

OUTSOURCING OF FREQUENT ITEMSET MINING WITH VERIFICATION AND USER PROVIDED BUDGET

Miss. Rashmi Kale¹, Prof. Kanchan Varpe²

1 Rashmi Kale, R. M. Dhariwal Sinhgad School of Engineering, Pune

2 Prof. Kanchan Varpe, R.M. Dhariwal Sinhgad School of Engineering, Pune

Abstract- *Data mining is a need of future, each association; data administrator does not have adequate assets, foundation to satisfy their client's needs. It is an organization perspective enables the data client's needs capacity, computational resources for outsource its mining task to an outcast association server. Outsourcing raises a real security and correctness issue in what way can the client of frail computational power state that the server returned right mining conclusion. It concentrates on the challenges of frequent item set mining, and proposes a proficient and functional authentication ways to deal and check whether the server has returned precise and complete frequent item sets. The proposed framework mines frequent item set according to user offered budget. The frequent item sets generate according to financial statement value for item sets. Also this system is improved with the concept of recommendation system. This returns the top k most appropriate items with respect to the provided financial value to the user.*

Key Words: Data mining as a service, cloud computing, security, result integrity verification.

1. INTRODUCTION

Cloud computing is one of the booming techniques which proving marginal computing services, gives the chance which is provided by data mining as outsourced service. Regardless of the possibility that the data mining-as-a service (DMaS) is precious to achieve refined information consider in a cost effective, end users falter to place full trust in Cloud preparing. It raises certifiable security issues. One of the essential security problems is the uprightness outcomes of the mining. There are various conceivable explanations behind the specialist organization to return off base answers. Like, the specialist co-op may want to increase its earning by computing with minimum assets while charging more. For instance, the master center might need to upgrade its pay by figuring with less resource while charging for extra. Thusly, it is needed to provide effective frameworks to make sure the result respectability of outsourced data mining algorithms.

The work concentrates on mining of frequent item set, vital issues related to information mining, as the primary outsourced information mining service. It mean to solve the specific issue of confirming whether the server has returned

right and complete frequent item sets. By accuracy, it implies that all item sets returned by the server are frequent. By fulfillment, it implies that no frequent item set is missing in the server's outcome.

The basic concept of presented technique is to develop a set of frequent item sets from real items as well as make utilization of these frequent item sets evidence of morality of the server's mining outcomes. It discards genuine things from the prior dataset for creating manufactured rare item sets also consolidates repeated things which present dataset for creating artificial successive items. In designed method the amounts of needed evidence frequent item sets are not depended on the size of the dataset and the amount of real frequent item sets, it is one of the decent properties of designed verification method. As developed method is suitable for verification of frequent mining on large datasets, it will become an advantage of our technique.

The concept of the system is to get verification of outsourced frequent item set mining outcomes. To do so we implement the probabilistic method to fetch the mining outcomes which does not provides the accuracy needed with high probability which is predefined. Basic concept is to derived a set of frequent item set from real items as well as make use of these item set as evidence for checking integrity of the server's mining outcomes.

We can say that the data mining is must in coming futures; it is not needed that the every data owner must have enough resources, system to fulfill the needs of users. Keeping this in mind storing of data mining is needed. Regardless of the fact that it is ideal to finish complex investigation on huge volumes of data in a cost effective manner, there are diverse certified security concerns of the data mining as-a service, for example, data mining performed at the server side and client needs to ask for the data mining. One of the issues is that when information is outsourced for mining reason, server has segment to the private beneficial data of the administrator and can access to sensitive data. In this manner security may break. Besides correctness is another issue rise.

The proposed framework incorporates:

- Design the probabilistic method to seize mining outcome that does not meet the predefined

Correctness and fullness with high possibility. The key thought is to build a set of (in) frequent item sets from genuine data set and utilize these (in) frequent item sets as proof to check the trustworthiness of the server's mining end result.

