International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 16 (2015) pp 37081-37084

Compact BitTable based Secured Association Rule Mining using Mobile Agent Framework

A.SALEEM RAJA
Research Scholar, Department of Computer Science, Engineering and Technology,
Bharathidasan University,Trichy, Tamil Nadu, India.
Email: asaleemrajasec@gmail.com

E.GEORGE DHARMA PRAKASH RAJ
Assistant Professor, Department of Computer Science, Engineering and Technology,
Bharathidasan University,Trichy, Tamil Nadu, India.

Abstract: Mobile Agent based distributed data mining has drawn attention from researchers in Internet computing recently. Mobile Agent based Distributed Association Rule Mining (MADARM) is task of generating globally strong association rules from frequent itemsets gathered from distributed databases using the intelligent mobile agents. Recently, few researchers proposed framework for secured distributed association rule mining by implementing cryptography algorithms to encrypt the mined data before passing site to site using mobile agents. Yet privacy and security of the mined data is still under research. This paper, we proposed the compact bit table based secured association rule mining using mobile agent framework(CBT-*fi*-SARMMA) which is based on the existing CBT-fi based Distributed Association Rule Mining using Mobile agent framework with security feature.

Keywords: Mobile Agent, Distributed Data Mining, Association Rule Mining

1. Introduction:

Goal of data mining is to identifying patterns and trends from large quantities of data. Most of the mining tools operated on the centralized database. Distributed nature of business databases paves the way of distributed and parallel data mining. Large scale distributed and parallel data mining, requires intelligent miner which can adopt for different mining strategy at each site and integrates the result seamlessly. This requirement integrates autonomous agent with distributed data mining. Agent can be treated as a computing unit that performs multiple tasks based on a dynamic configuration [1]. Integration of distributed data mining with mobile agent creates the new research direction and initiates several research problems in distributed data mining[2] such as reducing communication cost, handling multiple heterogeneous data sources, efficiency of incremental knowledge integration, scalability of the framework, data privacy & security, mobile agent security, fault tolerance and efficient data partitioning. This paper, we enhance the existing CBT-fi based MADARM framework with security feature. We assume homogeneous databases: All sites have the same schema, but each site has information on different entities. The goal has two folds

1. Produce association rules globally from the distributed sites
2. Secure the mined data (that is FI) moving from site to site by the MA.

Rest of the paper is organized as follows. Section 2 presents the overview of Association Rule Mining. Section 3 presents the overview of the mobile agent and existing mobile agent based secured association rule mining. Proposed framework is explained in section 4. Section 5 presents the experimentation and analysis. Finally we conclude the paper in section 6.

2 Association Rule Mining

Association rule mining is one of the data mining techniques, introduced by Agrawal et al in 1993 [3]. It finds the interesting association and/or correlation relationships among large set of data items. Discovering this association rules can guide the decision making. Association Rule Mining includes two major steps such as frequent item-sets (FI) mining and strong association rule generation. But complexity of FI mining is significantly greater than that of association rule generation. A typical and widely used example of frequent item-sets mining is to analyze supermarket transaction data, that is, to examine customer behavior in terms of the purchased products. Frequent sets of products describe how often items are purchased together. In addition to this frequent item-sets mining have applications in areas such as bioinformatics, fraud detection and web usage mining [4]. FIM algorithms generally classified into two types, candidate generation and pattern growth.

- Candidate generation algorithms (e.g. Apriori [3]) generates candidates based on previously identified valid item-sets.
- Pattern growth approaches (e.g. Eclat [6] and FP-growth [5]) eliminates the need of explicit candidate generation with special data structures for database representation and operations.

Apriori, Eclat, FP-growth are the basis of many other algorithms.

3 Secured association rule mining using mobile agents

This section presents the brief overview of mobile agents and existing agent based framework for secured association rule mining from distributed sites. Software Agents refers to intelligent program that performs certain tasks on behalf

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 16 (2015) pp 37081-37084
© Research India Publications. http://www.ripublication.com

of the user. Software agents endowed with the property of mobility are called Mobile Agents [MA]. MA is an autonomous transportable program that can migrate under its own or host control from one node to another in the heterogeneous network to perform a task. In other words, the program running at a host can suspend its execution at an arbitrary point, transfer itself to another host or request the host to transfer it to its next destination and resume execution from the point of suspension. Once the agent is launched, it can continue to function even if the user is disconnected from the network. MA not only moves from one host to another but also spawns new agents; interact with other stationary agents and searches services/resources [7][11]. Agents can support and enhance the knowledge discovery process in many ways. For instance, agents can contribute to data selection, extraction, preprocessing, and integration, and they're an excellent choice for peer-to peer parallel, distributed, or multisource mining. Agents are also a good match for interactive mining, human centered DM, service delivery, and customer service [8]. Few researchers deployed, the MA to mine the association rules from distributed sites with security and privacy. Some of the important contributions in this domain are presented below.

