International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Impact Factor (2012): 3.358

A Review of Privacy Clustered Mining Of Association Rules In Distributed Databases

Vanita Babane[1], Shital K.Somawar[2]

[1]Professor, Department of Computer Engineering, RMD Sinhgad School of Engineering, University of Pune, India

[2]Department of Computer Engineering, RMD Sinhgad School of Engineering, University of Pune, India

Abstract: ***With the use of protocol, secure mining of association rules in distributed database can be takes place. Some of the protocols based on the FDM(fast distributed mining) which consist of unsecured version of Apriori Algorithm. Thus different ways are presented by various researchers to overcome the problem of unsecure mining of association rules. In this paper a protocol containing two novel secure multiparty algorithms are proposed to overcome these problems.***

Keywords: Privacy preserving data mining, Distributed Computation, Association rules

1. Introduction

Here there is problem of secure mining of association rules in distributed databases. In that there are various sites that holds the homogeneous databases where databases contain same schema holds the information of different entities. Where the inputs are partial databases and the output is the list of association rules that hold in unified database with exceeding level of support and confidence. The aim is to find out association rules with predefined level of support and confidence and also protect the content of the information which is not only local but also more global.

So that to overcome the problem of security another protocol is proposed for secure computation of union of private subsets. The proposed protocol improves upon that in terms of simplicity and efficiency as well as privacy. In particular, our protocol does not depend on commutative encryption and oblivious transfer(what simplifies it significantly and contributes towards much reduced communication and computational costs). While our solution is still not perfectly secure, it leaks excess information only to a small number (three) of possible coalitions, unlike the protocol of that discloses information also to some single players. In addition, we claim that the excess information that our protocol may leak is less sensitive than the excess information leaked by the protocol.

2. Review of Different Methods

Previous work in privacy preserving data mining has considered two related settings. One, in which the data owner and the data miner are two different entities, and another,in which the data is distributed among several parties who aim to jointly perform data mining on the unified corpus of data that they hold.

In the first setting, the goal is to protect the data records from the data miner. Hence, the data owner aims at anonymizing the data prior to its release. The main approach in this context is to apply data perturbation [2], [11]. The idea is that the perturbed data can be used to infer general trends in the data, without revealing original record information. In the second setting, the goal is to perform data mining while protecting the data records of each of the data owners from the other data owners. This is a problem of secure multi-party computation. The usual approach here is cryptographic rather than probabilistic. Lindell and Pinkas [22] showed how to securely build an ID3 decision tree when the training set is distributed horizontally. Lin et al. [21]. discussed secure clustering using the EM algorithm over horizontally distributed data. The problem of distributed association rule mining was studied in [19], [31], [33] in the vertical setting, where each party holds a different set of attributes, and in [18] in the horizontal setting. Also the work of [26] considered this problem in the horizontal setting, but they considered large-scale systems in which, on top of the parties that hold the data records (resources) there are also managers which are computers that assist the resources to decrypt messages; another assumption made in [26] that distinguishes it from [18] and the present study is that no collusions occur between the different network nodes—resources or managers.

3. Proposed System

Propose an alternative protocol for secure computation of the union of private subsets. The proposed protocol improves upon that interms of simplicity and efficiency as well as privacy. In particular the protocol does not depend upon the commutative encryption and oblivious transfer. While this solution is still not secure, it leaks excess information only to a small number of coalitions, unlike the protocol that discloses the information of single users.

In addition, we claim that the excess information that our protocol may leak is less sensitive than the excess information leaked by the protocol.
There are several modules included:
1.User Module
2.Admin Module
3.Association Rule
4.Apriori Algorithm

1.User Module

In this module, privacy preserving data mining has considered two related settings. One, in which the data owner andthe data miner are two different entities, and another, in which the data is distributed among several parties who aim to jointly perform data mining on the unified corpus of data that they hold.

In the first setting, the goal is to protect the data records from the data miner. Hence, the data owner aims at anonymizing the data prior to its release. The main approach in this context is to apply data perturbation. He perturbed data can be used to infer general trends in the data, without revealing original record information.

