IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

Study of Data Mining Approach for Privacy Preservation

Priyank Pathak[1], Rajat Paliwal[2]

M.Tech Scholar, Department of CSE, RITS, Bhopal, India[1]

Professor, Department of CSE, RITS, Bhopal, India[2]

Abstract: In this paper reviewed the utility and application of data mining technique in the field of privacy preservation. Privacy preservation is technique for hiding of information and secured the information during transmission. Now a day's various technique of privacy preservation are used such as cryptography, k-animosity and other methods used for the hiding an information. Data mining provide verity of technique such as rule mining, clustering and classification, all these technique used for the process of privacy preservation. The noise adaptive and data transformation is well known technique for privacy preservation. The collection of different method of data mining and perform privacy preservation task is called collaborative mining technique for this task. The collaborative technique enhances the security strength of privacy preservation and decrease the loss of data during the transformation of data.

Keywords: Privacy Preservation, Data Mining, Collaborative Technique.

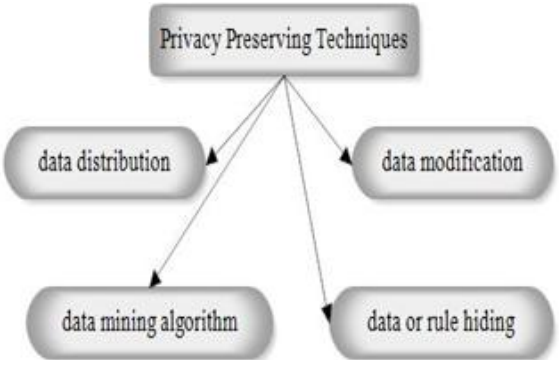
I. INTRODUCTION

Confidentiality and authentication of data is major issue in current scenario. For the confidentiality and authentication of data various technique are used such as cryptography, data Romanization, third party access control and many more method[1,2]. The conventional technique such as cryptography and other technique faced a problem of security issue in privacy preservation. Now a day's data mining technique play an important role for the privacy preservation [3, 4]. For the purpose of this used rule mining technique, classification technique and clustering technique. The rule mining technique is very important role in terms of transformation. The process of transformation changes the value of minimum support and confidence. And change the order of data associated with this range and hide the information [5, 6, 7].

Instead of these technique used clustering and classification for the process of privacy preservation. The process of clustering and classification such as decision tree and knn are used for this purpose. Now a day's principle of component analysis is used. The process of data privacy preservation proceeds in two different ways [8]. First act as sensitive raw data such as name, indemnifiers and some other important record. And other is sensitive information mined from database using data mining algorithm [9]. The process of data mining facilities the process of algorithms for modifying the original data in some way, so that the private data or private knowledge remain private even after the mining process. The problem that arises when confidential data can be derived from released data by unauthorized users is also commonly called the data duplication problem. Now a day's smc play an important role in privacy preservation in concern of third party communication [10]. They believe of all parties justify by the common factor of data analysis. The protocols of smc ensure that the communication party involve in proper manner [10]. In other words, unless proper incentives are set, current smc techniques cannot prevent input modification by participating parties. In section ii discuss the technique of privacy preservation. In section iii discuss the related work of data mining technique. In section iv discuss problem formulation and finally in section v discuss conclusion and future work.

II. PRIVACY PRESERVATION TECHNIQUE

The first scope of privacy preservation is data distribution. The categorization of technique basically based on the process and nature of data. The data distribution technique encompassed two different techniques such as vertical mining and horizontal mining [13, 15]. The process of data modification deals with the data transformation and random noise addition. The data mining algorithm such as clustering classification and various technique are applied for the process of privacy preservation.



IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

The rule hiding is also a sub class of data mining technique applied on the basis of support and confidence.

III.RELATED WORK

In this part discussed the related work in the field of privacy preservation using data mining technique and some other techniques. The data mining technique offers various algorithms for the process of privacy preservation. Association rule to play an important role in privacy preservation. Here discuss some work along their authors.

[1] In this paper author describes a Key Distribution-Less Privacy Preserving Data Mining system in which the publication of local association rules generated by the parties is presented. The association rules are securely combined to form the combined rule set using the (KDLPPDM) algorithm. The combined rule sets established are used to classify or mine the data. The results discussed in this paper compare the authenticity of the rules generated using the C 4.5 based KDLPPDM system and the C 5.0 based KDLPPDM system using receiver operating characteristics curves (ROC).

