International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 05 Issue: 05 | May-2018 www.irjet.net p-ISSN: 2395-0072

FOR MULTI AGENT DATA MINING IN PARALLEL AND DISTRIBUTED SYSTEM ESTIMATION SCHEME

Mr.D.karthik[1]. Mr.R.Sengamuthu [2] Mrs.R.Abirami[3]
[1, 2, 3] *Assistant Professor,*
Department of Computer Science, Govt Arts College ,Ariyalur

Abstract - *Multi agent data mining in parallel and distributed systems has been studied in various situations and they suffers with the problem of data availability, because they categorize the network nodes according to the type of data the nodes has and suffers with the accuracy of result producing according to the query submitted. We propose a modern approach which selects a set of nodes from where the data has to be retrieved based on multi attribute data availability scheme. The proposed approach identifies the set of values necessary to compute the query and based on that we compute the data availability of each node available in the network. Based on the data availability a set of nodes with more data availability factor is selected and number of agents is scheduled according to the data availability factor and the query is being executed. The proposed approach generates number of agents according to the number of locations being sorted and the agents are moved to the location to perform the specific query. The proposed approach has produced efficient results in all the factors of data mining like data reliability, data relevancy with less time complexity.*

Keywords: Data Mining, Multi Agent Approach, Parallel and Distributed Systems, Multi Attribute Data Availability.

1.INTRODUCTION

The amount of information is growing in day to day basis, and the information are distributed throughout the internet. The internet is the collection of network and each has different number of nodes and the information are distributed throughout the network and each node can represent according to the data it has. Even though the data nodes are distributed the retrieval system makes the difference. The retrieval is the process of extracting the information from the large set of distributed environment. The efficiency of the retrieval system is based on how relevant the system is producing the result.

Data Mining is the process of extracting useful information from large set of data set. In distributed environment the data will be available in various nodes in different nodes and the process of identifying the location of the data and retrieve then according to the query is defined as data mining. The parallel and distributed system is one which consists of number of nodes which are located in different positions and geographically. The geographically located nodes has various information about different concepts and to extract an information from the data nodes, the location of the data and the data nodes which has the data has to be identified.

The parallel and Distributed network has no fixed topology and the nodes are distributed throughout the network and each node may have various information's and there is no condition about the type of information the node has. Also in parallel and Distributed Systems, the query submitted has to be executed in different machines at the same time in order to reduce the execution time. The multi agents system is the platform where we can maintain any number of agents. Each agent has the behavior of mobility, that they can move from one location to another where the data is available. Using these agents we can perform query execution at multiple locations at the same time to reduce the time complexity of the system designed. The Java Agent Development Environment (JADE) is the platform which provides mobility of agents and can be controlled using Agent Control Messages. Using the language the agent can be moved from locations to other easily and they work based on events generated.

Whatever the node has information, the efficiency of the retrieval system is highly depending on availability and reliability of the query retrieval. So in order to improve the query efficiency the availability of the data has to be estimated and so that the efficiency of the retrieval could be efficient one.

2. Related Works

There are many approaches has been discussed for the retrieval of data from parallel and distributed systems and we discuss few of them here in this paper.Energy Efficient Wireless Sensor Network using Genetic Algorithm based Association Rules [1], considers the amount of data sent to sink is reduced using association rule mining and in turn to further reduce the energy consumption of the network; optimal routes are chosen to transmit data to the sink based on energy consumption. The proposed method is able to discover the association rules to make predictive analysis on node failure, asymmetric links. The rules found form the basis for coding solutions in the proposed genetic algorithm. GA is applied to generate balanced and energy efficient data aggregation spanning trees for wireless sensor networks.Span, which is an energy-aware spanning tree algorithm and Lifetime-Preserving Tree (LPT) are used to create data aggregation trees. The proposed GA extends

network lifetime. Energy-Efficient and Coverage-Aware Clustering in Wireless Sensor Networks [4], proposes a general framework that addresses both metrics for clustering algorithms in wireless sensor networks. The proposed framework is based on applying the principles of Virtual Field Force on each cluster within the network in order to move the sensor nodes towards proper locations that maximize the sensing coverage and minimize the transmitted energy. Two types of virtual forces are used: an attractive force that moves the nodes towards the cluster head in order to reduce the energy used for communication and a repulsive force that moves the overlapping nodes away from each other such that their sensing coverage is maximized. The performance of the proposed mechanism was evaluated by applying it to the well-known LEACH clustering algorithm.

