INTERNATIONAL JOURNAL OF PURE AND APPLIED RESEARCH IN ENGINEERING AND TECHNOLOGY

A PATH FOR HORIZING YOUR INNOVATIVE WORK

A REVIEW ON PRIVACY PRESERVATION AND COLLABORATIVE DATA MINING WITH ASSOCIATION RULE SHARING MODEL

MISS. YAMINI M. BABNEKAR, PROF. SHEETAL S. DHANDE

Dept. of Computer Science & Engineering, Sipna College of Engineering, Amravati.

Abstract: As the use of information and internet is widely used, the disclosure of information and its misuse needs to be providing with some security mechanism. The Privacy Preserving Data Mining (PPDM) system is a concern to these respects. The vertically partitioned data available with the parties involved which can provide exact mining results when using the collaborative data mining framework from huge database. To overcome the privacy issue in data disclosure this paper we use a Key Distribution-Less Privacy Preserving Data Mining (KDLPPDM) [1] system in which the publication of local association rules generated by the parties is published. The association rules are securely combined to form the combined rule set using the Commutative RSA algorithm and for data security in data mining. The combined rule sets established are used to classify or mine the data. The result of this paper performs the accuracy of the rules generated using the C5.0 based on the KDLPPDM system [1]. The KDLPPDM provides more security by not generating the private and public key that provided by past cryptographic approaches.

Keywords: Privacy Preserving Data Mining, Vertically Partitioned Data, Association Rules, C5.0 Algorithm, Commutative RSA.

Corresponding Author: MISS. YAMINI M. BABNEKAR

Access Online On:

www.ijpret.com

How to Cite This Article:

Yamini M. Babnekar, IJPRET, 2016; Volume 4 (9): 179-185

PAPER-QR CODE

INTRODUCTION

In different business or company homes and government bodies possess sure framework or infrastructures for maintaining brooding again knowledge collections for watching and processing it. The assorted knowledge extracted from its native or confined databases aren't comfortable to full pity given accomplishing the specified results. Therefore, such shortcomings do require a system or platform that could effectively collect the massive distributed knowledge and may perform the data mining to want information that would be analyzed expeditiously and exactly provide desire result, Such a situations place forth the requirement for privacy preservation in data processing (PPDM) systems. The main objective of the PPDM systems is to keep maintains the integrity of the data revealed and to realize economic data processing results. The privacy preservation in knowledge mining systems that presently exist secure the info exploitation varied mechanisms. Cryptography is the most widely used mechanism to increases data integrity as well as security[10]. Cryptography techniques in which the key generation and key exchange functions are performed. But now a days cryptography technique becomes problematic because as a result of this method has some disadvantages. To beat this disadvantages and enhancing the adversaries ineffective in this paper we tend to introduce a Key Distribution-Less Privacy Preserving Data Mining system (KDLPPDM). The accepting of the optimal data mining technique is also critical and must facilitate accurate analysis with the use of limited data which not provide exact solution and results of information.

The C5.0 data processing algorithm is taken into account for mining within the KDLPPDM system. (Key Distribution-Less Privacy Preserving Data Mining). This paper introduce Key Distribution-Less Privacy conserving data processing system techniques within which there is no key exchange mechanism and independent RSA algorithmic program which makes it strong even the presence of adversaries and will increase privacy on knowledge, similarly overcomes disadvantages of cryptography techniques. The Proposed KDLPPDM technique that considers a completely different protocol low-level formatting phases and earlier analysis has been targeted particularly on coping with privacy preservation in exceedingly centralized information. The Privacy conserving in data processing are primarily used on the tasks of classification, association rule mining and association rule mining perform varied action. The Association analysis involves the creation of associated rules, showing completely different attribute numerous worth and varied conditions that occur often in an exceedingly given set of data [2].

LITERATURE REVIEW

Many have contributed for the economical use of Privacy Preservation of information mining we tend to discuss some of the contributions given by Sweeney et al., [3] proposed a theme called k-anonymity approach which was after improved by Yu et al., [9].

