ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

Data Privacy Preservation Using Various Perturbation Techniques

S. Mayil[1] and Dr. M.Vanitha[2]

[1] Research Scholar, PG and Research Department of Computer Science,
J.J.College of Arts and Science (Autonomous), Pudukkottai,Tamilnadu,India.
Email: mayilma@yahoo.co.in
[2] Assistant professor, Department of Computer Science and Engineering,
Alagappa University,Karaikudi, India.
Email: mvanitharavi@gmail.com

Abstract: **Data privacy is a big issue in every field and it plays a major role in the field of data publishing. The data publishing is done for the purpose of data analysis so the information is collected from various organizations. The collected information should protect the record owner's identification. It is the main issue before publishing the data to other organizations. The privacy of the records should be maintained by the particular organization which is going to publish the data for others analytical purpose. There are various techniques involved in the privacy preservation of data publishing. Among them perturbation technique is an important method to perturb the data that can help to publish the data for further use of the records by other organizations. It is used for both data privacy and accuracy. In this paper we are going to discuss various perturbation techniques that are used for data privacy**.

Keywords: *data privacy, perturbation, privacy preserving, data mining, random data perturbation*

I.Introduction

Huge volumes of detailed personal data are regularly collected and analyzed by applications using data mining. Such data include shopping habits, criminal records, medical history, credit records, among others. On the one hand, such data is an important asset to business organizations and governments both to decision making processes and to provide social benefits, such as medical research, crime reduction, national security, etc. The threat to privacy becomes real since data mining techniques are able to derive highly sensitive knowledge from unclassified data that is not even known to database holders. Worse is the privacy invasion occasioned by secondary usage of data when individuals are unaware of "behind the scenes" use of data mining techniques.

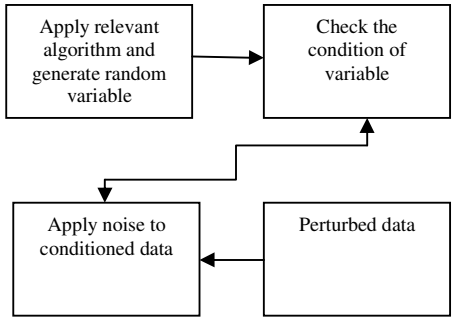
There are large number of applications like financial, educational and health sectors that requires privacy preserving data mining which can be further used for the mining of useful information. The data should maintain the reputation and regulation of data owners by maintaining the privacy of sensitive information and in another case it should also be helpful to discover new knowledge pattern. This is done when there is necessity for out sourcing the data to third parties by the record owners [1]. This can be done using various techniques such as anonymization, perturbation, etc., Anonymization is a heuristic method for privacy preserving data publishing [2] where data perturbation is a reconstruction method used in privacy preserving data mining [3].

B. Probability Distribution Method:

preserving data mining which can be further used for the mining of useful information. The data should maintain the reputation and regulation of data owners by maintaining the privacy of sensitive information and in another case it should also be helpful to discover new knowledge pattern. This is done when there is necessity for out sourcing the data to third parties by the record owners [1]. This can be done using various techniques such as anonymization, perturbation, etc., Anonymization is a heuristic method for privacy preserving data publishing [2] where data perturbation is a reconstruction method used in privacy preserving data mining [3].

A.Data Perturbation:

It is a technique for maintaining the data privacy. This technique changes the data record value without changing the underlying meaning of data. It uses two techniques Probability Distribution of data and Data distortion by building decision tree classifiers to add noise or any other data to the original data for the classification before data publishing which can be further reconstructed by the record owners knowing the sampling data used for changing the original data. This mainly deals with the confidentiality of the data [4], [5].



Implementing perturbed techniques

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

The Probability distribution technique converts the original data by the same distribution's sample data or by itself [6].

C.Value Distortion

The Value distortion changes the value by adding additive or multiplicative noise to the data directly to the original data [6].

II.Types of Data Perturbation

The above two method is applied in all types of data perturbation to protect the data privacy. The types are mainly classified into three types.

1) Rotation Perturbation:

In this method the value of the two attributes in the matrix are rotated but the meaning of the value is protected. The value distortion technique is applied by first selecting the pair of attributes then their values are distorted by rotating those two values [7].

2) Projection Perturbation:

The data perturbation is done by selecting the data value in high dimensional space and it is changed to the data in lower dimensional space randomly. This can be projected either column or row wise [8].

3) Geometric Perturbation:

his is a hybrid technique with the combination of rotation, translation and adding random noise value to the given data value in the matrix to provide quality of data preservation mainly for clustering [9].

