www.ijecs.in
International Journal Of Engineering And Computer Science ISSN:2319-7242
Volume 4 Issue 5 May 2015, Page No. 11845-11849

Non-Linear Data Perturbation Method for Privacy Preserving Data Mining

Dr.M.Gopichand
Professor &Head,Department of IT,
Vardhaman College of Engineering,Shamshabad
A.P,INDIA.
Email:gopi_merugu@yahoo.com

ABSTRACT

To hold effective data from big databases, Data Mining Methodologies are put on purpose. For recognizing the connections amid the item sets quickly, academicians and researchers are giving attention on the association rule mining amid many different data mining approaches. The subject of privacy matters when the data is spread amongst number of places or sites. The site owner doesn't desires to supply their data to other sites. But they are involved to have information about global or whole results existed from mining actions. A new representation is suggested which adopts mix up based protected calculation cryptographic practice considering that whichever locale or site can be dealt as Faithful site to find global association rules for horizontally divided databases.

KEYWORDS: Association rule mining, Big-Database, Confidentiality, Key dissimilarity, Frequent item sets.

I. INTRODUCTION

Possessing the opinion of the motivation to include privacy in data mining methods to defend the undisclosed data of the client, there gradually developed an innovative flow into data mining, which is protecting the privacy in the data mining. There exists a key dissimilarity amid usual data mining algorithms underneath a range of data mining techniques that is the proper algorithms deal with how to scrutinize the stored raw data and how to extract useful knowledge discovery patterns from the database, whereas in the later one, it primarily deals with the perceptive information of the user records where privacy aspect is the chief distress and it is measured to be central problem.

The core endeavour in many methods in the distributed environment for protecting the privacy in the data mining will be to allocate practical summative computations on the complete data collection by preserving different sites private information. Each site possessor is fascinated to work in partnership in acquiring mutual results, but not wholly having faith in other sites when it comes to the division of their individual data group. Any data mining scheme should persuade the imperative property that is privacy preserving of data/information. Principally in distributed data mining, privacy preserving is solitary fundamental aspect. For safeguarding the privacy in distributed data mining the method secure multi party computation has been used which will be a valuable approach.

To get hold of commonly favourable large-scale data mining objectives and to protect privacy in data mining makes use of an algorithm inside mining, exclusive of illuminating private data. For that reason, in several data mining applications confidentiality preserving has happen to an imperative concern.

Diverse lay down of tuples by way of alike set of attributes, in a database which is distributed in horizontal manner of whole database are positioned next to dissimilar sites. The relations which link items or which link item sets are able to be initiated truthful only on condition that rules are resoluted as an outcome of whole group of tuples commencing each site. And no data by any site owner desires to present and this creates the problem a tough thing that is extracting crucial information on or after every sites information exclusive of accessing individual records toward generating association rules. The proposed process solves this trouble via a cryptography procedure that is Mix up based secure sum along with Faithful site.

II. RELATED WORK

For the researchers, to ascertain correlations amid items or item sets, and to receive further attention in between all the data mining techniques, Association rule mining technique is used. The privacy is achieved by dividing data into multiple levels and encrypting it [1].

By the Association rule mining the matter of performance and demonstrating the issues involved with building highly-efficient methods, were discussed [2,3] additionally. Along with the inclusion of the notations of secure multi-party computations [4], the study of their importance to the ground of primary prototype and protection of privacy in data mining are offered through the authors included.

For analyzing the algorithms related to privacy protection in data-mining [5], the authors suggested an outline which is centered on the measurement of the diverse properties of these algorithms which are privacy protecting, is done in a way which is diverse assessment methods. Using cryptographic way [6], and to lessen the data spread as a result of totaling the burden towards the process of mining is introduced, through association rules by means of protected mining in a database which is partitioned in horizontal manner.

An enhanced form from kantarcioglu and Clifton's systems have been proposed by authors [7], which really suits as a stage for securing privacy in the area of distributed data mining. Couple of practices which are introduced by them, with or without reliable party, used to boost the safety against consent in the environment where the data is exchanged.

The alteration of the existing algorithm which is an innovative algorithm, furthermore, according to a model through minimal probability regarding collision is projected in [8]. Cryptographic ways were used by the authors, which is moreover used for the preservation of the privacy [9].

By establishing the benefits of the initial approach they suggested an innovative solution which safeguards the privacy of the data to get a grip on the method and the second way which utilizes secure cryptography solutions by having an expanded position based access along the method of reducing thrashing of privacy along with information.

