ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 4, Issue 12, December 2016

Survey on Outsourced Frequent Itemset Mining in Data Mining

Rashmi B. Kale[1], Kanchan M. Varpe[2]

M.E., Dept. of Computer, RMD Sinhgad School of Engineering, Pune, India[1]

Assistant Professor, Dept. of Computer, RMD Sinhgad School of Engineering, Pune, India[2]

ABSTRACT: Cloud computing is promoting the computing worldview in which information is outsourced to an outsider specialist third party server for data mining. Outsourcing raises a genuine security and correctness issue in what manner the customer of weak computational power can confirm that the server returned right mining result. It considers the server that is possibly untrusted and tries to escape from check by utilizing its earlier learning of the outsourced information.It propose proficient probabilistic and deterministic confirmation ways to deal with check whether the server has returned right and finish visit itemsets. Our probabilistic approach can get mistaken outcomes with high likelihood, while our deterministic approach measures the result accuracy with 100% confidently. It likewise plan productive confirmation strategies for both cases that information and mining setup are remodel. It exhibits the viability and productivity of our techniques utilizing a broad arrangement of exact outcomes on genuine data-sets. The proposed system extract frequent itemset according to user provided budget. The frequent itemsetsare generated according to budget value of itemset. It enhanced with the recommendation of top k most relevantitems with respect to the provided budget value to the user.

KEYWORDS: Cloud computing, data mining as a service, security, result integrity verification.

I. INTRODUCTION

Data mining is the strategy for investigating data from alternate points of view and summarizing it into valuable data that can be utilized to expand income, cuts costs, or both. Data mining programming is one of the diagnostic instruments for dissecting data. It permits clients to assess data from a wide range of measurements or edges, order it, and condense the connections is recognized. In fact, data mining is the technique for discovering connections or examples among many fields in extensive social databases. Data mining (the examination venture of the "Learning Discovery in Databases" process, or KDD), an interdisciplinary subfield of software engineering, the computational procedure of investigating examples in vast data sets including techniques at the crossing point of computerized reasoning, insights, and database frameworks. The general objective to choose promoting techniques for their item. They can utilize data to look into among contenders. Data mining translates its data into continuous examination that can be utilized to build deals, advance new item, or erase item that is not esteem added to the organization.

Pattern mining is a data mining technique that includes finding existing examples in data. In these setting designs regularly implies association rules. The first inspiration for looking association rules originated from the yearning to break down general store exchange data, that is, to inspect client conduct regarding the acquired items. For instance, an affiliation rules "Soft Drink⇒ potato chips (80%)" states that four out of five clients that purchased soft drink likewise purchased potato chips. With regards to pattern mining as an instrument to recognize buying pattern group set of the data mining technique is to concentrate data from a data set and change it into a reasonable structure for further utilize.

Data mining utilizes data from past information to dissect the result of a specific issue or circumstance that may emerge. Data mining attempts to assess information stored in data warehouses that are utilized to store that information that is being analyzed. That specific information may originate from all parts of business, from the generation to the administration. Supervisors likewise utilize information mining to settle on showcasing methodologies for their item. They can utilize information to thoroughly analyze among contenders. Data mining translates its information into real

DOI: 10.15680/IJIRCCE.2016. 0412093

time investigation that can be utilized to expand deals, promote new item, or erase items that are not profitable to the organization.

Frequent sets assume a fundamental part in numerous Data Mining tasks that attempt to discover interesting patterns from databases, for example, association rules, relationships, groupings, scenes, classifiers and clusters. The ID of sets of things, items, side effects and attributes, which regularly happen together in the given database, can be viewed as a basic task amongst the most fundamental assignments in Data Mining. The inspiration for looking frequent sets originated from the need to analyze supermarket grocery store exchange information, that is, to examine client conduct as far as the purchased items. Frequent sets of items portray how frequently items are obtained together.

