Literature Survey On Formation Of Association Rule Using Secure Mining

Vidisha H. Zodape, Leena H. Patil

Abstract: Data mining is the automatic extraction of previously unknown patterns from the database. In order to better serve the needs of web based applications and to find the associated data from web, the overview of privacy preserving in mining association rule and the private dataset is given in this paper. Here protocols are proposed to allow secure association rule on distributed database.

————————————◆————————————

I. Introduction

The Web is a vast, volatile and mostly amorphous data repository, which stores incredible amount of data, and also enhance the complexity of how to deal with the information stored in database. Users wish for the tool/search engine which will provide relevant information. Service providers will have to find the techniques to create the web site by minimizing the load to best serve the site to the different users. Business analyst wants the tool to analyze the behaviour of consumer needs. Mining is the process of finding out what users are looking for on the internet, some are interested in document file, and some users are interested in media file or images. This is the technique to find out the interesting usage pattern and best serve the information to the user. Here the method is introduced to form association rule [7] using combined apriori and FP Tree. Data mining techniques have been introduced successfully to retrive knowledge in order to support a variety of domains marketing, weather forecasting, medical diagnosis, and national security. But it is still a challenge to mine the data by protecting the private database of user. Most organizations want information about individuals for their own specific needs. However, different units within an organization themselves share the information. In such cases, each they must be sure that the privacy of the individual is not violated or that sensitive business information is not revealed. In order to provide security, records can be modified before the records are shared with anyone who is not permitted directly to access the data. This can be done by deleting from the dataset some identity fields, such as name and passport number in passenger information record.

II. Literature survey

In paper [1], the work focuses on web usage mining. As web is mostly amorphous data repository, and also enhance the complexity of dealing with the information from the different opinion of view, users, web service providers and business analyst. They have used apriori and improved FP tree to find association rule. Apriori -the classical mining algorithm is a way to find out certain potential, regular knowledge from the massive ones.

_____

- *Vidisha H. Zodape (Mtech IIIrd sem student), CSE Department, PIET, Nagpur, Maharashtra, India*
- *Leena H. Patil (Lecturer), CSE Department, PIET, Nagpur, Maharashtra, India*

Apriori algorithm [13] is the mining of frequent item set and association rule learning[11] over transactional databases. It scans the frequent item sets by scanning the database until those items appear often in database. This is used to find the association rule[7].The FP-Tree Algorithm, is an another way to find frequent patterns [15] without utilising candidate generations[15], therefore improving performance. It uses a divide-and-conquer strategy. The central part of this method is the usage frequent-pattern tree (FP-tree), which keeps the piece set association information. In simple words, this algorithm works as follows:

- It compresses the input database and get an FPtree instance to represent common items.
- It then divide the compressed database into a set of conditional databases, each one associated with one common pattern.
- Eventually, each database is extract separately.

Using this scheme, the FP-Tree decreases the cost to scan database looking for small patterns recursively and then combine them in the long common patterns, in large databases. In paper [2], they have considered the applications on business environment, its benefits is defined by collaboration, team efforts and partnership, rather than individual efforts. So the collaboration is most important because it brings mutual benefit. Sometimes, collaboration even occurs among competitors, or among companies that will give them an advantage over other competitors.
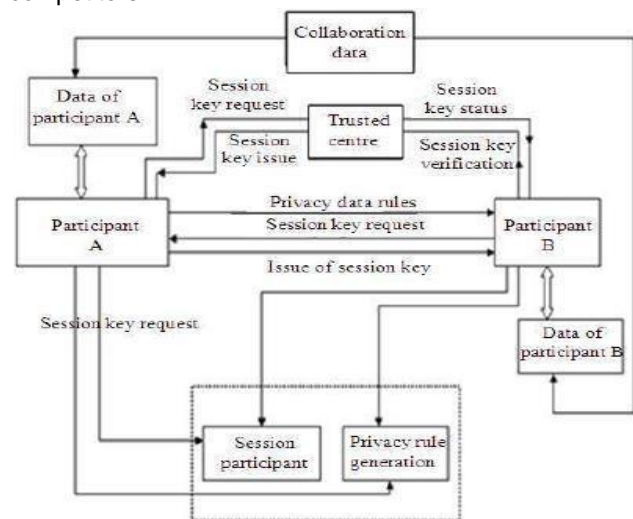

Figure: *Framework of secured multiparty computation process of collaborative data*