[2] In this paper, they first develop key theorems, and then base on these theorems, they analyse certain important privacy-preserving data analysis tasks that could be organized in a way that telling the truth is the best choice for any participating party. they have investigated what kinds of PPDA tasks are incentive compatible under the NCC model. Based on our findings, there are certain important PPDA tasks that are incentive driven. classifies the common data analysis tasks studied in this paper into DNCC or NonDNCC class. Most often, data partition schemes can make a difference in determining DNCC or Non-DNCC classifications.

[3] This paper proposed an article selection with privacy preservation in centralized network. Data can be preserved for privacy by perturbation approach as alias name. In centralized data evaluation, it makes data classification and feature choice for data mining decision model which generate the structural information of model in this paper. The application of gain ratio technique for improved performance of feature selection has taken to perform the centralized computational task. All articles don't need to preserve the privacy for confidential data for best model. The chi-square testing has taken for the classification of data by centralized data mining model using own processing unit.

[4] In this paper they review on the various privacy preserving data mining techniques like data modification and secure multiparty computation based on the different aspects. Data mining is such a technique which extracts the useful information from the large depository. Knowledge discovery in database (KDD) is another name of data mining. Privacy preserving data mining techniques are imported with the aim of extract the relevant knowledge from the large amount of data while protecting the sensible information at the same time.

[5] In this paper, they propose a generic PPDM framework and a simplified taxonomy to help understand the problem and explore possible research issues. they also examine the strengths and weaknesses of different privacy preserving approaches and summarize general principles from early research to guide the selection of PPDM algorithms. they conduct an extensive review on literature. they present a classification scheme, adopted from early studies, to guide the review process. As part of future work, they plan to apply the proposed evaluation framework to formally test a complete spectrum of PPDM algorithms.

[6] In this paper author discuss about the challenges in privacy-preserving data quality assessment. A two-party scenario is considered, existing of a client that wishes to test data quality and a server that holds the dataset. Privacy preserving protocols are presented for testing significant data quality metrics: completeness, consistency, uniqueness, timeliness validity. For semi-honest parties, the protocols ensure that the client does not explore any information about the data other than the value of the quality metric. The server does not explore the parameters of the client's query, the specific attributes being tested and the computed amount of the data quality metric.

[7] This paper propose a privacy preserving approach that can be applied to decision tree learning, without concomitant loss of accuracy. It defines an approach to the preservation of the privacy of collected data samples in cases where information from the sample database has been partially lost. This approach converts the original sample data sets into a class of unreal data sets, from which the original samples cannot be reconstructed without the entire class of unreal data sets. Meanwhile, a specific decision tree can be built directly from those unreal data sets. This unique approach can be applied directly to the data storage as soon as the first sample is collected.

[8] In this paper, they have given a review of the state-ofthe-art methods for privacy and analyze the representative technique for privacy preserving data mining and point out their advantages and disadvantages. Privacy preserving data mining has been studied broadly, because of the wide proliferation of sensitive information on the internet. They discuss method for Perturbation, K-Anonymization, condensation, and Distributed Privacy Preserving Data mining. While all the proposed methods are only approximate to our goal of privacy preservation, they need to further perfect those approaches or develop some efficient methods.

[9] In this work, they relax this assumption and expand the scope of perturbation-based PPDM to Multi-Level Trust (MLT-PPDM). In our context, the more trusted a data miner is the less perturbed copy of the data it can access. Under this setting, a malicious data miner may have enter

IJARCCE
ISSN (Online) 2278-1021
ISSN (Print) 2319 5940
International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 8, August 2016

to differently perturbed copies of the same data through various means, and may combine these different copies to jointly infer additional information about the original data that the data owner does not intend to release. Stopping such diversity attacks is the key challenge of providing MLTPPDM services. They address this challenge by properly associating perturbation across copies at different trust levels. They prove that our solution is robust against diversity attacks with respect to our privacy goal.

[10] In this paper author proposed framework provides a good basis for more accurate comparison of the given techniques to privacy preserving distributed data mining. In addition, this framework allows recognizing the overlapping amount for Divers approaches and identifying modern approaches in this field. At first, these techniques divided into three approaches of secure multiparty computation, secret sharing and perturbation and then every approach was being investigated. Accordance proposed evaluation framework, the premise of ensuring the privacy of how to plan an effective technique against malicious model and independent from the assumptions.

IV. PROBLEM FORMULATION

The problem of privacy preservation in the process of data evaluation in real time scenario, the inability of algorithm the process of data transformation faced a problem of extraction of data. The lacking of extraction most of data part is lost. The major issue in the form of social community data such as health sector data, statics data, and social website data [14]. In some case the organization develop the parallel lines of data for the purpose of communication.