In the partial data activity system the info that is to be revealed for data processing operations are reworked in such the simplest way that normally doesn't affected the results of mining operations and affected on desired result. Xindong Chinese et al.,[5] Tomasz Bujlow et al.,[6] and [4] thought of algorithmic program of C5.0 data processing for knowledge extraction and mining for enhancing higher accuracy and will increase potency of information, security of knowledge. The researchers had enforced C5.0 algorithmic program and this algorithm that perceive the performance of this method as compared to its predecessors with work done by researchers Xindong et al.,[5] and Ming-Jun et al., [7] UN agency introduced 2 algorithmic program ID3 and C4.5 severally for accomplishing classification. The distinctiveness, correctness and therefore the potency of the privacy preservation data processing technique that not solely provides security exploitation independent algorithms however also take away the burden arising as a result of key distribution reasons is essential straightforward use by unauthorized user, key assortment or storage overheads and re-keying method that was more thought of in Karst, N.J et al., [8]. this method performed well however broad functions with strong classification and rule generation techniques like C5.0 was lacked that's why not get actual result and even independent RSA quite approaches couldn't be bestowed in their work that not offers desired result. In this paper planned an extremely strong and economical system known as Key Distribution-Less Privacy preserving data processing (KDLPPDM) that not solely reduces the procedure complexness and overheads however additionally enhances the privacy and accuracy.

DIFFERENT TECHNIQUES USED

A. *Privacy Preservation in Data Mining (PPDM)*

The PPDM systems that currently exist secure the data using varied mechanisms. Cryptography is the most commonly used mechanism to achieve data integrity. To incorporate cryptographic techniques key generation and key exchange is an integral function to be achieved where in adversaries can benefit if improper techniques are adopted. To overcome this drawback and render the adversaries ineffective in this paper we introduce a Key Distribution-Less Privacy Preserving Data Mining (KDLPPDM) system. The adoption of the optimal data mining technique is also critical and must facilitate accurate analysis with the use of limited data. The C5.0 data

181

mining algorithm is considered for mining in the KDLPPDM system. The KDLPPDM system discussed in this paper considers no key exchange to establish the Commutative RSA algorithm which makes it robust even in the presence of adversaries. The Proposed KDLPPDM technique considers a multiple stages based function along with the system or protocol initialization phase. Earlier research has been predominately focused on dealing with privacy preservation in a centralized DB. The PPDM systems are mainly used on the tasks of classification, association rule. Association analysis involves the discovery of associated rules, showing attribute value and conditions that occur frequently in a given set of data.

B. Distribution-less privacy preservation Data Mining

The system initialization phase discusses the C5.0 algorithms along with it optimistic implementation with Commutative RSA based element based score matrix implementation for privacy preservation. Many researchers proposed many methods for privacy preserving association rule mining for both centralized and distributed databases for preserving huge data and its keep maintain security. This paper also describes the different dimensions of preserving data.

C. Vertically partitioned data

We present privacy preserving algorithm to mine association rules from vertically partitioned data that vertically divide various data for privacy concern [7]. The main goal of the privacy preservation is protecting the sensitive information in data while extracting knowledge from large amount of database as well as maintains security and privacy of data because privacy and security is major issue. Privacy is crucial issue when one wants to make use of data that involve individual sensitive information not expose to others or with unauthorized access. It present a privacy preserving algorithm to mine association rules from vertically partitioned data which partitioned data for easily manipulated and analyze the distributed data, by vertically partitioned, we tend to mean that each site contains some parts of a dealings of knowledge in which data is distributed not centralized. The vertically partitioned data available with the parties involved which can provide accurate mining results when using the collaborative data mining framework.

D. Association Rule Mining

The Association rule mining could be a technique in data processing that realize the regularities found in giant volume of knowledge moreover as associated varied data that already partitioned off. Association rule mining such a way might establish and reveal hidden data

that's non-public for a personal or organization. Association rule mining is used to find correlations between items sets efficiently. In the process of data mining the generation of association rule refers towards investigating the inter-relationships among numerous data sets available in datasets of the each individual parties. Here in this paper and the proposed KDLPPDM model the expected and exact results generating from association rules states the relationships amongst the datasets of participants or parties and demographic information etc.