- Design the deterministic method to grasp any wrong and uneven infrequent item set mining response with 100% possibility.
- The key thought of deterministic method is to require the server to develop cryptographic verifications of the mining outcome.
- Both rightness and fulfillment of the mining outcomes are measured against the verifications with 100% assurance.
- For both probabilistic and deterministic methodologies, give productive techniques to manage updates on both the outsourced data information and the mining setup.
- Implement recommender framework to suggest the top k data items from frequent data set as per the client offered finances.

In this paper we study about the related work done, in section II, the proposed approach, modules, its description, mathematical modeling, algorithms in section III, experimental setup, expected results, accuracy comparison graphs and results discussions in section IV and at final we provide a conclusion and future scope in section V.

2. LITRATURE SURVEY

Some past survey for related work has been done here:

“Trust-but-Verify: Verifying Result Correctness of Outsourced Frequent Item set Mining in Data-Mining-As-a-Service Paradigm” [1], Cloud computing is popularizing the computing structure in which information is outsourced to an outsider specialist organization (server) for information mining. Outsourcing, in any case, raises a genuine security issue: by what means can the customer of frail computational power confirm that the server returned rectify mining result. It focuses on the specific task of frequent item set mining. System considers the server that is potentially entrusted and tries to escape from verification by using its prior knowledge of the outsourced data. Efficient probabilistic and deterministic authentication methods make sure whether the server has returned accurate and entire frequent item sets. This probabilistic method can catch erroneous results with high possibility, while deterministic method deals the result accuracy with 100 percent confidence and it also plan proficient verification methods for both cases that the data and the mining setup are updated.

“Privacy-preserving frequent item set mining in outsourced transaction databases” [2], Cloud processing has encouraged another enthusiasm for a worldview called Data mining-as-a service. This worldview is gone for associations that do not have the specialized ability or the computational assets empowering them to outsource their information mining assignments to an outsider specialist organization. One of the primary issues in such way is the classification of the profitable information at the server which the information proprietor considers as private. It studies the problem of privacy preserving frequent item set mining in outsourced transaction databases. Author proposes a novel hybrid method to meet k-support anonymity based on statistical observations on the datasets.

“Privacy-Preserving in Outsourced Transaction Databases from Association Rules Mining” [3], Data mining-as-a-service has been carefully chosen as extensive research problem by researchers. An organization (data owner) can outsource its mining needs like resources or knowledge to a third party service provider (server). However, both the association rules and the items of the subcontracted operation database are reserved property of data owner. The data vendor encrypts its data, sends data and mining requests to the server, and accepts the correct designs from the encoded patterns expected from the server to protect the confidentiality. The difficulty of outsourcing operation database within a commercial privacy framework is calculated. We propose an attack model based on earlier facts and plan a scheme for confidentiality preserving subcontracted data mining. Scheme guarantees that each transformed data is dissimilar with respect to the assailant's prior statistics. The investigational results on actual operation database prove that methods are accessible, effectual and guard confidentiality.

“Secure Mining of the Outsourced Transaction Databases” [4], Data mining is a diagnostic device which permits clients to break down information, classes it and synopsis the connections among the information. It finds the helpful data from huge measure of social databases. Security is an essential issue while performing mining job towards the end of outsider third party service. At the point when information administrator outsources its mining needs to outsider specialist or third party, there must be a security. Outsider specialist third party has entry to the information and can take in delicate data from it. The outcome is security may break. As information is an essential private property of the administrator it must be secure. It displays the calculations utilized for mining reason and in addition for securing the information while executing between Data administrator and third meeting author. It gives secure, proficient and speedier outcomes. It gives secure, compelling and snappier results.

“Integrity Verification of Outsourced Frequent Item set Mining with Deterministic Guarantee” [5], focuses on the issue of result trustworthiness check for outsourcing of

frequent data set mining. Creator plan productive cryptographic methodologies that check whether the returned frequent item set mining results are right and complete with deterministic certification. The key of our answer is that the specialist third party provider develops cryptographic evidences of the mining outcomes. Both rightness and fulfillment of the mining results are measured against the evidences.