Writers deal with the situation of association rule mining in a database in divided in vertically partitioned manner [10], which is done by utilizing cryptography-based way and also gift the study of communication as well as protection. The annals of safe multiparty calculation in two milliners problem, to know whoever is wealthier exclusive of revealing their prosperity is demonstrated. Practices may be anticipated in twos milliners trouble along with multi party situation.

III. PROPOSED SYSTEM

The following subsection specifies the summary and Particulars of the proposed scheme which diverse phases in the proposed design.

III. a. Summary of Proposed System

Whenever the unique site named faithful site is there, the proposed scheme deals with the databases which are partitioned in a horizontal manner, which is having very advanced level of privileges or rights to obtain and accomplish the most assured or guaranteed responsibilities. Each part of the location in the proposed situation helps to produce the association rules which are global. These rules also produce recurrent or item sets, which are frequent as well as global. No site is all set to picture their item sets which are frequent, the size of the database and the overall supports towards the faithful site as well as the site owner.

For this kind of problem to be solved, the privileges which are unique are furnished through the method of the Faithful Site, in order to wholly grab regional item sets which are frequent without

even captivating the price of price from each and every site to summarize every single one site item sets which are frequent. Each site worker welcomes to present regional recurrent itemsets during unreadable structure to Faithful party to whoever the people trust to create combined recurrent list of itemsets. Mix up based protected calculation cryptographic practice was adopted by the proposed model to discover total association rules as a result of securing the confidentiality.

III. b. Particulars of the proposed scheme

In distributed environment while records are partitioned horizontally in the midst of dissimilar sites with a Faithful site is considered in this document. The diverse phases in the proposed design are follows:

Phase 1: The initial course of action is initiation, which is fulfilled by the Faithful site, plus it also sends appeal to learn frequent item sets in the direction of all sites by delivering public-key, Minimum supports.

An item set's Minimum Support can be referenced as the proportion of communications or transactions which hold the itemset, in the dataset.

Phase 2: All sites see nearest item sets which are frequent, in acknowledgment to the public key. Using the public key every site applies encryption algorithm for the newly produced set of item sets, which are frequent, to translate frequent item sets into encrypted form and send it to Faithful site.

Phase 3: By means of Private key, Faithful site then decrypts the every site's encrypted information and creates a combined list consisting of each site's restricted recurrent itemsets following elimination of duplicates.

Phase 4: For each item set, each site calculates partial support which is in the merged list which is predictable from Faithful site. For all the article sets each site then broadcasts to all other sites, its computed values listed in the merged list.

Phase 5: Total of all computed standards of every site is computed and sent to the Faithful site, for each item set.

Phase 6: For all item sets, whole addition from all the sites is received by Faithful site.

Phase 7: The distinctiveness of well-known Computed values as of all sites is confirmed by Faithful site. Faithful site stresses the entire owners to employ phase five one more time to acquire the acceptable results, if any incongruities are present.

Phase 8: Global Support (GS) value for each item set is computed by Faithful site. The item set is globally frequent, if the evaluated value of GS is greater than zero, otherwise it is globally infrequent.

Phase 9: Faithful site finds Actual Support for each global frequent item set, and transmits a catalog which includes their values to each site along with all total frequent itemsets.

Phase 10: With the aid of the globally frequent item sets all sites can produce correlation rules with diverse confidence values and cling to support values received from Faithful site.

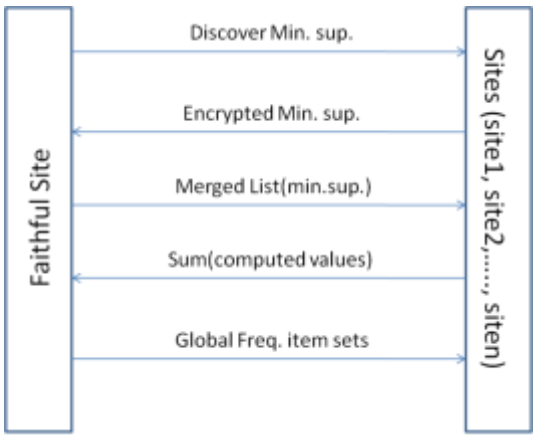

Figure 1: Sites communication in proposed model (Min. sup. = Minimum Support)

IV. RESULTS

The proposed solution is implemented in JAVA. The Mix up based secure sum concept which is employed in the computation of partial supports enhances the privacy by Partial Supports which are in disguised form. And they are broadcasted to the sites securely.

