MAD-ARM: Distributed Association Rule Mining Mobile Agent

Miss. Rajshree S Hingane[1], Prof. (Dr.) Suryakant Patil[2]

[1,2]Pune University

Abstract*—Rapidly development in the IT field and the problems occurred during the storage of the tremendous data is today's biggest problem. For discovering correlation between the large set of data items the distributed association rule mining plays a very important role. In present the focus of research is going on for improving the efficiency of the algorithm for association rule mining and increasing the speed of the mining process. Many researchers focuses on the improving the efficiency of association rule mining in distributed environment by deploying the intelligent agent for the mining frequent item sets and generate association rule. The existing work for the association rule is based on IDMA, EMADS communication. In our proposed work which is called as MAD-ARM, which attempts to reduce the communication overhead.*

I. INTRODUCTION

MAD-ARM: Distributed Association Rule Mining Mobile Agent.The main purpose of data mining is to discovering the potentially useful information embedded in databases. Data mining purpose this information because of the instantly increasing the data and the need of turning this data into useful knowledge. For this reason the data mining is the topic of attraction. For analyzing the volume of data there are several methods of data mining are implemented. One of the most important types of data mining which was introduced by Agrawal in 1993 is named as Association rule mining. The association rules mining method is found to be most interesting and correlation between large set of data items. This association rule helps in guiding the decision making algorithm. The market base analysis is one of the typical and widely used examples of association rule.

The two major processes which were included by the association rule mining is as follows:

1. Frequent item generation: the frequent item set is called as the item sets which satisfy the minimum support of threshold.
2. Rule generation: the rule generation is also called as the strong rules, because this rules all the high confidence rules which were found in the first step of the association rule.

Our main focus is on the frequent item sets mining which is greater than association rule mining in complexity.

II. LITERATURE REVIEW

In mining association rules between sets of items in large databases paper they present a large database of customer transactions. Each transaction consists of items purchased by a customer in a visit. They present an efficient algorithm that generates all significant association rules between items in the database. The algorithm incorporates buffer management and novel estimation and pruning techniques. They also present results of applying this algorithm to sales data obtained from a large retailing company, which shows the effectiveness of the algorithm.

Association Rule Mining in Distributed Environment paper proposed framework will attempt to reduce the communication overhead. For providing more security against unauthorized clients, we are using RC4 algorithm for encryption and decryption of messages which is going to be passed between the clients.

Parallel mining of association rules consider the problem of mining association rules on a shared nothing multiprocessor. We present three algorithms that explore a spectrum of trade-offs between computation, communication, memory usage, synchronization, and the use of problem specific information. The best algorithm exhibits near perfect scale up behavior, yet requires only minimal overhead compared to the current best serial algorithm.

In the Agent based distributed data mining: The KDEC Scheme briefly review and classify existing approaches to agent-based distributed data mining, propose a novel approach to distributed data clustering based on density estimation, and discuss issues of its agent-oriented implementation. The agent based approach for mining the strong association rules from distributed data sources highlighted by the recently proposed framework, AeMSAR (Agent Enriched Mining of Strong Association Rules).

An investigation into the issues of Multi-Agent Data Mining paper discusses a number of research issues concerned with the viability of Multi-Agent systems for Data Mining (MADM). The problem addressed by this thesis is that of investigating the usefulness of MAS in the context of DM. This thesis also examines the issues affecting the design and implementation of a generic and extendible agent-based data mining framework. Another method for promoting the ideas of high availability and high performance without compromising data or DM algorithm integrity the method which is promote this functionality is known as extensible multi agent data mining system.

III. DISTRIBUTED ASSOCIATION RULE MINING

To notice the need for parallel and distributed association rule mining

1. Including security and fault tolerance etc data are inherently distributed for diversity of practical reasons.
2. a lot of the association rule mining algorithm needs all data to be occupant in memory. This might be impracticable for large datasets. The one of the popular solution for this problem is Data partitioning.
3. Process high dimensional datasets, characteristically obtainable as input to the problem of association rule detection needs multiple processors parallel system.

a. Parallel and Distributed Association Rule Mining for Algorithms

These algorithms are based on the Apriori algorithm. Task parallelism and data parallelism are the normally used parallelism classes. The algorithms used for this association rule mining's are

1. Count Distribution (CD):- It is the representative algorithm of the data parallelism [3].
2. Data Distribution (DD):- It is the representative algorithm of the task parallelism [3].
3. Candidate Distribution:- Its efforts to reduce the synchronization and communication overhead in the count distribution (CD) and data distribution (DD) by selectively partitioning the database such that each processor has locally the data needed to process its candidate set.