“Publicly Verifiable Delegation of Large Polynomials and Matrix Computations, with Applications” [6], Outsourced calculations are becoming progressively significant due to the growth of Cloud Computing and the explosion of mobile devices. Since cloud suppliers may not be trustworthy a critical problem is the authentication of the reliability and accuracy of such computation, possibly in a public way, i.e., the result of a calculation can be proved by any third party, and needs no secret key similar to a digital signature on a message. It presents novel conventions for publicly provable secure outsourcing of Valuation of High Degree Polynomials and Matrix Exponentiation. Compared to earlier projected solutions, increase in skill and proposal security in a stronger model. It also deliberates many practical solicitations of the protocols.

“Market-oriented cloud computing: Vision, hype, and reality for delivering it services as computing utilities” [7], creator recognizes distinctive handling standards promising to pass on the vision of figuring utilities; describes Cloud processing and gives the plan to making market orchestrated Clouds by using advances, for example, VMs; gives considerations on showcase based resource organization procedures that incorporate both customer driven administration and computational hazard management to keep up SLA oriented resource allocation.

“Security in outsourcing of association rule mining” [8], 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) reduction of 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 used exclusively by the owner. It proposes substitution figure systems in the encryption of value-based information for outsourcing association rule mining.

“Protecting business intelligence and customer privacy while outsourcing data mining tasks” [9], author expressed that the data mining plays a vital role in decision making. Since many associations do not possess 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, make sure their business insight and client protection. Given approach can carry out high rightness in data mining by exchanging off the capacity prerequisite but privacy is not provided.

“Result Integrity Verification of Outsourced Frequent Item set Mining” [10], the information mining-as-a service (DMaS) standard empowers the information administrator (customer) that needs aptitude or computational assets to outsource its mining events to an outsider specialist organization (server). Outsourcing, in any case, raises a genuine security issue: in what way can the customer of weak computational power confirm that the server returned precise mining result. It concentrates on the issue of continuous item set mining, and proposes a productive and down to earth probabilistic confirmation ways to deal with check whether the server has returned right and finish visit item sets.

3. PROPOSED SYSTEM

Proposed system enhanced with user provided budget and recommendation of item set.

3.1 Problem Definition

For given outsourced dataset, develop a system to mine the frequent item sets and verify the correctness of outsourced frequent item set according to the user provided budget at user side. Also enhanced this system by providing recommendation of item to users for provided budget.

3.2 Proposed System Overview

The proposed system focus on the problem of verifying whether the server returned accurate and total frequent item sets. By accuracy, we imply that all data item sets returned by the server are recurrent. By entirety, we imply that no frequent item set is absent in the returned result and furthermore suggested that frequent item set. Here in propose system data owner read the dataset and apply probabilistic approach to catch mining result that does not meet the predefined accuracy and fullness requirement with high probability and deterministic approach to catch any incorrect/ incomplete frequent item set mining answer with 100% probability. The data owner send minimum support and dataset to cloud server and cloud server generate the frequent item set and also recommended top k list and send to the data owner. Data owner verify the correctness and completeness of a frequent items. The frequent item sets are generated according to budget value of item set and user get top k item set with respect to their provided budget value which is verified by the user.

Proposed system design model is as per following:

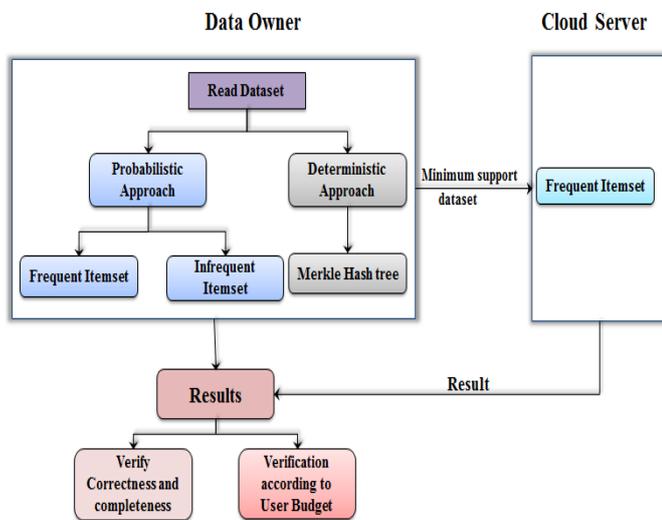


Fig – 1: Proposed System Architecture

1. Data Owner

Here Data owner take a dataset and read dataset which content probabilistic approach and deterministic approach.