Fuzzy-Genetic Algorithm Based Association Rules for Wireless Sensor Data [5], proposed to mine WSN data using association rule to extract patterns. Genetic algorithm and Fuzzy logic is used with association rule for effective extraction of rules from quantitative data. The proposed method uses INTEL dataset for evaluation. Fuzzy logic and Genetic algorithm is used with association rule for effective extraction of rules.

An object tracking scheme for wireless sensor networks using data mining mechanism [6], proposed an object tracking scheme for OTSNs using data mining approach. We have improved the Apriori algorithm for mining association rules and made it applicable to the OTSNs. The data mining algorithm is applied to the past movement information of the object and useful association rules are excavated, which are then used to predict the next location of the object. Our scheme predicts the next location of the object more accurately and increases the network lifetime. Experimental results have been conducted to evaluate the performance of our proposed scheme for OTSNs and they show that our scheme outperforms the existing schemes in terms of energy efficiency and accuracy of tracking.

Data Partitioning and Association Rule Mining Using a Multi-Agent System [8], explores and demonstrates (by experiment) the capabilities of Multi-Agent Data Mining (MADM) System in the context of parallel and distributed Data Mining (DM). The exploration is conducted by considering a specific parallel/distributed DM scenario, namely data (vertical/horizontal) partitioning to achieve parallel/distributed ARM. To facilitate the partitioning a compressed set enumeration tree data structure (the T-tree) is used together with an associated ARM algorithm (Apriori-T). The aim of the scenario is to demonstrate that the MADM vision is capable of exploiting the benefits of parallel computing; particularly parallel query processing and parallel data accessing. In addition the approach described offers significant advantages with respect to computational efficiency when compared to alternative mechanisms for (a) dividing the input data between processors (agents) and (b) achieving distributed/parallel ARM.

Multiagent System for Pattern Searching in Billing Data [9], present an agent-based pattern searching system using a distributed Apriori algorithm to analyse billing data. In the paper, we briefly present the problem of pattern mining. Next, we discuss related research focusing on distributed versions of Apriori algorithm and agent-based data mining software. Paper continues with an explanation of architecture and algorithms used in the system. We propose an original distribution mechanism allowing to split data into smaller chunks and also orthogonally distribute candidate patterns support calculation (in the same computation task). Experimental results on both generated and real-world data show that for different conditions other distribution policies give better speedup. The system is implemented using Erlang and can be used in heterogeneous hardware environment. This, together with multi-agent architecture gives flexibility in the system configuration and extension.

A Multi-Agent System for Context-Based Distributed Data Mining [10], describes an approach that aims to resolve this issue. Focusing on a key business problem--the prediction of customer behaviour--it presents a distributed multi-agent framework that deals with context heterogeneity via hierarchical modeling. The main elements of this work are to (1) provide a solution to the contextual heterogeneity problem in distributed data mining and (2) design and implement a hybrid distributed system for the proposed distributed data mining approach.

Extracting Peculiar Data from Multidatabases Using Agent Mining [11], discusses the peculiar data mining and agent mining. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Agent Mining: The Synergy of Agents and Data Mining [13], give an overall perspective of the driving forces, theoretical underpinnings, main research issues, and application domains of this field, while addressing the state-of-the-art of agent mining research and development. Our review is divided into three key research topics: agent-driven data mining, data mining-driven agents, and joint issues in the synergy of agents and data mining. This new and promising field exhibits a great potential for groundbreaking work from foundational, technological and practical perspectives.

