Volume 2 || Issue 3 || 06-08
https://doi.org/10.51916/aksh.2021.v02i03.002

A Cryptographic Technique for Mining Association Rules in Distributed Databases with Privacy Preserving

*1 Byreddi Saritha, 2A. Rama Mohan Reddy
1,2 Computer Science & Engineering, Sri Venkateswara University, Tirupati, India
Email: byreddi.saritha@gmail.com

Received on: 05th August 2021; Published on: 30th September 2021

Abstract

The trend of Data mining has started more than a five decades. The datamining is to extract the interesting patterns from the huge databases and data warehouses. Now a day the data is coming from the distributed sources from various technologies like IoT (Internet of Things), Sensor Networks and soon. To extract the knowledge from that kind of data sources is a huge gap and with privacy preserving has a lot of research scope based on the literature. This work is focused on extracting the association rules in the distributed databases with the privacy preserving. Privacy preserving is a technique of securing the sensitive data of the users. Sensitive data may a sensitive raw data and sensitive patterns. Here proposed a technique called PACDD, based on the apriori and two fish cryptography algorithm to extract the association rules with privacy preserving in distributed databases. The proposed system shows efficiency in terms of the communication cost and the run time of the system when compared with the existing systems.

Keywords

Data Mining, Distributed, Cloud Service, Association Rules, Frequent Patterns, Privacy Preserving and Cryptography.

Introduction

Data mining is evergreen area and it has applications in almost all the fields like education, business, industries and retail sectors. It is used to extract knowledge from the databases. The knowledge may be a classification and prediction of data, mining association rules, and outliner analysis. Now a day, due to advanced technologies like Internet of Things (IoT), the data is vast and distributed [2]. Now a day, the data is distributed and unstructured, to extract knowledge from this kind of databases, researchers has done some research and developed some algorithms. Extracting the association rules in distributed databases is one of the big challenges. Simultaneously mining association rules with privacy preserving is a research gap, particularly in the databases like distributed and unstructured data. Privacy preserving data mining techniques are divided into four divisions [3], can be a heuristic, encryption, based on databases centralized or distributed and mixed strategies. In distributed databases, different users will use the different type of data from different sources, and the user requirements to extract the knowledge also will be different in terms of support, confidence, etc. [4]. Many mining association rules with privacy algorithms/methods/techniques [5-8] has been proposed, but major thing is the efficiency. To solve the above issues his paper is focused on mining association rules with privacy preserving by using the cryptographic algorithm with aprioi has proposed and implemented and was discussed in detail in next sections.

The remainder of sections are, in section Related Work, Proposed System, Implementation, Results and Discussion and finally the Conclusion and Future Work was discussed.

Related Work

In [1], the authors developed an algorithm for datamining by using the min hashing/k-min hashing and sensitive hashing, hat which is used to increase the scalability and to reduce the network traffic. In [3], has developed the cryptographic based algorithms to mine the association rules by achieving the privacy using RSA public key cryptographic algorithm and homomorphic encryption keys algorithms. In [4], the problem to extract association rules from web data, that present in multiple sources, has been solved by using the technique of sharing and testing the knowledge patterns. In [5], the authors proposed a method based on load balancing with multi core frequent pattern mining algorithm. In [9-10], proposed a framework called MAD-ARM, based on multi agents, used to extract the association rules in distributed platform.

Proposed Work

The proposed system is to develop a technique that which will extract the association rules in the distributed databases with the privacy preserving by using the cryptographic technique. The proposed algorithm was shown in the Algorithm 1. The inputs for the algorithm are the dataset, minimum support, and minimum confidence for extracting the association rules, for achieving the privacy i.e. to secure the sensitive raw data, while extracting the association rules from distributed databases, the two fish algorithm was used. The proposed system will work

E–ISSN: 2582 7456

based on the apriori algorithm and two fish cryptographic algorithm for extracting the frequent patterns from the distributed databases.