2. Probabilistic Approach

This module is utilized to find mining result that does not meet the predefined rightness/fulfillment necessity with high likelihood. The key thought is to build an arrangement of (in) frequent item sets from genuine things, and utilize these (in) frequent item sets as confirmation to check the respectability of the server’s mining comes about.

3. Deterministic Approach

This module is used to catch any incorrect/incomplete frequent item set mining answer with 100% likelihood. The key thought of our deterministic arrangement is to require the server to build cryptographic evidences of the mining comes about. Both rightness and culmination of the mining results are measured against the proofs with 100% assurance.

4. Frequent Item-set and Recommendation

In this module the cloud server generated a frequent item set from the dataset and minimum support value and sends the generated frequent item set result and sends to the data owner in encrypted format. In contribution the system extract frequent item set according to the user provided budget and implement a recommendation method to recommend top k list from the generated frequent item set results.

5. Correctness and Completeness Verification

At last data owner verify the encrypted and recommended frequent item set result by computing their precision and recall and budget value.

3.3 Algorithms

Algorithm 1: AES Algorithm

The AES algorithm is used for the security of the dataset. The steps involved in AES are:

1. Key Expansion: - By using Rijndaels key schedule Round keys are derivatives from the cipher key.
2. Initial Round:- AddRoundKey where Each byte of the state is combined with the round key using bitwise xor.
3. Rounds

SubBytes : non-linear substitution

ShiftRows : transposition

MixColumns : mixing operation of each column.
AddRoundKey

4. Final Round: It contain SubBytes, ShiftRows and AddRoundKey

Algorithm 2: Proposed System Algorithm

Input: Dataset D, Minimum Support Ms, Budget Bg.

Output: Verified Frequent Itemset (FIS)

1. Read D
2. Enter MinimumSupport (Ms) and Bg
3. Finding frequent and infrequent itemset :
 $CI_FIS = \text{cal_frequentItemset}(D, Ms, Bg)$;
 $CI_IFIS = \text{cal_infrequentItemset}(D, Ms, Bg)$
4. Probabilistic Approach : Calculate Precision and Recall
 $\text{Precision} = \text{cal_precision}(CI_FIS, SR_FIS)$
 $\text{Recall} = \text{cal_recall}(CI_FIS, SR_FIS)$
5. Calculating Evidence Frequent and Infrequent Itemset

$ev_FrequentItemsetnn = \text{cal_evidenceFrequentItemset}(D, \text{Recall})$

$ev_infrequentItemset = \text{cal_evidenceInfrequentItemset}(D, \text{precision})$

6. Calculating Probability :

$$Pr_R = 1 - \text{Recall}^{evidence(I)}$$

Pr_P= 1 – Precision^{evidence(I)}

7. Calculating completeness of server returned result

8. Deterministic Approach

a) calculating inverted list.

IL=calculate_invertedList(D)

b) construct MHT

MHT = calculate_Merkle hash tree(D)

c) Checking correctness of the result

Verified frequent itemset = Verify(MHT,SR_FIS)

9. Recommendation:

Enter number of top rules, teqrulecalculating :

R_FrequentItemset = give_recomendation

(req_rule, VerifiedFrequentItemset)

3.3 Mathematical Modeling

System S = {Input, process, Output}

- **Input** : I= Dataset

- **Process** :

Probabilistic Approach

P = {F, N}

Where, P is represented as a set of Probabilistic Approach from dataset.

F= {F1, F2, F3,..., Fn}

Where, F is represent as a set of Frequent item set from dataset.

N= {N1, N2, N3,..., Nn}

Where, F is represent as a set of Infrequent item set from dataset.

Deterministic Approach

D= { M }

Where, D is represent as a set of Deterministic Approach from dataset.