Gongzhu Hu et.al. [9][10] proposed an agent based approach (MARDDMA) to mine the association rules from the distributed data sets across the multiple sites while preserving the privacy of the local datasets. This approach relies on the local systems to find the frequent itemsets that are encrypted and the partial results are carried from site to site. System has one agent server which can communicate to the local hosts through six types of agents (ESUA, DSUA ,ESA, DSA, BA and OA) created and dispatched by it. The agents are defined as

a. Encrypt Secure Union Agent (ESUA)– performs encryption of each locally frequent k-itemset at host i.
b. Dencrypt Secure Union Agent (DSUA)- travels through every host to pursue decryption.
c. Encrypt Sum Agent (ESA)- travels through all the hosts to obtain the encrypted support count.
d. Decrypt Sum Agent (DSA)- carries the array of RuleSet extracted from the returned ESA. It travels through all the host and let each host subtract the random number they generated when dealing with the ESA.
e. Broadcast Agent (BA)- When the DSA agent comes back to agent master the globally frequent k-itemsets (F_k) can be calculated from the decrypted ruleset then BA is used to carry F_k to each host to update their knowledge.
f. Over Agent (OA): Its used to notify all the hosts that algorithm has terminated.

Local Hosts: there are several local hosts where agents can visits and perform their tasks of local association rule mining, encryption and decryption etc.

Saleem et.al[12] presented mobile agent based association rule mining (MADARM) framework, which deploys mobile agent for mining FI in a sequential manner focus to reduces communication overhead between server and agents and security.

Saleem et.al[13] proposed framework, which uses compact bit-table to improve the FI mining from local site and integrate the knowledge based on MADARM framework. Entire framework designed on top of the IBM's Aglets workbench system.

Sharath et.al[14] proposed a system consists of an intelligent encryption and decryption module that transforms the items in the database, C, according to the scheme to an encrypted database C*. The module first encrypts some items in the database using substitution cipher method. In this method, a unit of the database item is replaced with another item. The transformation mapping that is made for each item is stored in a file so as to decrypt the returned result. After encryption, groups the items in the database depending on the total number of items in it. Then irrelevant items are detected and masked by this module. The indexes on each group are then jumbled or shuffled to obtain the transformed database. This encrypted database is send across to the mining server. On receiving a query, the server mines the encrypted data and returns back the result. The EN/DN module on receiving it decrypts and recovers the true patterns.

From the existing system, we observed the following

- Uses more agents to perform distributed mining which increase agent's communication as well as cost of communication.
- Uses existing FI algorithm such as Apriori-T, FP-Tree , FDM and Bit Table for local FI mining and uses the existing encryption algorithm such as substitution cipher, Pohlig-Hellman encryption scheme and RSA to secure the mined data.

4 Proposed framework

This section we present the proposed framework, which is based on the CBT-*fi* MADARM framework with security features. CBT-*fi* algorithm improves the FI mining and one-time padding encryption algorithm performs the secure mining. Entire framework designed on top of the IBM's Aglets workbench system. We begin with problem statements.

a. Problem Statement: Frequent Item-sets Mining (FIM)

Frequent item-sets mining is defined as follows:

Let $T = \{T_i | i = 1 \ldots. n\}$ be the set of transaction in the database D and let I= $\{I_i | i = 1 \ldots. m\}$ be the set of items and each transaction can be identified by a distinct identifier tid.

Definition 1: A set $X \in I$ is called an itemset. An itemset with k items is called a k-itemset.

Definition 2: The support of an item-set x, denoted as $sup(x)$, is defined as the number of transactions in which x occurs as a subset.

Definition 3: For a given D, let min_sup be the threshold minimum support value specified by user. If $sup(x) \geq \min\_sup$, item-set x is called a frequent item-set.

The task FIM is to generate all frequent item-sets in the database, which have a support greater than $\min\_sup$.

b. *Problem Statement: Distributed Frequent Item-sets Mining (DFIM)*

In distributed mining, global frequent item-sets are generated based on the local frequent item-sets collected from distributed sites.

