attributes varies after data distortion. CP represents the change of rank of the average value of the attributes. CP parameter is given by:

$$CP = \frac{1}{u}\sum_{i=1}^{u}\left| OD_i - \overline{OD}_i \right| \qquad \ldots\ldots (4)$$

Where OD_i and $\overline{OD_i}$ represent the rank of the average value of ith attribute before and after data distortion respectively.

d. CK- Similar to RK, CK is used to measure the percentage of the attributes that keep their rank of average value after data distortion. CK parameter is given by:

$$CK = \frac{1}{u}\sum_{i=1}^{u} Ck^i \qquad \ldots\ldots (5)$$

Where

$Ck^i = 1$ If $OD_i = \overline{OD_i}$

Otherwise $Ck^i = 0$

C. Utility Measure:

After the conduction of certain perturbation the data utility measures indicate the accuracy of data mining algorithms on distorted data. In this paper we choose the accuracy of a NBTree (Naive Bayes Classifier) [5] as our data utility measure.

V. Z-SCORE NORMALIZATION

Data transformation such as Normalization [4] is a data preprocessing tool used in data mining system. An attribute of a dataset is normalized by scaling its values. Normalization is particularly useful for classification algorithms involving neural networks, or distance measurements such as nearest neighbor classification and clustering. There are many methods for data normalization includes min-max normalization, z-score normalization and normalization by decimal scaling.

A. Z-Score Normalization::

Z-score normalization is also called zero-mean normalization. In Z-score normalization the values for an attribute **K** are normalized based on the mean and standard deviation of **K**. A value t of **K** is normalized to t' by the following formula:

$$t' = \frac{t - mean(K)}{std(K)} \qquad \text{….. (6)}$$

Where mean (K) = mean of attribute **K**
std (K) = standard deviation of attribute **K**

VI. THE PROPOSED ALGORITHM

Let N be a data matrix of dimension S X R, representing the original dataset. The rows of the matrix represent objects and the column of the matrix represent attributes.

Now the original data matrix N whose size is S X R, must be first transformed by Z-Score normalization transformation to get transformed matrix $\overline{N}$ whose size has the same size S X R as the original data matrix. The Z-Score normalization transformed each element of the original data matrix N into the specific range.

Now after applying the Z-Score normalization on the original data matrix, each element of the original data matrix N has been now perturbed into the specific range.

Now we have obtained a new matrix $\overline{N}$, which is very similar to the original data matrix N, but not identical. More importantly, $\overline{N}$ preserve the properties of N, so $\overline{N}$ can work as a distorted version of the original data matrix N.

Now the distorted data matrix is further shifted by multiplying it with a shifting factor, i.e. a negative number, to increase the security of data, because after applying the shifting factor (a negative number) on the distorted data matrix The order as well as the value of each element of the distorted data matrix was changed, i.e. the greater number become lesser and the lesser number become greater.

Algorithm: Data Perturbation Method for Privacy Preseving Data Mining Based On Z-Score Normalization

Input: Numerical Dataset, Shifting Factor ε_h (a negative number).

Output: Shifted Distorted Data Matrix.

Step 1: Initialized original data matrix N according to dataset, whose size is S X R. Where each row of the matrix N, indicate individual object and each column represent a particular attribute of these objects.

Step 2: The original data matrix N must be first transformed by Z-Score Normalization transformation. The Z-Score Normalization transformed each element of the original data matrix N into a specific range.

Step 3: Now after applying Z-Score Normalization transformation on original data matrix N, we get a transformed or distorted data matrix whose size is also S X R.

Step 4: The transformed or distorted data matrix is further shifted by multiplying it with a shifting factor ε_h, i.e. a negative number.

VII. EXPERIMENTAL RESULTS

We have conducted experiments to evaluate the performance of data distortion method. We choose four real-life Databases obtained from the University of California Irvine (UCI), Machine Learning Repository [6]. Datasets are the Iris, Glass Identification, Haberman's survival data, AND Bupa Liver Disorders Dataset. The summaries of the original database are given in Table [I] and Table [IV] show the performance of distorted method. We use WEKA (Waikato Environment for Knowledge Analysis) [7] software to test the accuracy of distorted method. The privacy parameters are measured by a separate Java programme. We have constructed the classifier for NBTree classification, and a 10-fold cross validation to obtain the classification results.

We apply our data perturbation method with shifting factor $\varepsilon_{h\,=}$ -5 on the real life dataset mentioned in table [I] and get the results. All these result shown in table [II].