All sites are not having any idea about the Mix up, random number which is assigned by Faithful site

to new sites and the database size of additional sites is also not recognized as well. The advantages of the proposed model are discussed below:

1.Faithful Site discovers the minimum support or the local frequent item sets of each site in an encrypted manner which can also be decrypted. So, due to the encrypted format, the Faithful Site cannot know the supports regarding any item sets related to sites.

2.By using the mix-up based protected calculation method, we can summarize the partial supports of the item sets. Now the partial supports are in mixes up form with unknown random numbers and symbols, which make the sites impossible to discover or reveal the size of the entire database or any other private information. So privacy to the individual data can be preserved by broadcasting the partial supports to each site.

3.Faithful site also cannot estimate the user's private data or database size, etc, after receiving total partial supports from the item sets.

4.Global frequent item sets, which are finally broadcast by the Faithful Site, to each site, are secured and the site also cannot predict the local support of whichever total recurrent itemsets.

The number of transactions and data transfer leads to the rise in the cost in communications in the distributed system environment.

5.The Faithful Site sends bulk data transfer at a time, hence reducing the number of transactions which is the indication of efficiency.

V. CONCLUSION

The consideration of difficulty in protecting privacy in association rule mining whenever the database is distributed or divided in a horizontal manner amid various quantity of sites by way of a Faithful site.

A new form of method is anticipated in this document which makes use of a method called a Mix up based secure sum cryptography procedure for finding the association rules which are global, exclusive of opening ones private details. The functionality of the proposed model is illustrated with a picture diagram.

The association rules which are global in the proposed system is broadcasted along with the global frequent item sets and with lowest amount of interactions.

VI. REFERENCES

[1] Enabling Multilevel Trust in Privacy Preserving Data Mining Yaping Li, Minghua Chen, Qiwei Li, and Wei Zhang. IEEE transactions on knowledge and data engineering, vol. 24, no. 9, september 2012.

[2] R Agarwal, T Imielinski and A Swamy, Mining Association Rules between Sets of Items in Large Databases, Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, page 207-210, 1993.

[3] Verykios, V.S., Bertino, E., Nai Fovino, I., Parasiliti, L., Saygin, Y., and Theodoridis, Y. 2004 . State-of-the-art in privacy preserving data mining. SIGMOD Record, 33(1):50–57.

[4] Y. Lindell and B. Pinkas, Secure Multiparty Computation for Privacy-Preserving Data Mining, The Journal of Privacy and Confidentiality (2009) , 1, Number 1, pp. 59-98.

[5] Elisa Bertino , Igor Nai Fovi no Loredana Parasiliti Provenza ,A Framework for Evaluating Privacy Preserving Data Mining Algorithms, Data Mining and Knowledge Discovery, 2005, 11, 121–154.

[6] M. Kantarcioglu and C. Clifto. Privacy-preserving distributed mining of association rules on horizontally partitioned data. In IEEETransactions on Knowledge and Data Engineering Journal, volume 16(9), pages 1026–1037.

[7] Chin-Chen Chang, Jieh-Shan Yeh, and Yu-Chiang Li, Privacy-Preserving Mining of Association Rules on DistributedDatabases, IJCSNS International Journal of Computer Science and Network Security, VOL.6 No.11, November 2006.

[8] Mahmoud Hussein, Ashraf El-Sisi,and Nabil Ismail, Fast Cryptographic Privacy Preserving Association Rules Mining on Distributed Homogenous Data Base, I. Lovrek, R.J. Howlett, and L.C. Jain (Eds.): KES 2008, Part

II, LNAI 5178, pp. 607–616, 2008.© Springer-Verlag Berlin Heidelberg 2008.

[9] Lalanthika Vasudevan , S.E. Deepa Sukanya, N. Aarthi ,Privacy Preserving Data Mining Using Cryptographic Role Based Access Control Approach, Proceedings of the International MultiConference of Engineers and Computer Scientists 2008 Vol I, IMECS 2008.

[10] Vaidya, J. and Clifton, C. 2002. Privacy preserving association rule mining in vertically partitioned data, 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM Press, pp. 639–644.

[11] A.C. Yao. Protocols for secure computations. In Proceedings of the 23rd Annual IEEE Symposium on Foundations of Computer Science, 1982.