S be the set of sites $S = \{S_i | i = 1 \ldots. n\}$ in distributed environment.

D be the set of horizontally partitioned data sets $D = \{D_i | i = 1 \ldots. n\}$ where D_i is the data set located in S_i.

$$D = \bigcup_{i=1}^{n} D_i$$

KS is the knowledge server where the global frequent item-sets are generated. Using global frequent item-sets, strong association rules will be generated.

LFI_i - Local Frequent Item-sets of site S_i

$LFISC_i$- LFI_i Support Count

GFI-global Frequent Item-sets

Entire transactional database is divided into n partitions $D = (D_i, i = 1 \ldots n)$ horizontally. Partitioned Datasets are located in n remote sites$(S_i, i = 1 \ldots. n)$. The framework contains KS (Knowledge Server), where the global association rule is computed, ***n***-stationary agent (SA) which are located in ***n***-distributed sites, ***n***-key distributor agent (KDA) and Frequent Item set Collect Agent (FICA). Each site has its own private key and stationary agent (SA), which computes the frequent itemset based on thres_val (minimum support count) using CBT-*fi* algorithm.

Functions of each component in the framework are explained below

KS is the knowledge server where the global frequent item-sets are generated. Using global frequent item-sets, strong association rules will be generated. KS generates ***n***-random number (p) and encrypts the p (EP) using private key and send to n-sites using KDA.

FICA launched from KS with three containers such as LFI (local frequent item-sets) , LFISC (local frequent item-sets support count), thres_val . FICA visits each sites $(S_i, i = 1 \ldots. n)$ and collect the encrypted LFI and LFISC. Finally it comes back to KS.

Each local sites, stationary agent (SA) generate the LFI using CBT-*fi* algorithm , once after receiving the FICA. SA decrypt the EP using its private key to obtain the p. Then using p, encrypt the LFI using one-time padding algorithm as shown in the algorithm 2.

Once FICA comes back to the KS, decrypt the local FI using one-time padding algorithm based on p and construct the compact bit table with rcv and bcv is constructed based in LFI, LFISC and thres_val. Using compact bit table with rcv, bcv , GFI is generated as shown in algorithm 1. Finally association rules are computed using GFI.

Algorithm 1: KS

```
function Server(list_of_sites, thres_val)
begin
MA=∅;
vector GFI=∅;
vector LFI=∅;
vector LFISC=∅
p=random(255);
EP=Encrypt(p, private_key);
Launch KDA(EP);
if (visted_sites<>∅)
        MA =launch FICA(list_of_sites, thres_val, LFI,
LFISC);
If(MA <> ∅)
            Begin
                  Decrypt(LFI,LFISC,p);
                  GFI=CBTFI(LFI,LFISC, thres_val) // global
            frequent item generation
            end
end
```

Algorithm 2:SA

```
begin
             Scan database D_i once.
            p=Decrypt(EP, private_key);
            Construct compact bit table using thres_val
            Compute FI with support count (SC)
            Encrypt(LFI,SC,p)
            UpdateFICA(FI,SC, SID)
end
```

5 Experimental Results

Experiments were conducted to show the performance of the two framework in terms of time taken to computing

GFI. We used two synthetic datasets which is based on retail database (http://fimi.ua.ac.be/data/retail.dat). The characteristics of two datasets that we used in the experiments are shown in table 1. Figure 1 and 2 shows the mining time of these dataset by varying threshold. We compare performance of CBT-*fi*-SARMMA with MARDDMA. The results show that CBT-*fi*-SARMMA performs better.

Table 1 Characteristics of two datasets

Name	D	T	N
T15D300N150	300	15	150
T20D600N200	600	20	200
D (Total number of transaction)			
T(Average number of items in a transaction)			
N (Total number of items)			

