A Survey on Privacy-Preserving Mining of Association Rule on Databases

Asavari G.Smart[1], P. M. Mane[2]

[1]Student of Dyanganga College of engineering, Savitribai Phule Pune University, India
[2]Assistant Professor at Dyanganga College of engineering, Savitribai Phule Pune University, India

Abstract: ***Many large businesses share their data, outsourcing for their business problem. In the field of data mining Privacy Preserving Data Mining (PPDM) is a research area related with the protecting private data. Nowadays, privacy preserving has widely used field of research. In the area of data mining PPDM is important field of research. In PPDM, various methods and techniques have been used till date. In this paper, we provide survey on Privacy Preserving Association Rule Mining. This is popular technique in pattern discovery methods in the field of privacy preserving data mining. In recent years, many algorithms have been proposed. We divide the proposals of privacy preserving association rule mining into three categories: reconstruction-based techniques, heuristic-based techniques, cryptography-based techniques. Finally, we conclude further research directions of privacy preserving algorithms of association rule mining by analysing the existing work.***

Keywords: privacy preserving algorithm, association rule mining, data mining, security, private data

1. Introduction

In the past two decades the scope of information technologies and the internet has brought attention of information into the hands of commercial companies. Data owners constantly seek to make better use of the data they possess, and utilize the data mining tools to extract useful knowledge from the large amount of data [1]. Privacy preserving data mining has become a most popular in data mining research. The main reasons behind it are the consequence of private data, boost technology and easy storage.[8]. In privacy preserving data mining, association rules are beneficial for examine and forecasting customer behavior and pattern of purchase or order. It plays an important part in market analysis, data of basket shopping, clustering of items, classification, and design of catlogs and store layout.

The association rule mining technique has received a great focus of attention since 1993 [2]. It is one of most promising pattern-discovery methods in the field of data mining. In recent years, many proposals and algorithms have been designed for it. At the same time, data mining algorithms are used for analyzing for the side-effects which incur in data privacy.

Thus, in the past few years several privacy-preserving techniques for association rule mining have also been introduced. Association rule mining is a technique in data mining that find regularities in large volume of data. Such a technique may identify data which is private from human or organization [3].

Association rule mining is a technique in data mining that recognizes the regularities found in data Privacy-preserving data mining using association rule turn in to the area of data mining that attempt to find protection for sensitive information from uninvited revelation.

By going through it in depth, in some of the algorithms with association rule mining, some techniques also can be applied to other data mining computations, such as decision tree inducers, association rule mining algorithms, clustering algorithms, and Bayesian networks etc. In this paper, we describe overview of privacy preserving association rule mining. The arrangement of rest paper is as follows: Section 2 classification of privacy preserving mining and introduces some secure association rule mining strategies; Section 3 lists and classifies some of these techniques; Conclusion.

2. Literature Survey

Objective of privacy preserving data mining is to develop an algorithm in such a way that private data and patterns, knowledge should remain same even after mining process. Various techniques are developed for privacy preserving. Many approaches are available for privacy preserving data mining but following dimensions are considered here to classify.

1. Data Distribution

This refers to a distribution of data. It can also be classified in to two types:

- Horizontal data distribution

In this type different database records re reside in different places [9].

- Vertical data distribution

This distribution refers to the cases where all the values for different attributes reside in different places [10].

1. Data modification

This refers to the data modification scheme. Modification is used to modify the original values of a database that needs to be released to the public and in this way to guarantee maximum privacy protection

2. Data mining algorithm

The data modification is taking place for the data mining algorithm. And this actually does the work of analysis and the design of the data hiding algorithm.

3. Data or rule hiding

The fourth dimension refers to what kind of data is to be protected, raw or aggregated. Protecting aggregated data in the form of rules is obviously more complex.

A huge amount of implementation of the privacy of data and knowledge are applied in association rule mining method. At present, privacy preserving association rule mining algorithms are divided into three categories according to privacy protection technologies [4].

A. Heuristic Based Technique

Heuristic based techniques generally used for the complexities issues. This technique solves the problem of selection of dataset to proper data modification. This technique generally proposed for the modification of data. For that, some methods are proposed called as data distortion method. It uses data distortion technique for the modification of confidential data [5].

B. Re-Construction Based Association Rule Mining

In order to perform the association rule mining many recently techniques has faces problem of privacy preserving. So these algorithms are implemented for perturbing data and then reconstruct the distribution. To deal with categorical data, Agrawal et al. in 2002 proposed a privacy protection approach on reconstruction-based association rule [7].

C. Cryptography-Based Techniques

The cryptography method is generally used for data encryption. There are many Cryptography-based approaches proposed to preserve the privacy of preserving data mining algorithms. Secure Multi-party Computation (SMC) approach, which is actually a Cryptography-based, is secure at the end of the computations. No party is aware of anything except its own input and the results. SMC method is a typical method [6].