II. Related work

The expanding capacity to create endless amounts of data presents specialized difficulties for productive data mining. Outsourcing data mining calculations to an outsider specialist server offers a financially effective alternative, particularly for datacustomers of constrained assets. It presents the data-mining-as-an administration (DMaS) worldview. It concentrates on regular itemset mining as the outsourced data mining undertaking. Frequent itemset mining has been demonstrated critical in numerous applications for example; advertise data analysis, organizing data study, and human quality affiliation think about. Past research has demonstrated that successive itemset mining can be computationally serious, because of the enormous hunt space that is exponential to data measure and also the conceivable hazardous number of found visit itemsets. In this manner, for those customers of constrained computational assets, outsourcing incessant itemset mining to computationally effective specialist third party is a characteristic arrangement. Some past survey for related work has been done here:

In [1], author distinguishes different processing paradigms promising to convey the vision of computing utilities; characterizes Cloud computing and gives the design to making market-arranged Clouds by utilizing technologies, for example, VMs; gives thoughts on market-based asset administration techniques that include both client driven management and computational hazard management to maintain SLA oriented resource allocation.

In [2], author prescribed that Outsourcing association rule mining to an outside specialist organization conveys a few critical focal points to the information owner. These incorporate (i) help from the high mining rate, (ii) reductionof demands in assets, and (iii) real centralized mining for multiple distributed owners.Furthermore, security is an issue; the specialist organization ought to be kept from getting to the genuine information since (i) the information might be connected with private data, (ii) the recurrence investigation is intended to be utilized exclusively by the owner. It proposes substitution figure systems in the encryption of value-based information for outsourcing association rule mining.

In [3], author expressed that the data mining plays a vital role in decision making. Since numerous associations do not possesses the in-house expertise of data mining, it is valuable to outsource information mining assignments to outer specialist organizations. However, most organization hesitates to do as such because of the worry of loss of business insight and client security. It introduced a Bloom channel based solution for empower associations to outsource their tasks of mining association rules, at the same time, ensure their business insight and client protection. Given approach can accomplish high rightness in data mining by exchanging off the capacity prerequisite but privacy is not provided

In [4], author showed a Privacy preserving data mining getting valid data mining results without learning the underlying data values has been receiving attention in the research community. It is not clear what privacy preserving means. It provides a system and measurements to examining the importance of security saving information mining, as an establishment for further research in this field but no system or an application is created

In [5], author build up a hybrid strategy towards security protecting frequent item set mining in outsourced exchange databases. For giving protection by means of k-support anonymity, the strategy includes a few things furthermore minimizes a few things. The reason for such frameworks is that the median minimizes the whole of mean absolute difference.

In [6], author fabricated a protection saving outsourced association rule mining solution for vertically partitioned databases. Our solutions protect data owner's raw data from other data owners and the cloud. Given system also assure the privacy of the mining results from the cloud. Compared with most existing solutions, our solutions leak less information about the data owner's raw data but outsourced comparison scheme in other setting is not provided

In [7], author gave efficient result integrity verification method which can provide deterministic guarantee for outsourced frequent item set mining. The key idea of the approach is to construct cryptographic proofs of all (in) frequent item sets. Authors have also discussed how to optimize the number of proofs to improve the performance. They also extended our study to the verification of maximal frequent item set mining.

In [8], author developed hybrid methodfor privacy preserving frequent item set mining in outsourced transaction databases. For providing privacy through k-support anonymity, the technique adds some items as well as removes few items. The reason for such an approach is, to the point that the median reduce the sum of mean absolute difference. In developed strategy, they additionally lessen the impact of outliers generating massive spurious frequent itemsets.

In [9], author showed methodfor outsourcing association rule mining to protect BI and customer privacy. The developed method is different from previous developed methods in which it can protect BI and customer privacy at the same time of outsourcing mining tasks, while maintaining the accuracy of mining outcomes. Next they did theoretical analysis on the false positive and false negative rates in data mining. At last investigated the trade-offs between mining precision and storage requirement butprivacy preserving techniques for mining specific types of frequent itemsets is not given.

In [10], author It analysed the issues of outsourcing the association rule mining task in a corporate privacy-preserving framework. It developed an attack model depending on background knowledge and devise a system for privacy preserving outsourced mining andminimize the number of spurious patterns

Wehave studied several studies on the data mining as well as on the frequent data set mining. We also identify the advantage and limitation of each approach along with techniques used for developing the system and given idea of the system which can be eliminates the cons of the existing systems

III. PROPOSED ALGORITHM

A. *DESIGN CONSIDERATIONS:*

- Initially data user reads the dataset.
- Approaches have been chosen accordingly.
- Frequent item sets get generated merkle hash tree get formed.
- Data security has been preserve with algorithm.
- Correctness and completeness get checked.
- Verification has been done with the user budget.

