International Journal of Scientific Research and Engineering Development-– Volume 2 Issue 6, Nov-Dec 2019

Available at www.ijsred.com

RESEARCH ARTICLE OPEN ACCESS

Privacy Preserving Data Mining Technique to Recover Association Rules Using Homomorphic Encryption Technique

Varsha Patel*, NamrataTapaswi**

*(Computer Science and Engineering, IPS Academy, Institute of Engineeringand Science Indore
Email: varsha1993patel@gmail.com)

** (Computer Science and Engineering, IPS Academy, Institute of Engineeringand Science Indore
Email: hod.compsc@ipsacademy.org)

----------**********----------

Abstract:

The data mining is a technique which is used for analysing the data and recovering the patterns for applications. A number of applications are available where the data mining techniques are being used such as for classification, decision making, pattern learning, etc. In this presented work the privacy preserving data mining techniques are explored and a new technique for privacy preserving rule mining is introduced. That technique usage the contributed data from different data suppliers and mine the association rules securely. In order to mine the association rules the apriori algorithm is used. Additionally to secure the data at client end the Homomorphic encryption algorithm is used. After mining the required association rules from encrypted data it is delivered to the associated parties. The received encrypted rules are decrypted at the end of client and decryption algorithm is implemented here to recover the decision rules based on contributed part of data. The implementation of the proposed system is given in client server approach. That technique is developed with the help of java technology. Finally using the different experiments the performance of system is measured in terms of time and space complexity. The results demonstrate the proposed technique is acceptable and secure due to Homomorphic encryption. Additionally for extending the given concept the future plan is also proposed in this work.

Keywords — ***association rule mining, privacy preserving rule mining, apriori algorithm, Homomorphic encryption, client end privacy management.***

----------**********----------

I. INTRODUCTION

Data mining is an approach for mining the centralized database for extracting the valuable patterns from data. This process needs to implement the different computational algorithms for finding the required patterns. But sometimes when we work with the real world datasets the privacy and sensitivity of data are needed to be maintained [1]. The branch of data mining which preserves privacy during the mining of data is known as privacy-preserving data mining techniques [2]. In this work, the main area of study is privacy-preserving data mining. In this context, the multiparty data association and association rule mining technique is tried to implement and explore. In literature where a number of parties want the conclusion from the aggregated data, the rules are mining for ease of mining and recovery of contributed attributes. However, for mining effective techniques are available for extracting the conclusion from data such as statistical approaches, opaque data models and others.

But the rule mining technique much suitable in such kinds of applications. Thus to mine, the association rules the apriori algorithm is selected for providing the security and privacy the Homomorphic encryption technique is used. However, the Homomorphic encryption techniques are cost-effective but the security is very sound. These

techniques are very valuable where the multiple parties are agreed for combining their data and mine common decision for planning and business purpose. But not a single party wants to disclose their data for securing the data owner's privacy. Thus such kinds of data mining techniques are suitable for the medical industry, education, hospitality, etc. now in these, a number of other industries are following the concept of PPDM.

II. PROPOSED WORK

The proposed work for privacy-preserving data mining using association rule techniques is detailed in this chapter. The proposed model and the relevant algorithm is described in this chapter.

A. System Overview

The privacy-preserving data mining technique is a sub-domain of data mining where the data analysis securely is the primary aim of the system. Here the data mining techniques are mostly applied in various source-based collected data. In this technique, the multiple parties are agreed to combine their data at the trusted party and mine them for preparing the common decisions based on data. That technique is frequently used in various business intelligence applications, medical domain data management, and many more. The different kinds of data mining algorithms are applicable but the rule mining techniques are suitable for understanding and providing the decision rules. Using these rules any organization can understand and recover the decisions by traversing the rules using the available attributes.