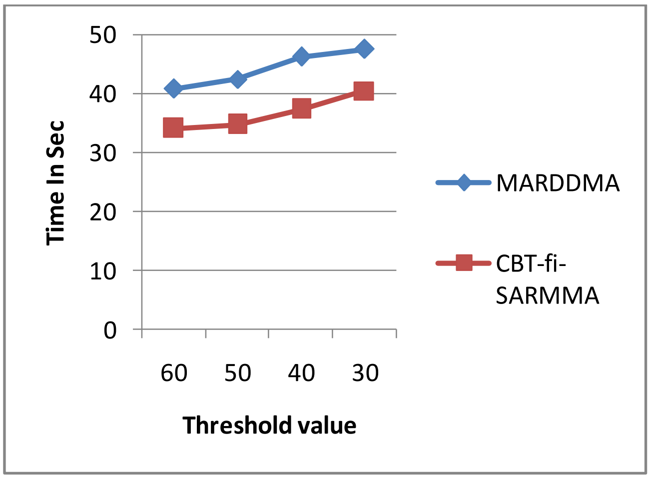

Figure 1. T15D300N150 with 3 sites

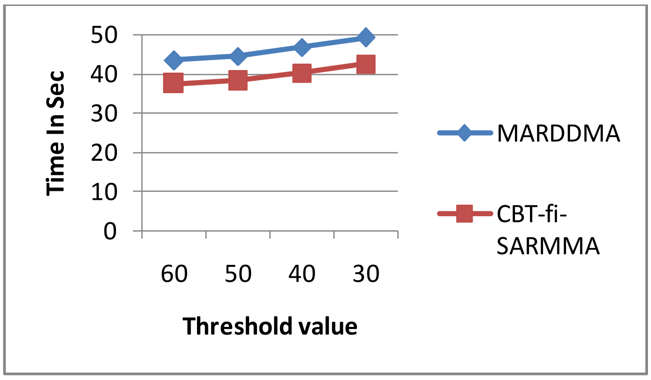

Figure 2. T20D600N200 with 3 sites

Conclusion

In this paper, we present the overview of secured association rule mining using mobile agent in distributed environment architecture and present the proposed CBT-*fi*-SARMMA . Finally, we compare the CBT-*fi*-SARMMA with MARDDMA and result shows CBT-*fi*-SARMMA performs better.

Reference:

[1]. L.Cao(ed.), Data Mining and Multi-agent Integration, DOI:10.1007/978-14419-0522-2_3, pp 47-58, Springer Science + Business Media, LLC 2009.

[2]. A. O. Ogunde, O. Folorunso, A. S. Sodiya, and J. A. Oguntuase, "Towards an adaptive multi-agent architecture for association rule mining in distributed databases," Adaptive Science and Technology (ICAST), 2011 3rd IEEE International Conference on 24-26 Nov. 2011, pp. 31 – 36 from IEEE Xplore.

[3]. R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, in: Proceedings of the 20th International Conference on Very Large Data Bases (VLDB'94), Chile, 1994, pp. 487–499.

[4]. Benjamin, Rainer, Wolfgang, Memory-Efficient Frequent-Itemset Mining, EDBT 2011 ACM, Uppsala, Sweden.

[5]. Christian Borgelt. "An Implementation of the FP-growth Algorithm".In International Workshop on Open Source Data Mining, 2005.

[6]. Christian Borgelt. "Efficient Implementations of Apriori and Eclat".In Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations, 2003.

[7]. G.S.Bhamra, A.K.Verma and R.B.Patel,"Agent Enriched Distributed Association Rule Mining: A Review". Springer Verlag Berlin Heidelberg, 2012.

[8]. Longbing Cao, Vladimir Gorodetsky, Pericles A. Mitkas, Agent Mining: The Synergy of Agents and Data Mining, IEEE Intelligent System,pp 64-72, 2009.

[9]. Gongzhu Hu and Shaozhen Ding, An Agent-Based Framework for Association Rules Mining of Distributed Data, Software Engineering Research, Management and Applications, SCI, vol- 253, pp 13-26, Springer-Verlag Heidelberg 2009.

[10]. Gongzhu Hu and Shaozhen Ding , Mining of Assocation Rules from Distributed Data using Mobile Agents, ICE-B 2009, PP 21-26 (2009).

[11]. G.S.Bhamra, A.K.Verma and R.B.Patel,"Agent based framework for Distributed Association Rule Mining: An Analysis". International Journal in Foundations of Computer Science and Technology (IJFCST), vol 5,no.1, January 2015.

[12]. Saleem and George, "MAD-ARM: Mobile Agent based Distributed Association Rule Mining", ICCCI'13, IEEE Conference 2013.

[13]. Saleem and George, "Compact BitTable based Distributed Association Rule Mining using Mobile Agent Framework",

[14]. Sharath Kumar J. and Maheswari N., "A Framework For Security in Data Mining Using Intelligent Agents", ARPN Journal of Engineering and Applied Sciences, Vol. 10, No. 6, April 2015