B. *DESCRIPTION OF THE PROPOSED ALGORITHM:*

In present day cryptography, AES is broadly embraced and upheld in both equipment and programming. Till date, no handy cryptanalytic assaults against AES have been found. AES has worked in adaptability of key length, which permits a level of 'future-sealing' against advance in the capacity to perform comprehensive key searches.AES can be

DOI: 10.15680/IJIRCCE.2016. 0412093

easily implemented using cheapprocessors and a minimum amount of memory. Very efficient Implementation was a keyfactor in its selection as the AES cipher.

IV. PSEUDO CODE

```
Step 1:  Key Expansion:  - Using Rijndael as key schedule Round keys are derived from the cipher key.

Step2:  If DistanceToTree(u) >DistanceToTree(DCM) and First-Sending(u) then

Step3:  Initial Round:-AddRoundKey where Each byte of the state is combined with the
round key using bitwise xor.

Step4:  Rounds

    •  SubBytes: non-linear substitution
    •  ShiftRows: transposition
    •  MixColumns: mixing operation of each columnAddRoundKey

Step 5:  Final Round: It contain SubBytes, ShiftRows and AddRoundKey

Step 6:  End.
```

V. PROPOSED SYSTEM

Figure 1 shows the system architecture of the proposed system. In proposed system the data owner reads the dataset. The data set is used as the input for the probabilistic approach which divides the given data set in frequent and infrequent data set.The key thought of our techniques is to build an arrangement of infrequent itemsets from genuine things, and utilize these infrequent itemsets as confirmation to check the uprightness of the server's mining result.