Data Publishing
Data publishing is a realistic problem related to the privacy preservation technique. In data publishing technique the data are performed in terms of output of user seen. For the publishing of data some mining technique are used such as classification and clustering.

Modification of result
In the process of privacy preservation used data mining algorithm compromised the result of evaluation such as classification method not find the transform value of result in data for the process of privacy pervasion.

Cryptographic Methods
In many cases, the data may be distributed across multiple sites, and the owners of the data across these multiple sites may wish to compute a common function. In such cases, cryptographic protocols may be used in order to communicate among the different sites, so that secure function computation is possible without revealing sensitive information.

Dimensionality
In real scenario of computing the processing of data is very high dimensions, the high dimensions data precede a very difficult problem for the privacy preservation. Now dimension reduction loss some data attribute during the process of data.

V. CONCLUSION

In this paper study of privacy preservation technique using data mining and some other collative technique of mining. The lacking of appropriate method of privacy preservation, it faced a problem of loss of data and dimension reduction and many more. The review in here, gives the information of securing the sensitive data and interest the secured the sensitive data from unauthorized user. Study of all technique reached only the limitation of data mining algorithm and impact of result of this entire algorithm. The lacking of many principle of data mining technique not cooperates properly for sensitive data for security issue. Now in future used some feature selection technique for incorporate mining technique for better preservation. IEEE LaTeX style files which have been used in the preparation of this template.

REFERENCES

[1] S KumaraSwamy,Manjula S , K R Venugopal, Iyengar S , L M Patnaik "Association Rule Sharing Model for Privacy Preservation and Collaborative Data Mining Efficiency" IEEE, 2014. Pp 1-6.
[2] Murat Kantarcioglu, Wei Jiang "Incentive Compatible Privacy-Preserving Data Analysis" IEEE, 2013. Pp 1323- 1335.
[3] He manta Kumar Bhuyan , Maitri Mohant , Smruti Rekha Das "Privacy Preserving for Feature Selection in Data Mining Using Centralized Network" IJCSI, 2012. Pp 434-440.
[4] Manish Sharma, Atul Chaudhary, Manish Mathuria, Shalini Chaudhary "A Review Study on the Privacy Preserving Data Mining Techniques and Approaches" IJCST, 2013. Pp 42-46.
[5] K. Srinivasa Rao, B. Srinivasa Rao "An Insight in to Privacy Preserving Data Mining Methods" CSEA, 2013. Pp 100-104.
[6] Julien Freudiger, Shantanu Rane, Alejandro E. Brito , Ersin Uzun PARC "Privacy Preserving Data Quality Assessment for High-Fidelity Data Sharing" ACM, 2014. Pp 1-9.
[7] Kumaraswamy S, Manjula S, K R Venugopal, L M Patnaik "A Data Mining Perspective in Privacy Preserving Data Mining Systems" IJCST, 2014. Pp 704-711.
[8] Beula Amalorpavam, N. Mookhambik "privacy preserving decision tree learning using unrealized data sets" IJITCSP, 2013. Pp 187-192.
[9] Ms.R.Kavitha, D.Vanathi "A Study Of Privacy Preserving Data Mining Techniques" IJCAIT, 2014. Pp 71- 77.
[10] Yaping Li, Minghua Chen, Qiwei Li , Wei Zhang "Enabling Multi-level Trust in Privacy Preserving Data Mining" IEEE, 2011. Pp 1-20.
[11] Somayyeh Seifi Moradi, Mohammad Reza Keyvanpour "classification and evaluation the privacy preserving distributed data mining techniques" Journal of Theoretical and Applied Information Technology, 2012. Pp 204-210.
[12] M. Kantarcioglu, R. Nix "Incentive Compatible Distributed Data Mining" Proc. IEEE Int'l Conf. Soc. Computing/IEEE Int'l Conf. Privacy, Security, Risk and Trust, Pp 735-742, 2010.
[13] M. Kantarcglu, C. Clifton "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data" IEEE Trans. Knowledge and Data Eng., vol. 16, no. 9, 2004, Pp 1026-1037.
[14] H. Kargupta, K. Das, and K. Liu "A Game Theoretic Approach toward Multi-Party Privacy-Preserving Distributed Data Mining" Proc. 11th European Conf. Principles and Practice of Knowledge Discovery in Databases, 2007, Pp 523-531.
[15] S. Mukherjee, H. Kargupta "Distributed Probabilistic Inferencing in Sensor Networks using Variational Approximation". J. Parallel Distrib.Comput., Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE Std. 802.11, 1997.