222

For this kind of collaboration, where all users want to share the data but want to secure the private data the method used namely Session based Secured Multiparty Collaborative Data Computation (SSMCDM). In this all participants are the authenticated users participating in data mining process and there is trusted third party which will provide session for the participants. In paper [3], the goal is to find association rule with minimum support at least s and minimum confidence at least c and minimizing the information disclose about the private database. They have proposed FDM [8] algorithm to securely mine the association rule on horizontally partitioned database. Initially all parties already having the locally s frequent itemset. Then the candidate set is generated using apriori algorithm. Local pruning is done to compute support and count. And then mining result is broadcasted. Here UNIFI-KC protocol is used to provide the privacy. First all parties will add to their private dataset a fake item set to hide the original size of database. Then all parties together will compute encryption of private subset followed by the computing encryptions of each itemset of subset. Then the union of subset is computed. Finally fake itemset is removed by decryption. In paper [4], they have worked on finding rules on vertically partitioned database where the items are distributed and single entity is split over the different sites.

Figure: *Proposed Privacy Preserving Ranking*

For privacy preserving, distributed ranking is applied on individual sites and the results are merged in order to get final ranking result. The scalar projection of entities is transferred to common sites based on weight vector. Support vector machines optimization is applied to find minimum weighted vector. This is used to find the vector projection and the values are then sent to the coordinate to combine results and ranking by adding all scalar projections from all sites. In paper [5], proposed new protocols which allow distributed subgroup discovery

while providing security of the individual databases. Property describes a prototypical implementation and present experiments that demonstrate the feasibility of the approach. Subgroup discovery technique is used in applications based on fraud detection or clinical studies. It gives top k patterns, which uses quality of functions which is based on relative high accuracy of itemsets. Subgroup discovery [9] is very potential to provide more understanding than numerical methods like SVM or neural network. In supervised descriptive itemsets are sometimes subsumed with the classical techniques. But subgroup discovery does not provide privacy for multi party so cannot secure the information leak. In paper [6], commutative encryption algorithm is proposed for privacy preserving. Privacy may restrict parties to share information where data is distributed at various place. This paper gives cryptographic techniques to mine association rules on horizontally distributed database. Commutative cryptography has two phases. In first one each party encrypts its itemset and pass it to other. This is then passed to other party to encrypt itemset and so on until all parties have encrypted all itemsets. Then common party will receive it to remove duplicates. Finally each party will decrypt each itemset and the result is common itemset.
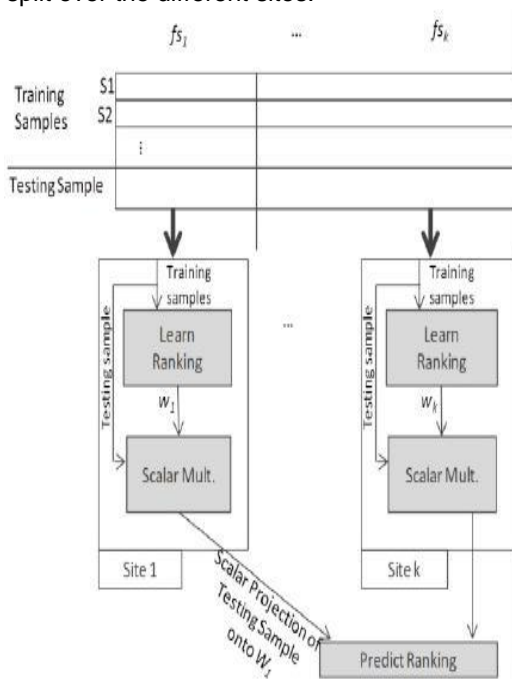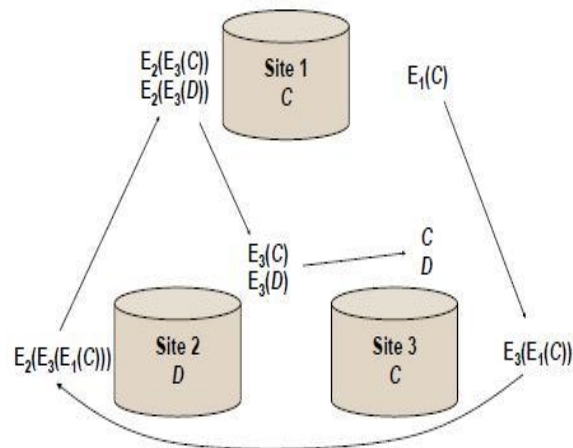