Algorithm 1: PARCDD

```
1.  Start
2.  Input: dataset, minimum support, minimum confidence
3.  If   data ==sensitive
          Twofish ()
    Else
          preparedata ()
4.  patterns()
5.  End
```

Here first all the association rules were extracted locally in all databases and second extracted globally by securing the sensitive data while extracting the association rules. Thus, the proposed system can achieve privacy preserving while extracting the association rules.

Implementation

The implementation was done in the distributed platform called Hadoop [1]. Here the Abalone dataset was used, which consists of 4177 instances and 8 attributes, that which was downloaded from the UCI machine learning repository for implementing the proposed technique. The proposed algorithm was implemented on the abalone dataset and compared with the existing algorithm, which is based on the cryptographic algorithm called blowfish, for achieving privacy preserving while extracting the association rules in distributed databases. The figure 1 shows the working model of the Hadoop. After implementing the obtained results were shown and discussed in the section 4.

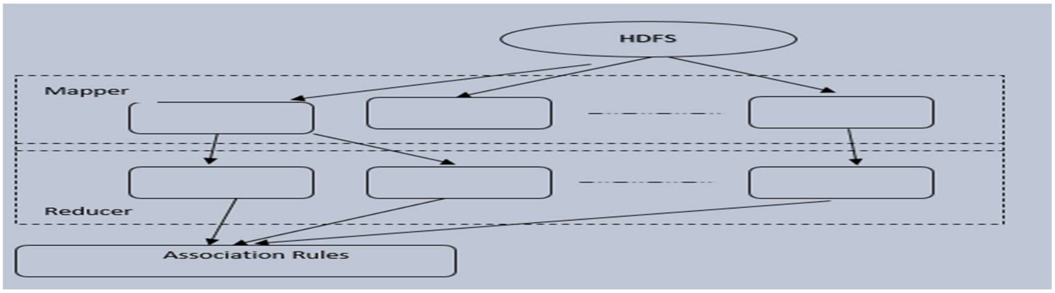

Figure 1: Hadoop Working Model

Results and Discussion

Here the results were generated after implementing in a distributed platform called Hadoop, the obtained results were compared with the technique called PARCDDC. The results show the effectiveness in terms of the communication cost and the runtime. In terms of the communication cost, the proposed system shows 4 percentage more than the existing system and in terms of the runtime the proposed system shows 7 percentage more than the existing system. The figure 2 and figure 3 shows the results in terms of the communication cost and runtime.