A Bounded and Adaptive Memory-Based Approach to Mine Frequent Patterns From Very Large Databases [12], could use only a bounded portion of the primary memory and this gives the opportunity to assign other parts of the main memory to other tasks with different priority. In other words, we propose a specialized memory management system which caters to the needs of the ARM model in such a way that the proposed data structure is constructed in the available allocated primary memory first. If at any point the structure grows out of the allocated memory quota, it is forced to be partially saved on secondary memory. The secondary memory version of the structure is accessed in a block-by-block basis so that both the spatial and temporal

localities of the I/O access are optimized. Thus, the proposed framework takes control of the virtual memory access and hence manages the required virtual memory in an optimal way to the best benefit of the mining process to be served. Several clever data structures are used to facilitate these optimizations.All the above discussed methods have the problem of data retrieval and its accuracy about time complexity.

3.Proposed Method

There are many stages present in the proposed approach namely Agent Scheduler, Multi Attribute Data Availability Estimation, Query Processor and so on. We discuss each of the functional components in detail in this section.

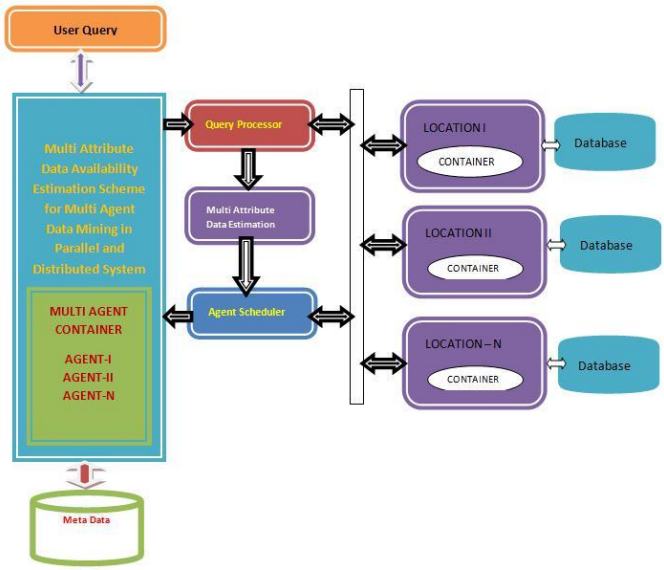


Figure 1: Proposed System Architecture

Query Processor:

The input query submitted by the user is handled by the query processor, and from the input query we identify the set of objects and values required to perform the query processing. There are number of data objects required to perform the query processing and the method maintains the meta data of data present in the different nodes of the network. With the identified required data objects and the meta data the method performs the multi attribute data availability using which the query processor choose set of locations. For the locations selected to perform the query processing, number of agents will be computed and the agents are optimized according to the data availability and the distance parameters to execute the query. finally the executed results will be given to the user as a final result.

Algorithm:

Input: Meta Data Md, Query Q.

Output: Result Res.

Step1: start

Step2: Identify set of all data objects required from the query Q.

Objects ob = ∑ ( ) ∈

Step3: for each object o from Ob

Identify set of location where it is available.

Data locations Dl = ∫ ( ) ∈ =1

end

Step4: for each location L from Dl compute data availability Dam. End

Step5: Choose set of locations with most DAM.

Step6:perform agent handling.

Step7: Return results.

Step8: stop.

Multi Attribute Data Estimation:

The multi attribute data estimation is the process of estimating the data availability of each node present in the network. The input query may require number of values but the data location may have only limited results and has to be estimated. For each location identified we estimate the data availability using the number of values need to process the query and the number of values the location has. The computed values will be given to the other process to perform agent handling.

Algorithm:

Input: Object set Os, Location L.

Output: Data availability measure DAM.

step1: start

Step2: Compute number of data values present in the location L.

Nop = ∫ ∑ ∈

Step3: Compute number of data values available in L.

Noa = ∫ ∑ ( ) ∈

Step4: Compute availability measure DAM.

DAM = × 100

Step5: stop.