3. Requirements of a PPDM Algorithm

A. Accuracy

The accuracy is nearly identified with the data loss coming about because of the hiding technique: the less is the data loss; the better is the information quality. PPDM algorithm has need to keep up high precision to lessen data loss[1].

B. Completeness and Consistency

Completeness assesses the level of missed information in the sterilized database. Deficient information has a huge effect on information mining comes about and weakens the information mining calculations from giving an exact representation of the underlying information [1].

C. Scalability

It is an alternate critical perspective to survey the execution of a PPDM algorithm. Specifically, scalability depicts the proficiency patterns when information sizes increment [1].

D. Data quality

It is a vital part of PPDM. High quality information that has been arranged particularly for information mining assignments will bring about valuable information mining models and yield. On the other hand, low quality information has a noteworthy negative effect on the utility of information mining results[1].

E. Security

It is the level of insurance against harm, loss of data, and wrongdoing. There are two principle methodologies with respect to how to manage the issues of security that emerge today. The primary is a lawful and arrangement approach whereby associations are constrained by the way they store and utilization information focused around protection law and open strategy. It commonly lives up to expectations by assessing situations and choosing if the security rupture brought about by utilizing the information within a given way is advocated or not. The second approach is innovative, and gives authorized protection ensures through cryptographic means. This methodology has the capacity of empowering the information to be utilized while averting protection ruptures [1].

4. Conclusion

In this paper, we introduced diverse PPDM procedures, necessities and issues and emphasize innocent protection protecting strategies to convey ones and the routines for taking care of on a level plane and vertically partitioned information. While all the purposed strategies are just estimated to our objective of privacy preserving, we have to further impeccable those methodologies or create some proficient systems.

The work presents here, which demonstrates the always expanding enthusiasm of analysts in the range of securing touchy information and learning from malevolent clients. At present, protection saving is at the phase of improvement. Numerous privacy preserving algorithm of association rule mining are proposed, on the other hand, privacy preserving algorithm innovation needs to be further complex due to the multifaceted nature of the security issue. We close three examination bearings of privacy preserving association rule mining by examining the current work later on.

Acknowledgement

This research is supported by College authority and guide as well as University of Pune. We are also thankful for support of all family and friends.

References

[1] R.Natarajan, Dr.R.Sugumar, M.Mahendran, K.Anbazhagan, "A survey on Privacy Preserving Data Mining", International Journal of Advanced Research in Computer and Communication Engineering Vol. 1, Issue 1, MARCH 2012.

[2] Agrawal R, Imielinski T, Swami A, "Mining association rules between sets of items in large databases," In:Buneman P, Jajodia S, editors.Proceedings of ACM SIGMOD conference on management of data.Washington, DC, 1993, pp.207–216.

[3] Agrawal D., and Aggarwal C.C (2007), "On the Design andQuantification of Privacy Preserving Data Mining Algorithms‟,Proceedings of the 20th ACM Symposium on Principles of DatabaseSystems, pp. 247-255.

[4] Vassilios S. Verykios, Elisa Bertino, et al., "State-of-the-art in Privacy Preserving Data Mining," SIGMOD Record, Vol. 33, No. 1, March2004, pp.50-57.

[5] Agrawal, R., and Srikant (2007), "Privacy Preserving Data Mining‟,Proceedings of the 19th ACM International Conference on KnowledgeDiscovery and Data Mining, Canada, pp. 439-450.

[6] C.Clifton,MuratKantarcioglou,XiadongLin,andMichaed Y.Zhu,"Tools for privacy preserving distributed data mining," SIGKDD Explorations 4 (2002), no. 2.

[7] SrikantR,Agrawal R, et al. , "Privacy preserving miningof association rules," In: Proc. of t he Eighth ACM SIGK2DDInternational Conference on Knowledge Discovery and Data Mining,ACM Press,2002, pp.217–228.

[8] F. Giannotti, L. V. Lakshmanan, A. Monreale, D. Pedreschi, and H. Wang, "Privacy-preserving outsourcing of association rule mining,"ISTI-CNR, Pisa, Italy, Tech Rep. 2009-TR-013, 2009.

[9] M. Kantarcioglu and C. Clifton, "Privacy-preserving distributed miningof association rules on horizontally partitioned data," *IEEE Trans.*Knowledge Data Eng., vol. 16, no. 9, pp. 1026–1037, Sep. 2004.

[10] D. W.-L. Cheung, V. Ng, A. W.-C. Fu, and Y. Fu.Efficient mining of association rules in distributed databases. Transactions on Knowledge and Data Engineering, 8(6):911–922, Dec. 1996.

Author Profile

Asavari Smart received the B.E. degree in Computer Engineering from KIT's College of Engineering Kolhapur in 2009 and pursuing M.E during 2013-2014.