Table [II], [III], [V], [VI] show the performance of Iris, Bupa liver disorder, Haberman's Survival data and Galss identification respectively. In Haberman's Survival data and IRIS datasets the value difference VD is vary with respect to shifting factor $\varepsilon_{h.}$ But in case of Glass Identification and Bupa liver disorder dataset, the value difference VD as well as accuracy have changed with respect to Shifting factor ε_h because after applying shifting factor the value of each element of distorted matrix $\overline{D}$ was changed. Therefore we carefully choose the correct shifting factor ε_h for better result.

Table I: The summary of the database

Database	Number of Instances	Number of Features	Number of Classes
IRIS	150	4	3
Haberman's Survival data	306	3	2
Bupa Liver Disorders	345	6	2
Glass Identification	214	10	7

Table II: Relation Between ε_h, Privacy Parameter and NBTree Accuracy of IRIS dataset Original Accuracy of NBTree is: 94%

ε_h	VD(i)	RP	RK	CP	CK	ACC %
-1	1.10312	74.74333	0	2.0	0	92.6667
-2	1.78552	74.74333	0	2.0	0	92.6667
-3	2.62937	74.74333	0	2.0	0	92.6667
-4	3.52043	74.74333	0	2.0	0	92.6667
-5	4.43030	74.74333	0	2.0	0	92.6667

Table III: Relation Between ε_h, Privacy Parameter and NBTree Accuracy of Bupa liver disorder dataset Original Accuracy of NBTree is: 66.087%

ε_h	VD(b)	RP	RK	CP	CK	ACC %
-1	0.97017	172.40966	0	3.0	0	65.5072
-2	0.99854	172.40966	0	3.0	0	65.5072
-3	1.08053	172.40966	0	3.0	0	65.2174
-4	1.20525	172.40966	0	3.0	0	65.2174
-5	1.36100	172.40966	0	3.0	0	65.2174

Table IV: How the privacy parameters and accuracy vary in four datasets

Data	**VD**	**RP**	**RK**	**CP**	**CK**	**Acc in %**
Iris (Original)	-	-	-	-	-	94
Iris (Perturb)	4.43030	74.74333	0	2.0	0	92.6667
Bupa Liver Disorders (Original)	-	-	-	-	-	66.087
Bupa Liver Disorders (Perturb)	1.36100	172.40966	0	3.0	0	65.2174
Haberman's survival data (Original)	-	--	-	-	-	72.549
Haberman's survival data (Perturb)	1.08037	151.98257	0	1.33333	0.33333	72.549
Glass dentification (Original)	-	-	-	-	-	94.3925
Glass dentification (Perturb)	1.00833	101.2514	0.00654	5.0	0	94.3925

Table V: Relation Between ε_h, Privacy Parameter and NBTree Accuracy of Haberman's Survival Data Original Accuracy of NBTree is: 72.549%

ε_h	VD(h)	RP	RK	CP	CK	ACC %
-1	1.01013	151.98257	0	1.33333	0.33333	72.549
-2	1.02336	151.98257	0	1.33333	0.33333	72.549
-3	1.03957	151.98257	0	1.33333	0.33333	72.549
-4	1.05863	151.98257	0	1.33333	0.33333	72.549
-5	1.08037	151.98257	0	1.33333	0.33333	72.549

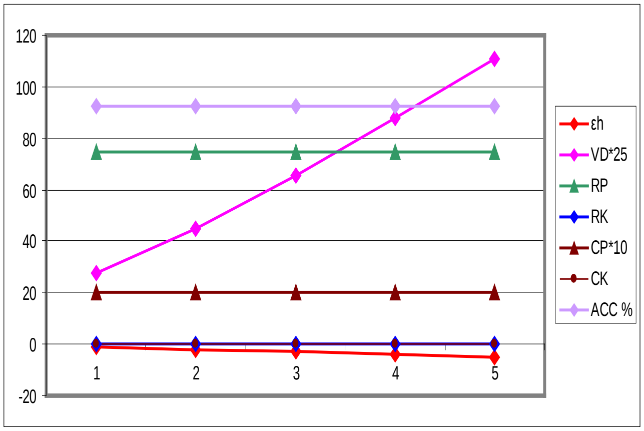

Figure 7.1: Relation between accuracy, parameter and Shifting Factor on Iris dataset.