IV. AGENT BASED DISTRIBUTED DATA MINING

To integrating the knowledge which is discovered out of data at different geographically distributed network sites with minimum amount of network communication and maximum of local computation [4] is the objective of many existing distributed data mining. Few architectural models were proposed which are still very open for further research to achieve this objective. In general, there are two architectural model used in the development of Distributed Data Mining (DDM) system namely.

1. Client-Server (CS) :- The client side requests are sent to DM server that determines the required data sources and collects data from different locations and brings all the required data for the specified mining process to the DM server.
2. Software Agent (SA):- refers to intelligent program that performs certain tasks on behalf of the user. Software agents endowed with the property of mobility are called Mobile Agents [MA].

The advantages of the MA model over Client-Server model are:

- Increasing scalability.
- Reduces the network traffic and save the network bandwidth by allowing only shared resultant data to be carried over the network.
- MA can operate asynchronously.

Multi-agent systems (MAS) are communities of software entities, operating under decentralized control, designed to address (often complex) applications in a distributed problem solving manner. MAS can make use of algorithms without necessitating their transfer to users, thus

contributing to the preservation of intellectual property rights Albashiri et al. (2010).

V. RELATED WORK

Here in this section we present the survey on mobile agent based distributed association rule mining.

IDMA Wang et al. (2003) architecture, this architecture shows association rule mining for both mobile agent based and incremental based mining. The discussed system included the system which is known as knowledge discovery management system, the knowledge discovery sub system, the data mining mobile agent and the last one local knowledge based.

The knowledge based management system dispatches the mobile agent data mining mobile agent to each site. After that for executing the mission of the data mining, the above mobile agent moves to the knowledge discovering sub system. For refreshing the local knowledge based data and for getting the local large dataset we obtained the local association rule mining.

The mobile agents counts the all the set of local large tem sets and their support for the knowledge discovery management system. After collecting all the mobile agents to the knowledge distributed management system the potential global items set can be calculated by counting the minimum and maximum support of the item sets.

For reducing the time to compute global frequent item sets [10] The mobile agent based distributed data mining method used an algorithm, the steps of this algorithm is as follows:

1. Mining the local frequent item sets at each and every sites and send this mining to the central site for calculating global frequent item sets.
2. In the above step we calculate the central site, this site is calculating the candidate global frequent item sets and this calculated item sets sends to the all sites. Local sites which accept this values calculates the candidate global frequent item sets and after calculating send the item sets back to the central sites for completing the process of finding global frequent item sets.

Another method for promoting the ideas of high availability and high performance without compromising data or DM algorithm integrity the method which is promote this functionality is known as extensible multi agent data mining system Coenen et al. (2010). The mention framework is very flexible and extensible for data mining platform. For building the collaborative DM approaches the proposed method is allowed for this operation. There are number of DM scenarios which the proposed framework applied. The scenarios are: classifier generation and Meta association rule mining.

Distributed BitTable Multi-Agent Association Rule Algorithm (Distributed BMAS algorithm) Hossain et al. (2011) combines the association rules as a data mining technique, the Bitable data structure that was proved to be a very efficient data structure for mining frequent itemsets and the Multi-agents technique to decrease the time needed for candidate generation and the support count processes. The BitTable data structure was implemented before the first iteration. This had a great impact on the algorithm performance. This algorithm was implemented and tested on two synthetic centralized datasets and five distributed real world benchmark datasets and has proved better performance and execution time.

The agent based approach for mining the strong association rules from distributed data sources highlighted by the recently proposed framework, AeMSAR (Agent Enriched Mining of Strong Association Rules) Bhamra et al (2012). This framework consists of one central site ($S_{CENTRAL}$) where global knowledge is computed and n distributed sites where horizontal partitioned transaction datasets are stored. Synthetic Transactional Data sets are generated and stored at each distributed site using a tool called TDSGenerator. $S_{CENTRAL}$ acts as the agent launching station from where mobile agents are dispatched carrying some information and returned back with results. Mobile as well as Stationary agents are stored in Agent Pool at this site. A Central Security Agency (CSA) at this site assigns a legal certificate to each mobile agent previous to its launch and when that agent reaches at a node in its route genuineness of this certificate is verified again so that no malicious agent can attack local node. There are five agents in the architecture, three of these are MAs and other two are stationary intelligent agents to perform different tasks. Mobile Agents are:

1. Local Frequent Itemset Generator Agent (LFIGA)
2. Local Knowledge Generator Agent (LKGA),

3. Total Frequent Itemset Collector Agent (TFICA).

These agents maintains dynamic itinerary, whenever required this can be updated at any node at any time in the itinerary. These agents maintain two containers- Result container and State container. One for transporting result data across the network and other for state variables and their intermediate values. Stationary Agents are:

1. Global Frequent item set Generator Agent (GFIGA) and
2. Global Knowledge Generator Agent (GKGA).

VI. MAD-ARM

This section presents the proposed framework called Mobile Agent based Distributed Association Rule Mining (MAD-ARM) on transactional data. This framework constructs on IBM's Aglets Workbench System. The main objective of this paper is, to reduce the communication cost involved in MADM framework and also guarantee the agent security.

In our MAD-ARM frame work, we used horizontally partitioned dataset across the site, which are homogenous. This framework consists of one knowledge server (K-Server) where the global association rule is computed, some stationary agent (SA) and one mobile agent. Entire transactional database is divided into n partitions MAD-ARM based on IDMA and AeMSAR framework. Partitions are located in n remote sites. Each site has stationary agent, which computes the frequent itemset based on the min_sup (minimum support count). K-Server launches the one mobile agent (MA) which carries secret-key, list of sites (LS), empty GFIL (Global Frequent Itemset List) and min-sup.

This MA visits the each site as per the LS. In each site, SA verifies the key and GFIL then it updates the GFIL. Once this process is completed the MA will move to the next site. Finally it brings back the GFIL to the K-Server. Based on the GFIL, Association Rule will be generated. In this proposed framework GFIL is computed in each site. MA carries GFIL to each site so it reduces communication between K-Server and each site.

VII. CONCLUSION

In our proposed work, we describe about the basic overview of distributed association rule mining and also we discussed about the architecture of agent based distributed data mining. We also discussed about the framework of the existing method of the data mining. In our proposed work we designed the framework for mobile agent based distributed association rule mining. The framework is designed for reducing the communication overhead. This framework ensures the security for the mobile agent.

REFERENCES

[1] O. Ogunde, O. Folorunso, A. S. Sodiya and G. O. Ogunieye, " A review of some issues and challenges in current agent based distributed association rule mining", Asian Journal of Information Technology, 2011.

[2] E. I. Ariwa, M. B. Senousy and M. M. Medhat, "Information and E-business model application for distributed data mining using mobile agents", Proceedings of the international conference WWW/Internet, USA,2003.

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

[4] Jaturon Chattratichat, John Darlington, Moustafa Ghanem, and ET. al, "Large Scale Data Mining: challenges and Responses", Proceedings of the 3th International Conference on Knowledge Discovery and Data Mining, 1997.

[5] Kamala Ali Albashiri, Frans Coenen, and Paul Leng, "An investigation into the issues of Multi-Agent Data Mining" Ph.D-Thesis 2010.

[6] Matthias Klusch, Stefano Lodi and Gianluca Moro, "Agent based distributed data mining: The KDEC Scheme".

R. Agrawal, T. Imielinski, and A. Swami, "Mining Associations between Sets of Items in Massive Databases," Proceedings of the ACM SIGMOD, Washington DC, 1993.

[7] Rakesh Agrawal and John C. Shafer, "Parallel Mining of Association Rules", IEEE Transactions on Knowledge and Data Engineering, 1996.

[8] U. P. Kulkarni, P. D. Desai, Tanveer Ahmed, J. V. Vadavi and A. R. Yardi, "Mobile Agent Based Distributed Data Mining", ICCIMA, 2007.

[9] Walid Adly Atteya, Keshav Dahal and M. Alamgir Hossain, "Distributed BitTable multi-agent

Association Rules Mining Algorithm" , Springer-Verlag, KES 2011, Part I, LNAI 6881.

[10] Yun-Lan Wang, Zeng-Zhi Li and Hai-Ping Zhu, "Mobile Agent Based Distributed and Incremental Techniques for Association Rules". In Proceeding of the Second International Conference on Machine Learning and Cybernetics", 2003.