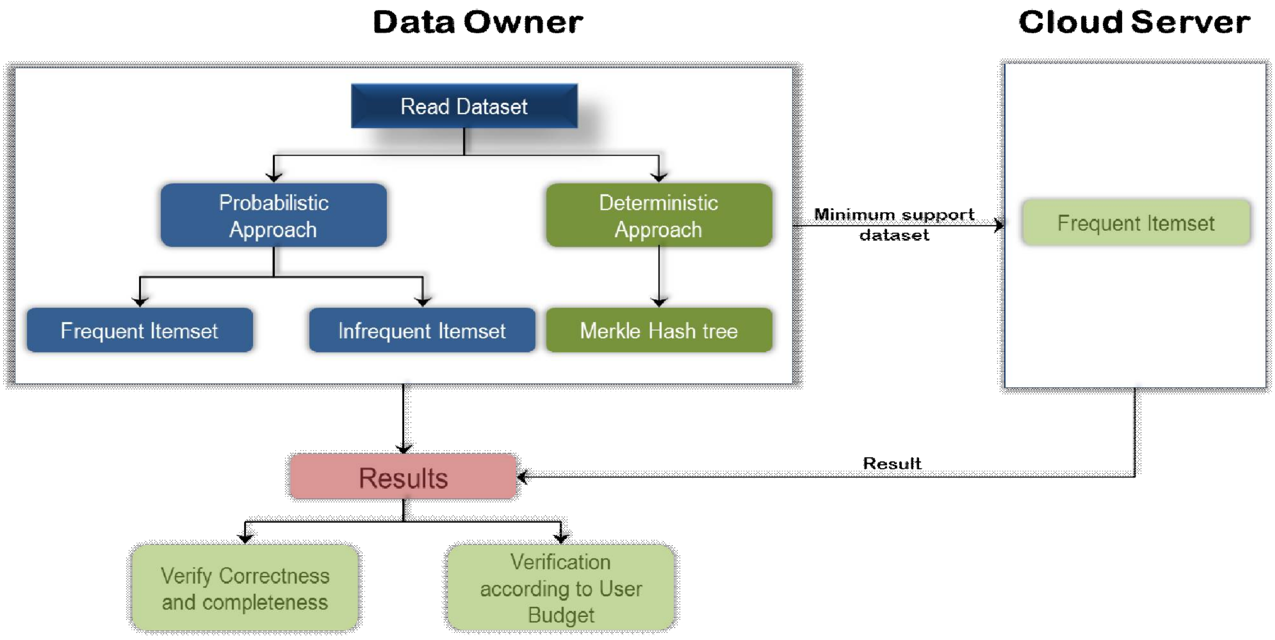

Fig 1. Propose System

It expels genuine things from the first dataset to build manufactured proof occasional itemsets. Embed duplicates of things that exist in the dataset to develop counterfeit confirmation visit. The deterministic approach is used for the authentication check which is performed by using Merkle Hash Tree. It measure the execution of evidence development at the server side and confirmation at the customer side also, investigated different variables that effect the confirmation execution of our deterministic approach, including different mistake proportion, visit itemsets of various lengths, and diverse database sizes. Same time the data owner sends the requirement to the cloud server where the frequent item sets are computed and send to the data owner where the results are compared with the data owner's needs. The results are verified according to the user provided budget and output is displayed to the user.

VI. CONCLUSION

From this survey, we have studied several studies on the data mining as well as on the frequent data set mining. We also identify the advantage and limitation of each approach along with techniques used for developing the system and given idea of the system which can be eliminates the cons of the existing systems.The data owner reads the dataset and is used as the input for the probabilistic approach which divides the given data set in frequent and infrequent data set.The deterministic approach is used for the authentication check which is performed by using Merkle Hash Tree.Same time the data owner sends the requirement to the cloud server where the frequent item sets are computed and send to the data owner where the results are compared with the data owner's needs. The results are verified according to the user provided budget and output is displayed to the user.

VII. ACKNOWLEDGEMENT

The authors wish to thank all the referees involved in the above mentioned work of this survey. Finally, an honorable mention goes to the entire staff of Dept. of Computer, RMD Sinhgad School of Engineering,WarjePune and for their understandings and supports on this survey.

REFERENCES

1. R. Buyya, C. S. Yeo, and S. Venugopal, "Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities," in Proc. IEEE Conf. High Performance Comput. Commun., Sep. 2008, pp. 5–13.
2. W. K. Wong, D. W. Cheung, E. Hung, B. Kao, and N. Mamoulis, "Security in outsourcing of association rule mining," in Proc. Int. Conf. Very Large Data Bases, 2007, pp. 111–122.
3. L. Qiu, Y. Li, and X. Wu, "Protecting business intelligence and customer privacy while outsourcing data mining tasks," Knowledge Inform.Syst., vol. 17, no. 1, pp. 99–120, 2008.
4. C. Clifton, M. Kantarcioglu, and J. Vaidya, "Defining privacy for data mining," in Proc. Nat. Sci. Found. Workshop Next Generation Data Mining, 2002, pp. 126–133.
5. I. Chandrasekharan, P. K. Baruah and R. Mukkamala, "Privacy-preserving frequent itemset mining in outsourced transaction databases," Advances in Computing, Communications and Informatics (ICACCI), 2015 International Conference on, Kochi, 2015, pp. 787-793.
6. L. Li, R. Lu, K. K. R. Choo, A. Datta and J. Shao, "Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases," in IEEE Transactions on Information Forensics and Security, vol. 11, no. 8, pp. 1847-1861, Aug. 2016.
7. B. Dong, R. Liu and W. H. Wang, "Integrity Verification of Outsourced Frequent Itemset Mining with Deterministic Guarantee," 2013 IEEE 13th International Conference on Data Mining, Dallas, TX, 2013, pp. 1025-1030
8. I. Chandrasekharan, P. K. Baruah and R. Mukkamala, "Privacy-preserving frequent itemset mining in outsourced transaction databases," *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Kochi, 2015, pp. 787-793.
9. Qiu, Ling, Yingjiu Li, and Xintao Wu. "Protecting business intelligence and customer privacy while outsourcing data mining tasks." *Knowledge and information systems* 17.1 (2008): 99-120.
10. Giannotti, Fosca, et al. "Privacy-preserving mining of association rules from outsourced transaction databases." *IEEE Systems Journal* 7.3 (2013): 385-395.