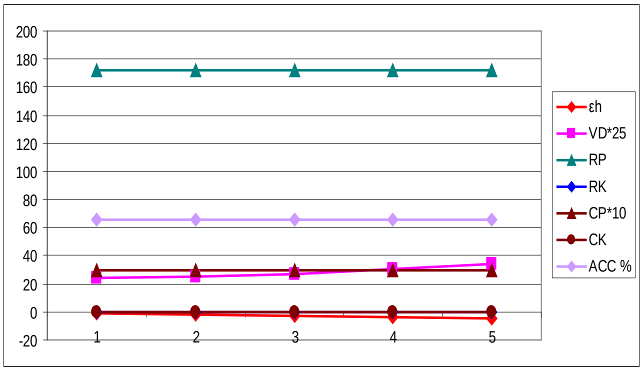

Figure 7.2: Relation between accuracy, parameter and Shifting Factor on Bupa Liver Disorder dataset.

Table VI: Relation Between ε_h, Privacy Parameter and NBTree Accuracy of Glass Identification Dataset Original Accuracy of NBTree is: 94.3925 %

ε_h	VD(g)	RP	RK	CP	CK	ACC %
-1	0.99939	101.2514	0.00654	5.0	0	95.7944
-2	0.99992	101.2514	0.00654	5.0	0	94.3925
-3	1.00159	101.2514	0.00654	5.0	0	94.8598
-4	1.00440	101.2514	0.00654	5.0	0	94.3925
-5	1.00833	101.2514	0.00654	5.0	0	94.3925

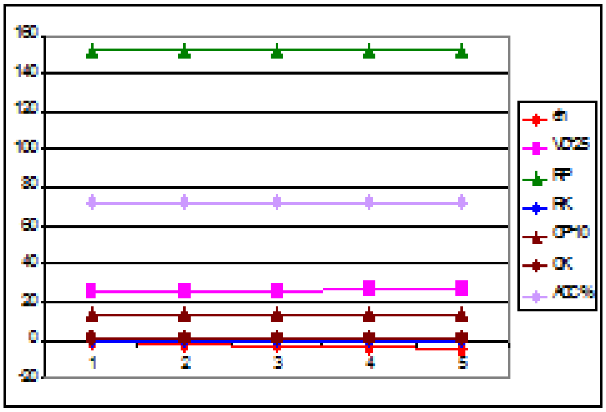

Figure 7.3: Relation between accuracy, parameter and Shifting Factor on Haberman's Survival dataset.

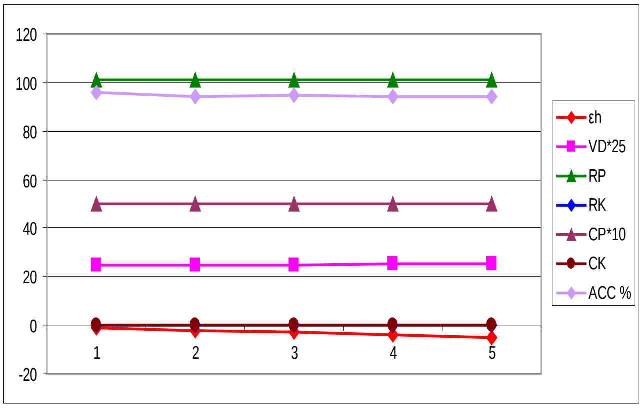

Figure 7.4: Relation between accuracy, parameter and Shifting Factor on Glass Identification dataset.

VIII.CONCLUSION AND FUTURE WORK

In this paper we proposed Z- Score normalization transformation based privacy preserving data distortion method. We conducted the experiment on four real life datasets and the experimental result show that the Z- Score normalization transformation based data distortion method is very effective for privacy preserving data mining. The privacy parameters used in this work show that the proposed method is very effective to concealing the confidential information. In addition, the proposed method also maintain the performance of data mining technique after data distortion, it is interesting to use the other normalization methods like Normalization by decimal scaling and also compare its result with Z- Score normalization.

IX. REFERENCES

[1]. M. Chen, J. Han, and P. Yu, "Data mining: An Overview from a database Prospective", IEEE Trans. on Knowledge and Data Engineering, vol. 8, no. 6, pp. 866-883, Dec. 1996.

[2]. S. Xu, J. Zhang, D. Han, J. Wang, "Data distortion for privacy protection in a terrorist analysis system", Proceeding of the IEEE International Conference on Intelligence and Security Informatics, pp. 459-464, 2005.