The proposed work is motivated for designing a secure and efficient technique for mining association rules from the multi-party data supplier. Each party contributes a part of data which are communicated at the server and the server first aggregates that data for further processing. In this context for providing security at the party end, the Homomorphic encryption technique is applied. Additionally over the encrypted data the association rule mining namely apriori algorithm is applied. In this technique, the entire data is processed at the server end and the client is responsible for data security. After data processing or association rule generation the rules are distributed to all the associated parties. The contributed part of data which are used for generating rule is recoverable at the same party end who contributed that part of data. Therefore the proposed privacy-preserving data mining technique is providing the service for controlling security and privacy at the data owner's end.

B. Methodology

The proposed system is described in this section; figure 2.1, contains the phases of data processing using the proposed privacy preserving data mining technique.

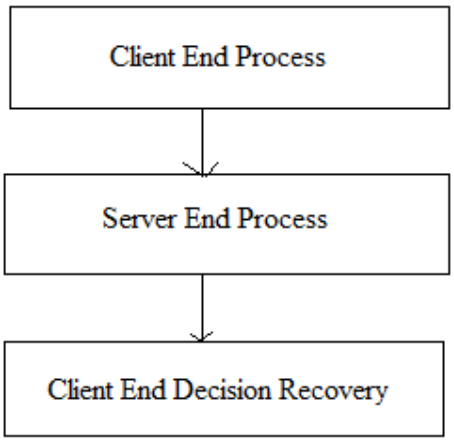

Fig.2.1 Proposed System

According to the given diagram 2.1, the proposed system is divided into three modules. The client selects their data at their own end and then connects to the server when the connection is established the user key is obtained by the client. The client is usages that key and produces the cipher text for the data. The server accepts the connection, collect the data, the process using the apriori algorithm generate rules. And finally, at the client end, the rules are extracted and recovered using the decryption process. The details of the proposed modules for privacy-preserving data modelling are described as:

No more than 3 levels of headings should be used.All headings must be in 10pt font.Every word

in a heading must be capitalized except for short minor words as listed in Section III-B.

1) ***Client End P:***The client end process is simulated to the figure 2.2, The model contains the required steps of data processing and their relationship among the server and client.

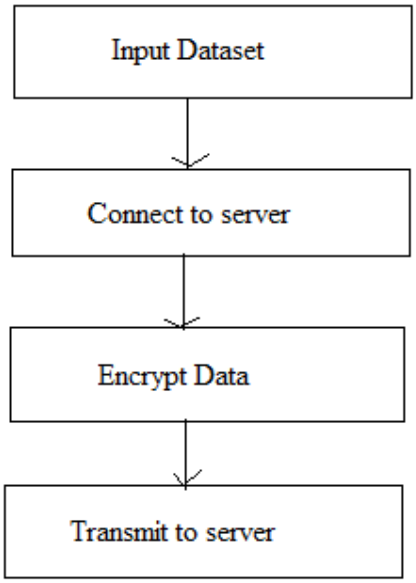

Fig. 2.2 Client End Process

Input dataset: all the concerned parties who want to combine their data with other parties confidentially and mine the data for concluding the decision rules. In this context, the user selects the part of their own data for sharing with the trusted server. The selected data is first read by the system and then the next processes are initiated.

Connect to server: the entire process is a client-server data management system therefore the processed data from the client end is needed to be forward at the server end. Therefore first need to establish connectivity from the server thus client creates a request to the server for accepting the connection and data. As the connectivity established the next process of encryption is taken place.

Encrypt data: the encryption is a technique which is used for securing the confidential data. That is a low cost and low maintenance technique for offering security and privacy over data. That transforms data from readable format to unreadable format. In this work, a Homomorphic algorithm namely Paillier algorithm is applied for preserving data confidentiality and privacy.

Transmit to server: the encrypted data attributes are reorganized into a file and transmitted to the server, where the server combines the obtained data from different parties are preparing a combined data set for processing.

2) ***Server End Process:***The server side process of system is demonstrated in figure 3.1, which provide the flow of steps which are involved for processing the data.