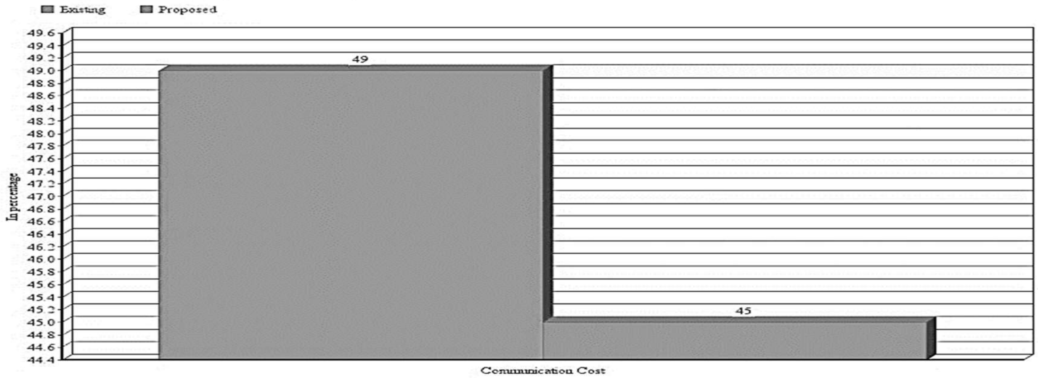

Figure 2: Communication Cost of Existing and Proposed Systems

https://doi.org/10.51916/aksh.2021.v02i03.002

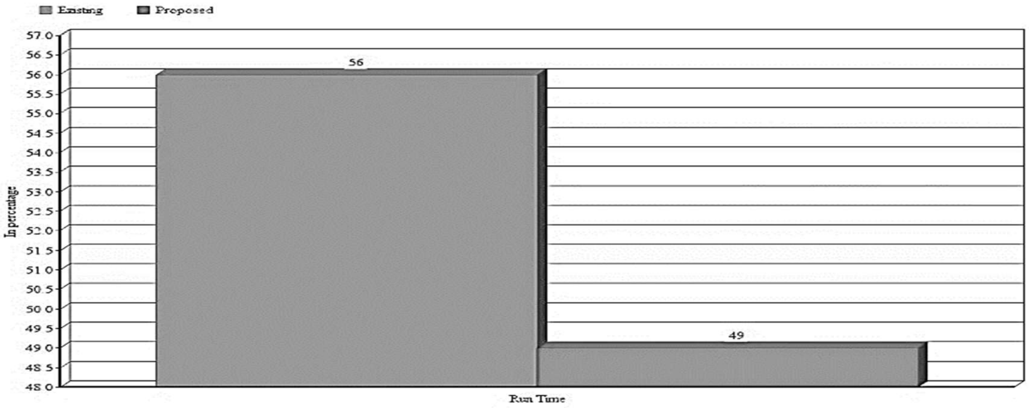

Figure 3: Run Time of Existing and Proposed Systems

Conclusion and Future Works

The cryptographic and Apriori algorithm was used in the proposed system. Here the sensitive raw data of the user was secured and association rules were extracted. In future the sensitive patterns may be secured with different cryptographic algorithm and different association rule algorithm with randomization.

References

[1] Ferro, Alfredo, et al. "Distributed randomized algorithms for low-support data mining." 2009 IEEE International Symposium on Parallel & Distributed Processing. IEEE, 2009.

[2] Baseer, K. K., et al. "Internet of Things: A Product Development Cycle for the Entrepreneurs." Helix 10.02 (2020): 155-160.

[3] Gui, Qiong, and Xiao-hui Cheng. "A privacy-preserving distributed method for mining association rules." 2009 International Conference on Artificial Intelligence and Computational Intelligence. Vol. 4. IEEE, 2009.

[4] Bai, Shilei, et al. "Hypothesis Testing Based Knowledge Discovery in Distributed Multiple Data Sources." 2010 International Conference on Internet Technology and Applications. IEEE, 2010.

[5] Yu, Kun-Ming, and Shu-Hao Wu. "An efficient load balancing multi-core frequent patterns mining algorithm." 2011IEEE 10th International Conference on Trust, Security and Privacy in Computing and Communications. IEEE, 2011.

[6] A. M. J. Md. Zubair Rahman and P. Balasubramanie, “An Efficient Algorithm for Mining Maximal Frequent Item Sets,” Journal of Computer Science 4 ISSN 1549-3636, 2008, pp.638-645.

[7] Bodon, F. “A fast Apriori implementation,” Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations, 2003.

[8] G. Liu, H. Lu, Y. Xu, and J. X. Yu, “Ascending Frequency Ordered Prefix-tree: Efficient Mining of Frequent Patterns,” In Proc. 8th Int. Conf. Database Systems for Advanced Applications, 2003, pp. 65-72.

[9] Raja, A. Saleem, and E. George Dharma Prakash Raj. "MAD-ARM: Mobile agent based distributed association rule mining." 2013 International Conference on Computer Communication and Informatics. IEEE, 2013.

[10] Walid Adly Atteya, Keshav Dahal and M. Alamgir Hossain, “Distributed Bit Table multi-agent Association Rules Mining Algorithm”, Springer-Verlag, KES 2011, Part I, LNAI 6881.

E–ISSN: 2582 7456