Agent Handling:

Agent handling is the process of generating number of agents according to number of locations where the data is available. Unlike other methods, the proposed method generates number of agents based on the data availability measure (DAM) and the number of instances available in the data location and the distance between the other locations. Sometimes there are locations which has little amount of information and the locations may be very closure. Generating a specific agent to the little data constrained network reduces the resource utilization and to remove this, the method generates agents and initialize them with different locations according to the data availability and the number of close locations. Generated agents are moved to the remote container and process the queries to return the results to the user.

Algorithm:

Input: Locations L, Data Availability Measures DAMS.

Output: Results Res.

Step1: start

Step2: for each location l from L

create agent Ag.

Initialize agent with location l.

Ag(location)=l.

if DAMS(l) < DTh then //data threshold
Identify neighbor locations with less
Dam.

Add location to the Agent.

end

end

Step3: Move all agents to the remote location.

Step4: Return the results to the user.

Step5: stop.

4.Results and Discussion

The proposed multiagent based data mining technique using multi attribute data estimation for parallel and distributed systems has been implemented using Java agent development environment with 500 locations and containers. We have tested the proposed approach with various numbers of locations and various size of data locations. The proposed approach has produced efficient result both in time complexity and result efficiency.

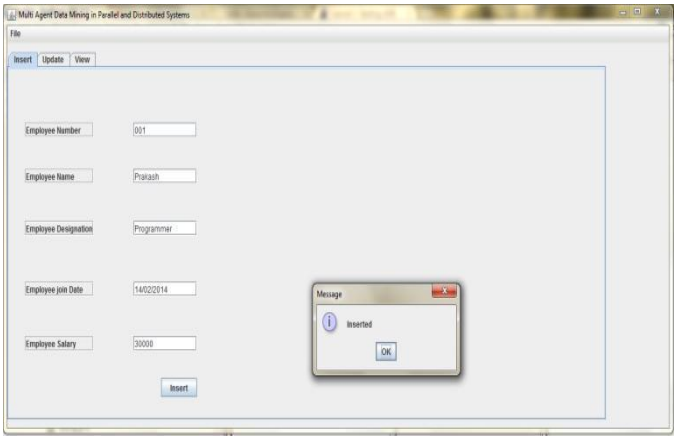


Figure 2: snapshot of result produced by proposed method.

The figure2, shows the snapshot of result produced by the proposed method, where the data is updated in different locations of distributed network using number of mobile agents.

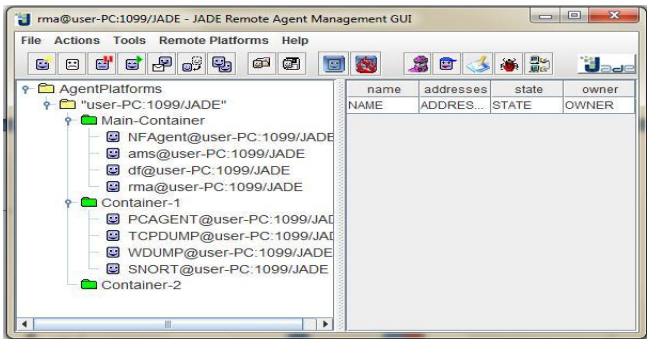
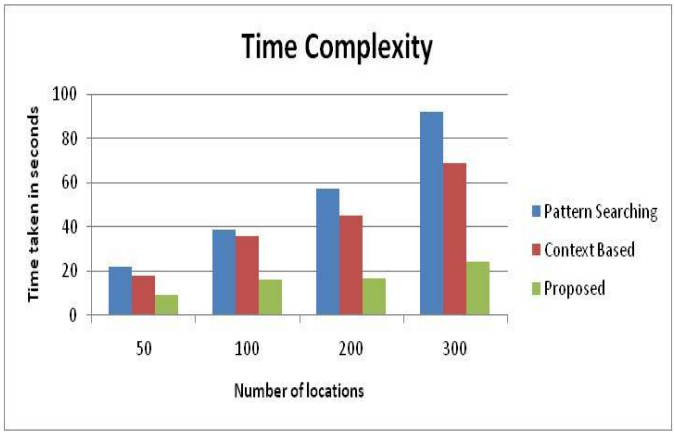