II. Data Perturbation

In all these methods the data perturbation is done by using random value matrix for multiplicative or additive noise to the data perturbation before data publishing to protect the privacy of data [10].

A. Rotation Perturbation

1) Keke Chen, Ling Liu [11]:

In random rotation perturbation technique, multiple column data values are converted in single column transformation. Therefore it leads to the query of privacy in multi dimensional rotation perturbation. In this technique the geometric distributions of random matrix should be considered to increase the privacy of the data value. A new approach for this problem with unified metric for privacy and also protects multi column privacy. This metric helps to analyze the issues against ICA-based reconstruction attacks of rotation perturbation. In this paper, it is also proved that higher privacy and also accuracy of rotation-invariant classifiers are protected than other perturbation techniques.

2) Zhenmin Lin, Jie Wang ; Lian Liu ; Changjiang Zhang [12]:

In this paper the disadvantage of random rotation perturbation technique is handled because of the issue that this technique cannot hold the geometric properties of the rotation random matrix used for the classification in order to preserve privacy. So here the data matrix is partitioned vertically and for each partitioned data sub-set matrix the random rotation matrix is used to perturb the original value of the data. These data is used for data analysis purpose by third parties. This centralized algorithm also handles the privacy issue by maintaining the geometric properties of the matrix and the accuracy of classifiers used for classification before publishing the data.

3) Li Liu, Murat Kantarcioglu and Bhavani Thuraisingham [13]:

A new classifier named C4.5 decision tree classifier is proposed for preserving the data privacy and also to classify and reconstruct the original data with high accuracy from the perturbed data. It decreases the computation and communication cost in terms of time. There are some data mining techniques that can be applied to perturbed data directly because the perturbation process that will preserve the nature of the data. Data mining classifier such as Naïve Bayes classifier can applied to the additive perturbation data and Euclidean based data mining tools directly, For example, .k-Nearest Neighbor Classifier, Support Vector Machines, and Perceptrons Neural Network can be applied to the multiplicative perturbation data. But the information loss will be reduced in the process and the other issue is performance in the applied process. On comparing with other techniques perturbation with random data rotation provides high privacy along with decision tree classifiers leaving the hard distribution issue solved.

B.*Projection Perturbation*

1) Yingpeng Sang, Hong Shen , Hui Tian [14]:

In privacy preserving data mining. Random Projection (RP) is gaining more concern for its high efficiency. As mentioned in [8], the original dataset with m attributes when multiplied k times by a random matrix then there can be k (m>k) series of data that are perturbed and can be published for data analysis. In this paper different prior knowledge are considered before reconstructing the data based on Underdetermined Independent Component Analysis (UICA) and on Maximum A Posteriori (MAP) methods used for correlation but the attributes should be mutually independent. It is proved that UICA and MAP outperforms the Principal Component Analysis (PCA) when used for perturbations by increasing the privacy of data and provides higher security.

C. Geometric Perturbation

1) Keke Chen, Ling Liu [15], [16]:

Geometric perturbation is very effective perturbation technique for privacy preserving data publishing on single-party. In this paper, the multiparty privacy-preserving collaborative mining of the geometric perturbation is proposed. For this approach three protocols namely simple, negotiation and Space adaption protocols uses either randomized or optimized algorithms for the generation of perturbation. The performance analysis of these three protocols shows that simple

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

protocol there are issues like cannot provide assurance of privacy to data providers and encryption is used so it cannot be easily shared by the public for data publishing but space adaption protocol can provide better scalability, flexibility for the distribution of data and overall satisfaction in protecting privacy for large space adaption protocol gives better result in improving privacy when there is multiple data providers.

A random geometric perturbation approach is proposed to preserve privacy on data classification. In random geometric perturbation $G(X)=RX+\Psi+\Delta$ there are three linear combinations, rotation, translation and distance perturbation Geometric perturbation in data-mining preserve the geometric class boundaries of the perturbed data. A multi-column privacy evaluation model and also unified privacy is designed and proposed to address the invariant classifier problems and to protect the naive-inference attacks, ICA-based attacks, and distance-inference attacks.

2) Keke Chen, Gordon Sun, Ling Liu [17]:

The improvement of accuracy and assurance of privacy can be improved by using Task/model-oriented perturbation.

For a bunch of popular data classification models Geometric perturbation can be used because the trained and tested data of perturbed datasets similar to the trained and tested original dataset in accuracy. In this paper potential attacks of geometric perturbation are analysed and proved that random optimization method with geo-metric perturbation will provide a satisfactory privacy assurance in compromise with less accuracy. This framework helps to analyse more attacks and also for optimization of the geometric perturbation.