In the second setting, the goal is to perform data mining while protecting the data records of each of the data owners from the other data owners.

2.Admin Module
In this module, is used to view user details. Admin to view the item set based on the user processing details using association role with Apriori algorithm.

3.Association Rule:
Association rules are if/then statements that help uncover relationships between seemingly unrelated data in a relational database or other information repository. An example of an association rule would be "If a customer buys a dozen eggs, he is 80% likely to also purchase milk."

Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships.Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true.

4.Apriori Algorithm:
Apriori is designed to operate on databases containing transactions. The purpose of the Apriori Algorithm is to find associations between different sets of data. It is sometimes referred to as "Market Basket Analysis". Each set of data has a number of items and is called a transaction. The output of Apriori is sets of rules that tell us how often items are contained in sets of data.

Algorithm - Fast Distributed Mining (FDM)
The FDM algorithm proceeds as follows:
1) Initialization
2) Candidate Sets Generation
3) Local Pruning
4) Unifying the candidate itemsets
5) Computing local supports
6) Broadcast Mining Results

4. Conclusion

As the mining association rules securely is very important. Proposed protocol improves the efficiency and privacy in mining of association rules. Also supports the novel multiparty protocol for computing the union of private subsets that each of interacting user holds. Other ingredient is a protocol that tests the inclusion of an element held by one user in a subset held by another.

References

[1] R. Agrawal and R. Srikant, "Privacy-Preserving Data Mining,"Proc. ACM SIGMOD Conf., pp. 439-450, 2000.
[2] A.V. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy Preserving Mining of Association Rules," Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD),pp. 217-228, 2002.
[3] M. Kantarcioglu and C. Clifton, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data," IEEE Trans. Knowledge and Data Eng., vol. 16, no. 9, pp. 1026-1037, Sept. 2004.
[4] M. Kantarcioglu, R. Nix, and J. Vaidya, "An Efficient Approximate Protocol for Privacy-Preserving Association Rule Mining," Proc. 13th Pacific-Asia Conf. Advances in Knowledge Discovery and Data Mining (PAKDD), pp. 515-524, 2009.
[5] X. Lin, C. Clifton, and M.Y. Zhu, "Privacy-Preserving Clustering with Distributed EM Mixture Modeling," Knowledge and Information Systems, vol. 8, pp. 68-81, 2005.
[6] Y. Lindell and B. Pinkas, "Privacy Preserving Data Mining," Proc. Crypto, pp. 36-54, 2000.
[7] J. Vaidya and C. Clifton, "Privacy Preserving Association Rule Mining in Vertically Partitioned Data," Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD), pp. 639- 644, 2002.
[8] J. Zhan, S. Matwin, and L. Chang, "Privacy Preserving Collaborative Association Rule Mining," Proc. 19th Ann. IFIP WG 11.3 Working Conf. Data and Applications Security, pp. 153-165, 2005.
[9] T. Tassa and E. Gudes, "Secure Distributed Computation of Anonymized Views of Shared Databases," Trans. Database Systems, vol. 37, article 11, 2012
[10] T. Tassa and D. Cohen, "Anonymization of Centralized and Distributed Social Networks by Sequential Clustering," IEEE Trans. Knowledge and Data Eng., vol. 25, no. 2, pp. 311-324, Feb. 2013.

Author Profile

Vanita Babane received the B.E. and M.E degrees in Computer Engineering from MBES COE Ambajogai and VIT Pune in 2005 and 2013, respectively. Currently she is working as Assistant Professor of Computer Engineering Department in RMD SSOE Pune, India

Shital K.Somawar Research Scholar RMD Sinhgad School of Engineering Warje, Pune, University of Pune.She received B.E. in Information technology from Information Technology Department of Tatyasaheb Kore Institute of Technology, Warananagar, Kolhapur from Shivaji University. Currently she is persuing M.E. in computer engineering from RMD Sinhgad School Of Engineering Warje, Pune ,University of Pune.