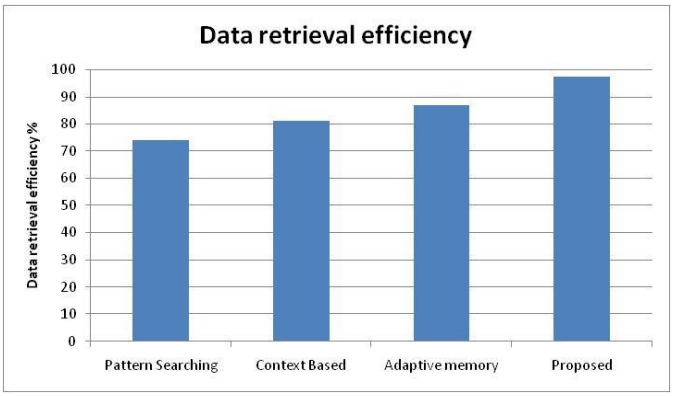
Figure 3: Snapshot of agent containers

The figure3, shows the snapshot of agent container and number of agents available at the container of the proposed system.



Graph 1: shows the time complexity of different approaches.

The graph1 shows the time complexity produced by different algorithms, it shows that the proposed approach has less time complexity than other pattern searching and context based approaches.



Graph 2: Comparison of data extraction accuracy

The Graph 2, shows the comparison of data extraction accuracy of different methods. It shows clearly that the proposed method has produced more accuracy than others.

Conclusion

We proposed a naval multi agent approach for data mining in parallel and distributed systems using multi attribute data estimation scheme and agent scheduling approach The method has process the query text to identify the data objects necessary to complete the query. For identified data objects, set of locations are identified using meta data available and for each location available we compute the data availability measure. Based on the DAM and the locations number of agents are generated and initialized with number of locations. The initialized agents are moved to perform query processing and the results will be returned to the user. The proposed method has produced efficient and accurate results with low time complexity.

References

[1]Abirami and P Priakanth. Article: Energy Efficient Wireless Sensor Network using Genetic Algorithm based Association Rules. International Journal of Computer Applications 91(10):8-12, April 2014.

[2]hong, S. K., Gaber, M. M., Krishnaswamy, S. and Loke, S. W. "Energy Conservation in Wireless Sensor Networks: A Rule-Based Approach", Knowledge and Information Systems, Vol. 28, No. 3, pp. 579-614, 2011.

[3]araiya, K., Kant, K. and Gupta, N. "Wireless Sensor Network: A Review on Data Aggregation", International Journal of Scientific and Engineering Research, Vol. 2, No. 4, pp. 1-6, 2011.

[4]ahedAwad, Energy-Efficient and Coverage-Aware Clustering in Wireless Sensor Networks, Scientific Research, Wireless Engineering and Technology, 2012, 3, 142-151

[5]Abhirami, Fuzzy-Genetic Algorithm Based Association Rules for Wireless Sensor Data International Journal of Computer Science and Telecommunications Volume 3, Issue 11, November 2012.

[6]en-Hwa Liao, An object tracking scheme for wireless sensor networks using data mining mechanism, Network Operations and Management Symposium (NOMS), 2012.

[7]uhaimenul Adnan and Reda Alhajj, "A Bounded and Adaptive Memory-Based Approach to Mine Frequent

[8]Patterns From Very Large Databases", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 41, NO. 1, FEBRUARY 2011.

[8]amal ali anshari, Data Partitioning and Association Rule Mining Using a Multi-Agent System, International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2, Issue 5, September 2013.

[9]Lucauze Bevan, Multiagent System for PatternSearchinginBillingData,Multimedia Communications, Services and Security Communications in Computer and Information Science Volume 368, 2013, pp 13-24

[10]Yang zing, A Multi-Agent System for Context-Based Distributed Data Mining,