IV. Conclusion

In this paper, we have analysed various perturbation techniques that can be used for privacy preserving data publishing. Data publishing is recently gaining greater impact hence the data privacy is also gaining greater impact which should be maintained in the data publishing. So we discussed in detail about the issues and its merits of each type of perturbation technique. We mainly focused on the random data perturbation which is implemented to increase efficiency in perturbation techniques and also discussed about the improvement of privacy from the attackers. In the analysis of these perturbation techniques, it is considered that geometric distribution provides higher privacy and also the data utility compared to other techniques in privacy preserving data mining.

REFERENCES

[1] Agrawal, C. and Yu, P.S., "General Survey of Privacy Preserving Data Mining Models and Algorithms",
Privacy-Preserving Data Mining Advances in Database Systems Volume 34, pp 11-52,2008.

[2] Fung, B.C.M., Wang, K., Chen, R., and Yu, P.S., "Privacy-Preserving Data Publishing: A Survey of Recent Developments," ACM Computing Surveys", Vol. 42, No. 4, pp. 1-53, 2010.

[3] Sweeney, L., "k-Anonymity: A Model for Protecting Privacy," Int'l J. Uncertainty, Fuzziness and Knowledge-Based Systems, Vol. 10, No. 5, pp. 557-570, 2002.

[4] Stanley, R. M., Oliveira and Osmar R. Za¨iane, "Privacy Preserving Clustering by Data Transformation", Journal of Information and Data Management", Vol. 1, No. 1, 2010.

[5] Agrawal, D., and Aggarwal, C.C., "On the Design and Quantification of Privacy Preserving Data Mining Algorithms," Proc. of 20th ACM SIGMOD-SIGACT-SIGART Symp. on Principles of Database Systems (PODS'01), pp. 247-255, 2001.

[6] Verykios, S., Bertino, E., Fovino, I.N., Provenza, L.P., Saygin, Y., and Theodoridis, Y., "State-of-the-Art in PPDM", ACM SIGMOD Record, Vol 3, pp. 50-57, 2004.

[7] Stanley,, M., Oliveir, p., and Osmar, R.,"Data Perturbation By Rotation for Privacy Preserving Clustering", University of Albetra, Technical Report TR 04-17, 2004.

[8] Kun, L., Hillol, K., and Jessica, R., "Random Projection-Based Multiplicative Data Perturbation for rivacy Preserving Distributed Data Mining", IEEE Transactions on Knowledge and Data Engineering, Vol. 18, pp.58-64, 2006.

[9] Jie Liu, Yifeng XU, "Privacy Preserving Clustering by Random Response Method of Geometric Transformation", Fourth International Conference on Internet Computing for Science and Engineering (ICICSE), pp. 181-188, 2009.

[10] Agrawal, R., Srikant, R., "Privacy-Preserving Data Mining", Proceedings of the 2000 ACM SIGMOD international conference on Management of data, pp. 439-450, 2000.

[11] Keke Chen, Ling Liu, "Privacy Preserving Data Classification with Rotation Perturbation", Fifth IEEE International Conference on Data Mining, 2005.

[12] Zhenmin Lin, Lexington, KY, Jie Wang, Lian Liu , Changjiang Zhang, "Generalized Random Rotation Perturbation for Vertically Partitioned Data Sets", IEEE Symposium on Computational Intelligence and Data Mining, 2009.

[13] Li Liu, Murat Kantarcioglu and Bhavani Thuraisingham, "Privacy Preserving Decision Tree Mining from Perturbed Data", IEEE Proceedings of the 42nd Hawaii International Conference on System Sciences, 2009.

[14] Yingpeng Sang , Hong Shen , Hui Tian, "Effective Reconstruction of Data Perturbed by Random Projections", IEEE Transactions on Computers, Vol. 61 , No. 1, Jan. 2012.

[15] Keke Chen, Ling Liu, "Geometric Data Perturbation For Privacy Preserving Outsourced Data Mining", Springer Knowl Inf Sys, 2010.

[16]Keke Chen, Ling Liu, "Privacy-preserving Multiparty Collaborative Mining with Geometric Data Perturbation", IEEE Transactions on Parallel and Distributed Computing, 2009.

[17] Keke Chen, Gordon Sun, Ling Liu, "Towards Attack-Resilient Geometric Data Perturbation", Proceedings of the Seventh SIAM International Conference on Data Mining, 2007.