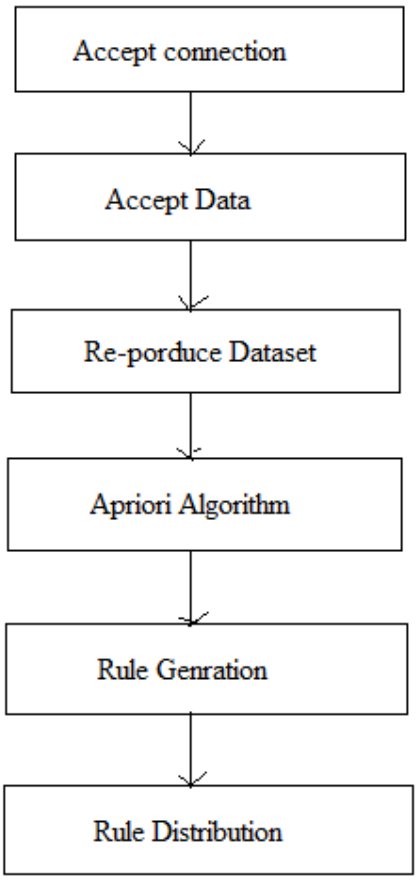

Fig. 3 Server End Process

Accept connection: the system is prepared to work in a multiple party environment therefore more than one client can connect with the server, when a data contributor is connected the server help to establish a connection. During this, the server system generates a unique ID for the connected party and sends the acknowledgment for the connection.

Accept data: after connection establishment, the client is able to communicate the data. The data is secured previously when they make a request for connection. After encryption of data, the server accepts the data from all the parties.

Re-produce dataset: the accepted data from all the data sources are combined together according to the available class labels and other parties' data. This process of combining all party data is termed here as the reproduction of the dataset.

Apriori algorithm: in literature different kinds of techniques available for mining the association rules. Among them, the apriori algorithm is much popular for association rule mining. The proposed technique here applies the apriori algorithm for

mining association rules. The apriori algorithm is explained in the previous chapter. The classical format of the apriori algorithm is used in this experimentation.

Rule generation: the computed frequent patterns are used in this phase for generating association rules. These rules are used for classification and prediction purposes.

Rule distribution: the generated rules are available at the client end and can be used by different parties. Therefore the server opens their connections for all the concerned parties to extract the rules from their own end.

3) ***Client End Decision Recovery:***The extracted rules from the server are distributed equally to all the connected data parties. These rules are used with the decryption process for recovering the contributed part of data and decisions. The encrypted part of the information is recovered due to the cryptographic algorithm by the target party. But the parties can only view the part of the decision variable which are contributed by own.

C. Proposed Algorithm

The algorithm is basic steps that are used for processing the data and generating the security and privacy-preserving rules. Thus two different activities using the algorithms are described in tables 2.1 and 2.2 First here for the server system and second is used for client end.

Table 2.1 Proposed Algorithm

Input: accepted dataset D Output: generated association rules R_n
Process: 1. $forall\ (Parties)$ a. $Connection.open()$ b. $Accept.Connection(x)$ c. $Temp_n = ReadSocket(x)$ d. $D.Add(Temp_n)$ 2. $endfor$ 3. $R_n = Apriori.GenratRules(D)$ 4. Return R_n

Table 2.2 Proposed Algorithm for Client

Input: data set D Output: recovered rule
Process: 1. $SendConnectionToServer()$ 2. $ReceiveConnnection()$ 3. $D_n = ReadData(D)$ 4. $A_n = HomoMorphic.Encrypt(D_n)$ 5. $SendToServer(A_n)$ 6. $wait(\ )$ 7. $if(connection == enable)$ a. $R_n = readRules$ b. $R_n = DecryptRules(R_n)$ 8. $endif$ 9. Return R_n

III. RESULTS ANALYSIS

This chapter provides the performance evaluation of the proposed privacy-preserving association rule mining technique. Therefore, two key parameters are focused on evaluation namely time requirements and memory usages. Both the performance factors are reported in this chapter.