E. Collaborative Data Mining

The major goal of the collaborative data mining with different parties to come together. In the proposed KDLPPDM system is to achieve better data mining results so as to establish useful analysis and conclusions from the data partitioned in them. We are using the collaborative filtering approach, with the help of which one can perform data mining efficiently using the vertically partitioned data. The objective of the collaborative data mining it better when compared to the metric, with respect to those solutions that cannot be achieved using non collaborative data mining. The solutions require evaluation, comparison and approaches for combination. Collaboration requires communication and implies some form of community of data. Data is distributed in various places either different site, data warehouse, server, needed to combined those data collaboration of data mining help to collaborate the data from different places.

F.   Key - Distribution- Less Privacy Preserving Data Mining (KDLPPDM)

The expose of information and its misuse of the data in Privacy Preserving Data Mining (PPDM) systems is a concern to the parties involved. To overcome the privacy issue in data disclosure this paper describes a Key Distribution-Less Privacy Preserving Data Mining system in which the publication of local association rules generated by the parties is published. This paper introduces the Key Distribution-Less Privacy Preserving Data Mining model to preserve the privacy and provides an environment to attain desired knowledge extraction in which the data is not shared for deriving knowledge gather the rules generated at each party are shared through cryptographic means for   data mining.

CONCLUSION

In this paper, we proposed the KDLPPDM System which overcomes the overheads arising due to key exchange and key computation by adopting the Commutative RSA cryptographic algorithm. By using the KDLPPDM system we try to secure the data and preserve its privacy much more effectively than the key distribution algorithm. The parties are reluctant to share

183

the original local data amongst one and other, hence the proposed KDLPPDM believes in sharing the locally generated data mining rules providing for enhanced privacy preserving features. The use of association rule mining using the C5.0 data mining algorithm is adopted in the KDLPPDM to achieve higher mining accuracy, higher speed and for rule generation. By using the vertically portioned data and the accuracy of the association rule mining algorithm, the collaborative data mining can be possible leads to the better performance.

ACKNOWLEDGMENT

REFERANCES

1. Association Rule Sharing Model for Privacy Preservation and Collaborative Data Mining Efficiency. S KumaraSwamy1,Manjula S H1, K R Venugopal1, Iyengar S S2, L M Patnaik3

2. Department of Computer Science and Engineering 1University Visvesvaraya College of Engineering, Bangalore University, Bangalore

3. C5.0 Algorithm to Improved Decision Tree with Feature Selection and Reduced Error Pruning. Rutvija Pandya Diploma Computer Engineering Department, Gujarat Technological University Atmiya Institute of Tech & Sci Rajkot.

4. L.Sweeney, "k-anonymity: A model for protecting privacy," International Journal of Uncerainty, Fuzziness and Knowledge-Based Systems (IJUFKS), vol. 10, 5, pp.557-570, 2002

5. http://www.rulequest.com/see5-in

6. Xindong Wu, Vipin Kumar, J.Yang, Hiroshi Motoda, Geoffrey Philip S. Yu, et al., "Top10 algorithms in data mining" Knowledge and Information System.vol.14, 2008.

7. Tomasz Bujlow, Tahir Riaz,"A method for classification of network traffic Algorithm," in Workshop on Computing and Networking IEEE, pp. 237-240,2012.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

8. Privacy Preserving Association Rule Mining in Vertically Partitioned Data. Jaideep Vaidya Department of Computer Sciences Purdue University West Lafayette, Indiana jsvaidya@cs.purdue.edu

9. Karst, N.J., Wicker, S.B.," On the Rekeying Load in Group Key Distributions Using Cover-Free Families," IEEE Transactions on Information Theory, vol. 58 , no 10, pp. 6667 - 6671 ,2012.

10. C. C. Aggarwal and P. S. Yu, "A condensation approach to privacy preserving data mining," in Proceedings of the International Conference on Extending Database technology (EDBT), vol. 2992, pp. 183- 199,2004

11. William Stallings, Cryptography and Network Security: Principles and Practice (5th Edition), Prentice Hall 2010