M= { M1, M2, M3,..., Mn }

Where, M is represent as a set of generated merkle hash tree and M1, M2, M3,...,Mn are number of nodes.

Frequent Item-set Result :

Fr= { Fr1, Fr2, Fr3,..., Frn }

Where, F is represent as a set of Frequent item set Result.

Recommendation :

R= { R1, R2, R3,..., Rn }

Where, R is represent as a set of Recommendation Result to verify Correctness and Completeness of item set.

C= { C1, C2,..., Cn }

Where, C is represent as a set of Correct and Complete frequent item set according to user budget.

Verification :

Precision :

Let F be the real frequent item sets in the outsourced database D, and F^S be the result returned by the server.

$$P = \frac{|F \cap F^S|}{F^S}$$

i.e., the percentage of returned frequent item sets that are correct.

Recall :

$$R = \frac{|F \cap F^S|}{F}$$

i.e., the percentage of correct frequent itemsets that are returned.

- **Output :**

O = {O1, O2, O3,..., On }

Where, O is represent as a set of Frequent and recommended item set which are correct and complete.

4. RESULTS AND DISCUSSION

Proposed systems experimental setup and results discussed here with the help of comparison graphs.

4.1 Experimental Setup

The framework is built with the assistance of Java framework on Windows platform. The development tool is utilized as Net beans IDE. The framework does not require a specific hardware to run; any standard machine is fit for running the application.

4.2 Expected Result

Table 1 represents the accuracy in % of frequent item set mining outcomes over outsourced data. It is distinctly demonstrated that the proposed system is more accurate than existing system.

Table 2: Accuracy Comparison

System	Memory Required
Existing system Without Budget	65%
Proposed system With Budget	85%

Following figure 2 shows the accuracy comparison of the proposed system with the existing system. Proposed system is more accurate. X-axis represents the system names and Y-axis represents the accuracy in %.

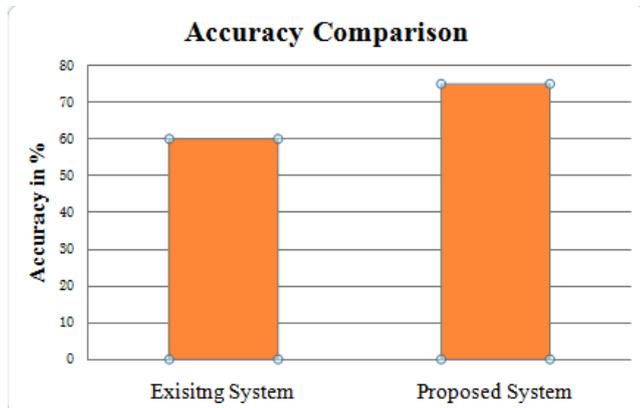


Fig - 2: Accuracy Comparison Graph

Following figure 3 shows the graph of the proposed system with the existing system. Proposed system is more accurate. X-axis represents the retail dataset and Y-axis represents the accuracy in %.

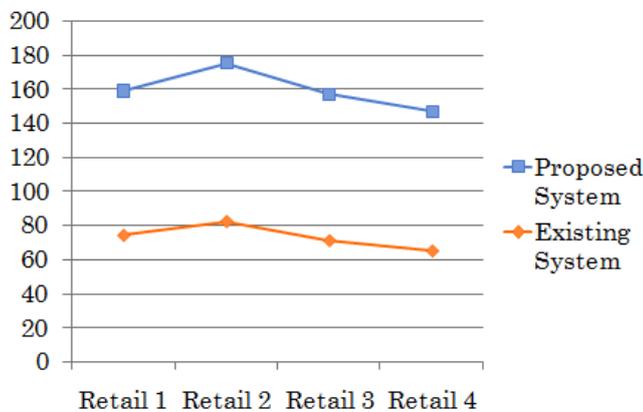


Fig - 3: Accuracy Graph

5. CONCLUSION AND FUTURE SCOPE

We present two integrity verification methods for outsourced frequent item set mining. The probabilistic verification technique constructs evidence of (in) frequent item sets. The deterministic technique requires the server to construct cryptographic evidences of the mining outcome. The accuracy and fullness are measured against the proofs with 100% assurance. Our experiments show the efficiency and effectiveness of our methods. System verifies the accuracy of outsourced frequent item set mining outcomes. It performs the integrity authentication for outsourced frequent item set mining. It recommends most appropriate

item sets to user offered budget. This system gives accurate result and save time as well as memory.