A. Time Requirements

The time requirements of an algorithm are also known as the time complexity for algorithms. In this work for processing the input dataset and obtaining the association rules time requirements are measured. That is computed using the following Eq.

$$TimeRequirements = Algorithmendtime - Starttime$$

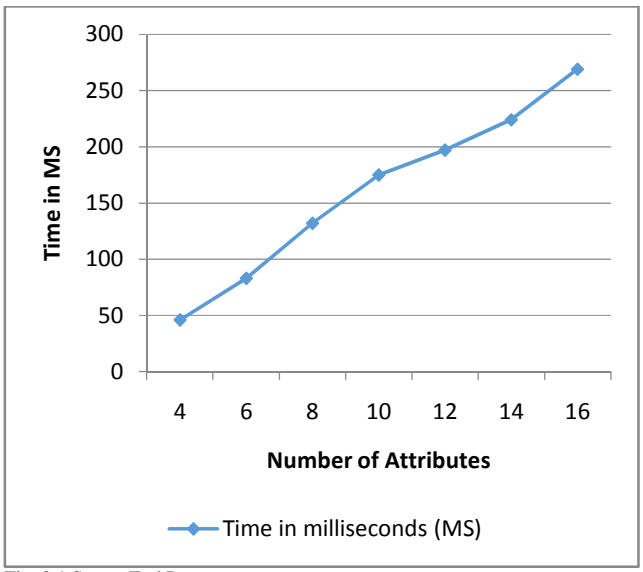

Fig. 3.1 Server End Process

The time requirement of the proposed privacy-preserving association rule mining technique using secure Paillier's algorithm is reported in table 3.1, additionally, their line graph representation is given in figure 3.1. In order to provide a line graph, the X-axis contains the number of attributes used for rule mining and the corresponding consumed time is notified on Y-axis. The time requirements of the algorithm are given here in terms of milliseconds. According to observed results as the number of time requirements, is increased when the number of attributes for processing and the number of transactions for processing is increased

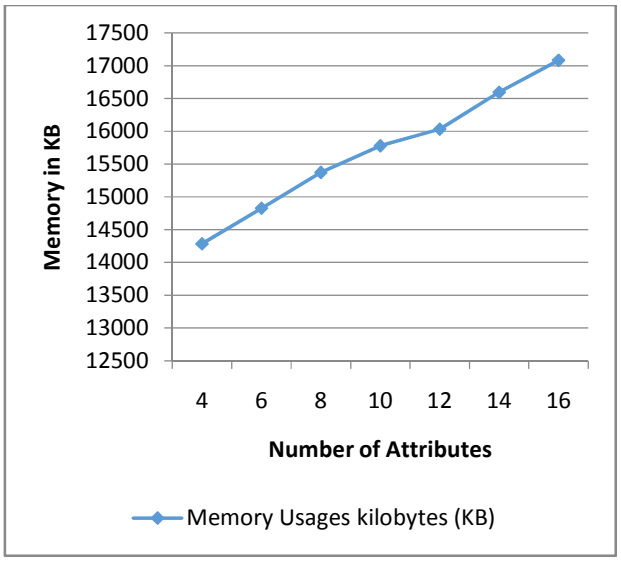

Fig. 3.2Memory Usages

Table 3.1 Time Requirement

Number of Attributes	Time in milliseconds (MS)
4	46
6	83
8	132
10	175
12	197
14	224
16	269

Table 3.2 Memory Requirement

Number of attributes	Memory Usages kilobytes (KB)	
4	14284	
6	14826	
8	15372	
10	15778	
12	16031	
14	16594	
16	17082	

The obtained results for the proposed privacy-preserving association rule mining technique is provided using table 3.2 additionally that is visually represented in a line graph as given in figure 3.2. The X-axis of the line graph prepared on the basis of the number of attributes involved in association rule mining and the Y-axis shows the corresponding measured main memory usages. The measurement of main memory utilization is given here in terms of kilobytes (KB). According to the developed results, the association rule mining space requirements are increases with the number of attributes and number of transactions.