[3]. W. Frankes and R. Baeza-Yates, "Information Retrieval: Data Structures and Algorithms", Prentice–Hall, Englewood cliffs, NJ, 1992.

[4]. J. Han, and M. Kamber, "Data Mining: Concepts and Techniques", Second edition, 2006, Morgan Kaufmann, USA.

[5]. Ian H. Witten, Eibe Frank,"Data Mining Practical Machine Learning Tools and Techniques", Second Edition, 2005.

[6]. UCI Machine Learning Repository http://archive.ics.uci.edu/ml/datasets.html

[7]. The Weka Machine Learning Workbench. http://www.cs.waikato.ac.nz/ml/weka

[8]. Jie Wang, Weijun Zhong, Jun Zhang, "NNMF- Based Factorization Techniques for High-Accuracy Privacy Protection on Non-negative-valued Dataset", Proceeding of IEEE Conference on Data Mining, International Workshop on Privacy Aspects of Date Mining (PADM2006), pp.513-517, 2006.

[9]. T. Dalenius and S.P. Reiss, "Data-Swapping: A Technique for Disclosure Control," Journal of Statistical Planning and Inference, vol. 6, pp. 73-85, 1982.

[10]. Saif M. A. Kabir, Amr M. Youssef, Ahmed K. Elhakeem, "On data distortion for privacy preserving data mining", Proceedings of IEEE Conference on Electrical and Computer Engineering (CCECE 2007), PP. 308-311, 2007.

[11]. Shuting Xu, Shuhua Lai,"Fast Fourier transform based data perturbation method for privacy protection", Proceeding of IEEE International Conference on Intelligence and Security Informatics, pp. 221-224, 2007.

[12]. Lian Liu, Jie Wang, Jun Zhang,"Wavelet-Based Data Perturbation for Simultaneous Privacy-Preserving and Statistics-Preserving", Proceeding of IEEE International Conference on Data Mining Workshop, PP. 27-35, 2008.

[13]. Zhenmin Lin, Jie Wang, Lian Liu, Jun Zhang, "Generalized random rotation perturbation for vertically partitioned data sets", Proceeding of the IEEE Symposium on Computational Intelligence and Data Mining, pp:159- 162, 2009.

[14]. Bo Peng, Xingyu Geng, Jun Zhang, "Combined data distortion strategies for privacy-preserving data mining", Proceeding of the IEEE International Conference on Advanced Computer Theory and Engineering (1CACTE), PP. V1-572 - V1-576, 2010.

[15]. S. Xu, J. Zhang, D. Han and J. Wang, "Singular value decomposition based data distortion strategy for privacy protection", ACM Journal of Knowledge and Information Systems, vol. 10, no. 3, pp. 383-397, 2006.

[16]. Weimin Ouyang, Qinhua Huang, "Privacy Preserving Sequential Pattern Mining Based on Secure Multi-party Computation", Proceedings of the 2006 IEEE International Conference on Information Acquisition, pp. 149-154, August 20 - 23, 2006.

[17]. C. K. Liew, U. J. Choi, and C. J. Liew, "A Data Distortion by Probability Distribution", ACM Transaction on Database Systems (TODS), vol. 10, no. 3, pp. 395-411, Sep. 1985.

[18]. R. Agrawal and R. Srikant, "Privacy-preserving data mining", Proceeding of the ACM SIGMOD Conference on Management of Data, pp. 439–450, May 2000.

[19]. L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems, vol. 10, no. 5, pp. 557-570, 2002.

[20]. K. Chen and L. Liu, "Privacy preserving data classification with rotation perturbation", Proceeding of the 5th IEEE International Conference on Data Mining (ICDM 2005), pp. 589-592, 2005.

[21]. Yingjie Wu, Zhihui Sun, Xiaodong Wang, "Privacy Preserving k-Anonymity for Re-publication of Incremental Datasets", 2009 World Congress on Computer Science and Information Engineering, pp. 53 – 60,2009

[22]. Chieh-Ming Wu, Yin-Fu Huang and Jian-Ying Chen, "Privacy Preserving Association Rules by Using Greedy Approach", World Congress on Computer Science and Information Engineering, pp 61-65, 2009

[23]. Bo Peng, Xingyu Geng, Jun Zhang, "Combined data distortion strategies for privacy-preserving data mining", Proceeding of the IEEE International Conference on Advanced Computer Theory and Engineering (1CACTE), pp. V1-572 - V1-576, 2010.