In future the system can enhance to authenticate the data that is store on cloud and to improve the time to retrieve the frequent item sets for user provide budget.

ACKNOWLEDGEMENT

It is my privilege to acknowledge with deep sense of gratefulness to my guide Prof. Kanchan Varpe for her kind cooperation, valuable suggestions and capable guidance and timely help given to me in completion of my paper. I express my appreciation to Prof. Vina M. Lomte, Head of Department, RMDSSOE (Computer Science) for her constant support, suggestions, assistance and cooperation.

REFERENCES

- [1] B. Dong, et al, "Trust-but-Verify: Verifying Result Correctness of Outsourced Frequent Itemset Mining in Data-Mining-As-a-Service Paradigm," in IEEE Transactions on Services Computing, 9(1), pp. 18-32, Jan-Feb 2016.
- [2] Chandrasekharan, et al, "Privacy-preserving frequent itemset mining in outsourced transaction databases," in International Conference Advances in Computing, Communications and Informatics (ICACCI), 2015, pp. 787-793.
- [3] Deokate and Waghmare, "Privacy-Preserving in Outsourced Transaction Databases from Association Rules Mining," in International Journal of Engineering Research and General Science, 2(6), pp. 821-825, 2014.
- [4] Kumbhare and Chobe, "Secure Mining of the Outsourced Transaction Databases," in International Journal of Computer Science Engineering and Technology(IJCSET), 4(1), pp. 250-252,2014.
- [5] Dong, et al, "Integrity Verification of Outsourced Frequent Itemset Mining with Deterministic Guarantee," in IEEE 13th International Conference on Data Mining, Dallas, TX, 2013, pp. 1025-1030.
- [6] Fiore and Gennaro, "Publicly Verifiable Delegation of Large Polynomials and Matrix Computations, with Applications," in Proc. ACM Conf. Comput. Commun. Security, 2012, pp. 501-512.
- [7] Buyya, et al, "Market-Oriented Cloud Computing: Vision, Hype, and Reality for Delivering IT Services as Computing Utilities," in Proc. IEEE Conf. High Performance Comput. Commun., 2008, pp. 5-13.

- [8] Wong, et al, "Security in outsourcing of association rule mining," in Proc. Int. Conf. Very Large Data Bases Endowment, ACM, 2007, pp. 111-122.
- [9] Qiu, et al, "Protecting Business Intelligence and Customer Privacy while Outsourcing Data Mining Tasks," in Knowledge Inform. Syst. Springer- Verlag, London, 17 (1), pp. 99-120, 2008.
- [10] Dong, et al, "Result Integrity Verification of Outsourced Frequent Itemset Mining," in Stevens Institute of Technology, Hoboken, NJ, USA, 2013.
- [11] Canetti, et al, "Practical delegation of computation using multiple servers," in Proc. 18th ACM Conf. Comput. Commun. Security, 2011, pp. 445-454.
- [12] Wong, et al, "An Audit Environment for Outsourcing of Frequent Itemset Mining," Proc. Int. Conf. Very Large Data Bases Endowment, ACM, 2, pp. 1162-1172, 2009.
- [13] Molloy, et al, "On the (In)Security and (Im)Practicality of Outsourcing Precise Association Rule Mining," in Ninth IEEE International Conference on Data Mining, 2009, pp. 872-877.
- [14] Qiu, et al, "An Approach to Outsourcing Data Mining Tasks while Protecting Business Intelligence and Customer Privacy," in Sixth IEEE International Conference on Data Mining, 2006, pp. 2702-2707

BIOGRAPHIES



Miss. Rashmi Kale, ME Second year
Student of R.M. Dhariwal Sinhgad
School of Engineering, Pune