IV. CONCLUSION

The implemented system for the privacy-preserving association rule mining is accomplished successfully. That offers security at the client end by using Homomorphic encryption; in addition to that server-side there is no probability to be data leak due to association rule mining is processed over the encrypted data. Therefore the proposed technique is fully secure and trusted.

A. Future Work

The aim of the proposed work is to understand the privacy-preserving data mining technique for discovering the association rules. In this context, the model is developed and evaluated successfully. In the near future, the following work is proposed for designing a new and enhanced system.

1. The work is aimed for multiparty based decision rule mining which increases the amount of data dimensions significantly, therefore need to handle the data dimensions,
2. The association rules are accurate for prediction and classification but it is computationally expensive therefore need to explore more rule mining techniques for improving the resource complexity.

REFERENCES

[1] *D. A. Adeniyi, Z. Wei, Y. Yongquan, "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method", Applied Computing and Informatics (2016) 12, 90–108*

[2] R. Mendes, J. P. Vilela, "Privacy-Preserving Data Mining: Methods, Metrics, and Applications", Vol. 5, 2017, 2169-3536, 2017 IEEE

[3] M. V. Ahluwalia, A. Gangopadhyay, Z. Chen, Y. Yesha, *"Target-Based, Privacy Preserving, and Incremental Association Rule Mining"*, IEEE Transaction on Services Computing, DOI 10.1109/TSC.2015.2484318, IEEE

[4] "Chapter 3: Data Mining: an Overview"

[5] M. J. Zaki, W. MeiraJr, *"Data Mining and Analysis Fundamental Concepts and Algorithms"*, Cambridge University Press Hardback, 2014 [Book]

[6] M. Goebel, L. Gruenwald, "A Survey of Data Mining and Knowledge Discovery Software Tools", ACM, 1999

[7] N. adhabPadhy, Dr. P. Mishra, *"The Survey of Data Mining Applications and Feature Scope"*, International Journal of Computer Science, Engineering and Information Technology (IJCSEIT), PP. 43-58 Vol.2, No.3, June 2012.

[8] N. Sundaravaradan, M. Marwah, A. Shah, N. Ramakrishnan, *"Data mining approaches for life cycle assessment"*, In Sustainable Systems and Technology (ISSST), 2011 IEEE International Symposium on, pp. 1-6. IEEE, 2011.

[9] F. Gorunescu, "Data Mining: Concepts, Models, and Techniques", Springer, 2011.

[10] Zhao, Yijun. "Data mining techniques." (2015).

[11] "Data Mining Tutorial: Process, Techniques, Tools & Examples"

[12] M. Tiwari, R. Singh, S. K. Singh, *"Association–Rule Mining Techniques: A general survey and empirical comparative evaluation"*, International Journal of Advanced Research in Computer and Communication Engineering, Vol. 1, Issue 10, Dec. 2012

[13] A. Mittal, A. Nagar, K. Gupta, R. Nahar, *"Comparative Study of Various Frequent Pattern Mining Algorithms"*, International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 4, April 2015

[14] *"Laboratory Module 8: Mining Frequent Itemsets – Apriori Algorithm"*, available online at: http://software.ucv.ro/~cmihaescu/ro/teaching/AIR/docs/Lab8-Apriori.pdf

[15] F. Armknecht, *"A Guide to Fully Homomorphic Encryption"*, IACR Cryptology ePrint Archive 2015 (2015): 1192.

[16] J. Sen, *"Homomorphic encryption: theory & applications"*,arXiv preprint arXiv: 1305.5886 (2013).