Figure: *Determining Global Candidate Set*

In second phase all locally frequent itemsets are experimented to see if they are globally s frequent. When common itemset is computed first site will choose any value X, and the exceeded amount of support is added to X. This is passed to other party and same is repeated everyone add exceeded support. The result is then compared, if it exceeds value X then it is globally supported.
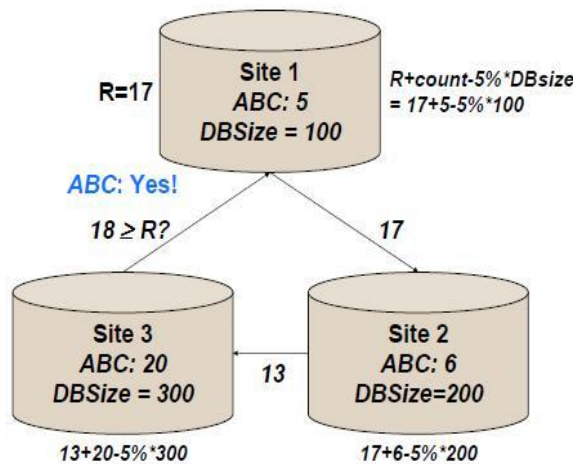
Figure: *Determining If Itemset Exceeds Threshold*

III. Conclusion

The conclusion is that privacy preserving can be applied on certain limitations in data mining algorithm. Some classes of data mining cannot be generalized. Also this survey shows that the research in this area is increasing.

REFRENCES

[1] Ms. Monal saxena "Association rules Mining Using Improved Frequent Pattern Tree Algorithm", International Journal of Computing, Communications and Networking, Volume 2, No.4, October - December 2013

[2] M.J. Freedman, K. Nissim, and B. Pinkas,"Privacy Preserved Collaborative Secure Multiparty Data Mining," Proc. Int'l Conf. Theory and Applications of Cryptographic Techniques (EUROCRYPT), pp. 1-19, 2012.

[3] T. Tassa and E. Gudes, "Secure Mining of Association Rules in Horizontally Distributed Databases," IEEE Trans. Database Systems, vol. 37, 2014.

[4] 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, 2012.

[5] H. Grosskreutz, B. Lemmen, and S. R€uping, "Secure Distributed Subgroup Discovery in Horizontally Partitioned Data," Trans. Data Privacy, vol. 4, no. 3, pp. 147-165, 2011.

[6] T. Tassa and D. Cohen, "Privacy preseving distributed mining of association rules on horizontally partitioned data ," IEEE Trans. Knowledge and Data Eng., vol. 25, no. 2, pp. 311-324, Feb. 2013.

[7] A R "Fast Algorithms for Mining Association Rules", Sep 12-15 1994, Chile, 487-99, pdf, 1-55860-153-9.

[8] Mannila H,"Efficient algorithms for discovering association rules mining." conference Knowledge Discovery in Databases (SIGKDD). 181-83.

[9] Tan, P. N., M. St., V. Kumar, "Introduction to web Mining", Addison-Wesley, 2013, 769pp.

[10] I. H. Witten and E. Frank, Data Mining: Practical Machine Learning Tools and Techniques with Java Implementation, 2nd ed. San Mateo.

[11] Huang, H., Wu, X.. Association analysis with one scans of web data bases. Paper submitted at the IEEE On Data Mining, Japan.

[12] R. Jin "An Efficient Implementation of Apriori Association web mining," Proc. Workshop on High Performance Data webMining, Apr. 2011.

[13] J. H and M. Kaber, "association mining:" 2014.

[14] Han J "Mining frequent patterns without candidate rules mining technique," in the national seminar of the international web of data, ACM Press, pp. 4-11-2004

[15] E-H. Han, G. Caryopsis "Scalable Data web mining for Association web Rules," IEEE Trans. Eng., vol. 12, no. 3, July 2012.

[16] Brin S., R. Mot, J.D. Ullman, "web item set counting and implication rules

[17] Association mining in data base", in Proceedings of the ACM SIGMOD International Conference on Management of Data, pp.289-294, 1999.

[18] Masseglia F., "Using Data Mining Techniques on Web Access Logs to Dynamically Improve Hypertext Structure language", In ACCM Web Letters, Vol. 10 No. 9, pp.13-19, 2011.