A Comparative Study of Intelligent Agent Techniques for Distributed Data Databases

M Dukitha and A Banumathi

December 5, 2019

A COMPARATIVE STUDY OF INTELLIGENT AGENT TECHNIQUES FOR DISTRIBUTED DATA DATABASES

[1]Mrs.Dukitha.M, [2]Dr.A.Banumathi

[1]Assistant Professor, [2]Assistant Professor,

[1]Department of MCA, Er.Perumal Manimekalai College of Engineering, Hosur

[2]Department of Computer Science, Government Arts College, Karur

[1]dukitham@gmail.com, [2]Banukarthikeyan7811@gmail.com

[1]9486344558 ,[2]75986 03260

Abstract

Intelligent agent-based systems have become a growing approach to decision-making in the business applications. Analyzing with the maximum knowledge enriches the rapid and accurate decisions in the competitive business environment. With the explosive development and growth of accessing information from the internet makes integration of multiple database data from diverse locations gives voluminous database and constructive effects in the decision making. Distributed data mining can accomplish a mining task with databases in diverse locations. Instead of moving of agents to all locations moving of agents to maximum density data locations is time consuming, space consuming and also increase of accuracy in decision making. This paper focuses on the comparative analysis of using intelligent-agent for distributed data mining, and density estimation based location selection for distributed data mining.

Keywords-Intelligent–agents, Distributed data mining,Density estimation,Decision making, Location selection

1. INTRODUCTION

Data Mining is used to extract and discover hidden patterns, correlations, and relationships within the data which are used to predict the behaviors and future trend analysis for organizational decision making. Applications are in different scenery and from different capacity. For accurate and globalized decision-making information to be extracted from multiple databases from distributed locations. The external data helps to improve the quality of decisions. Distributed Data Mining (DDM) is the global data from decentralized data sources. The field of DDM deals with the challenges in analyzing distributed data and DM takes place both locally at each distributed sites and integrates the local knowledge into a global model in order to discover global knowledge. Classification accuracy is based on a global model created within a given strategy. The aim of the DDM is to extract a useful pattern from the distributed heterogeneous database for the purposes of decision making. Figure 1 shows the distributed data mining framework.
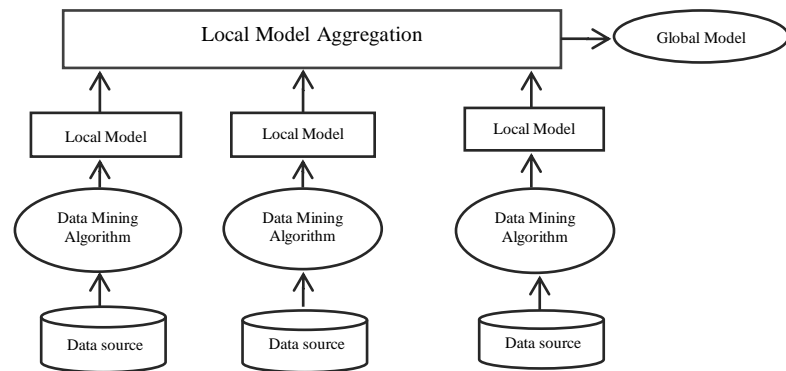
Figure 1 Distributed Data Mining Framework

Retrieving of data from distributed databases can be done by intelligent agents. An Intelligent agent is an autonomous software agent or a computer program that carries out tasks on behalf of users and without the intervention of the users. An agent consists of two parts: the code part, which compiles the instructions that define the behavior of the agent and its intelligence, and the current state of execution of the agent. All agents are autonomous, has its degree of control over its own actions. Agents have a task and act in accordance with that task. An agent can coordinate their actions, cooperate with other agents and act on their own idea to achieve their task and also adapt their behavior based on the changing environment. The mobile agents move from machine to machine to be closer to the data they may need to process and do so without network delays.

All the locations may not have sufficient data. Moving agents to all the locations is a time consuming process. The problem of decision making is identifying specific location where the data is present for the specific query. Using Intelligent – Agent for density estimation based Location Selection (DELS) method is the one which performs the selection of optimal location based on density estimation of attributes. It uses various metadata (MD) to represent the availability of data in different locations and computes the attributes density estimation. The agent can move to a maximum density location and fetch the data for analysis and knowledge discovery.

2. RELATED WORKS

The researchers are giving more attention in developing agent based distributed system and also efficient retrieval of data from optimal locations. Few of them are

Liu et al (2011) proposed a framework for an automated data mining system based on intelligent agents. This framework provides the base knowledge about agent system and for the developing intelligent agent-based automated data mining system for ordinary users who are lacking the knowledge of data mining.

The framework called mobile agent based distributed association rule mining (MAD-ARM) is presented, which attempts to reduce the communication overhead and ensures the mobile agent security.

Laftah Al-Yaseen et al (2015) proposed three agents, namely, coordinator, analysis, and communication agent. The basic concept underpinning the utilized MAS is dividing the largely captured network dataset into a number of subsets and distributing these to a number of agents depending on the data network size and core CPU availability. The results show that compared to the current methods, the Intrusion detection system in a multiagent system (MAS-IDS) reduces the IDS processing time by up to 70%, while improving the detection accuracy.

An intelligent agent is implemented to equalize the human behavior and beliefs. Ronald & Sterling (2005) designed the agent-based model of pedestrian behavior in a real world environment. It defines the unpredictable nature of human decision making. Human decision exhibits different behaviors depending on their knowledge on that environment and other personal characters. BDI agent-based architecture model is implemented in this simulation which is useful for high-level decision making. Here the advantage of the agent is found that the capable of doing several things concurrently without trouble.

3. INTELLIGENT   AGENT   FOR DISTRIBUTED DATA MINING

In modern years, intelligent agent perceptions have been applied in decisive support systems (DSS) for trade users to solve large-scale conventional problems by using DDM. The intelligent concepts applied in distributed decision support system (DDSS) encapsulate the containers in different locations and perform a flexible and autonomous action to achieve the desired objectives. The major significance in using agent-based DDSS are, it improves timely decision making, consistency in decisions, explanations and justifications for specific recommendations, enhanced management of uncertainty, and formulation of organizational knowledge. Examples of medical harms that can be determined by intelligent DDSS are medical diagnostics amplification, patients medical data congregation, medical knowledge exploration, and medical decision support, the pro-active assistance of the physicians, reducing medical errors, improving patients' safety and decreasing cost.

A single agent is not expertise with all the task of problem-solving. Different agents are experts in managing different areas of problem solving, and it can coordinate with other agents to achieve the desired task. At the time of the query, the DDSS created and migrates the agents to different containers and perform the data processing. MAS in DDSS is composed of users interface agent, knowledge management agent, task management agent, communication agent, data mining agent, rule-definer agent, knowledge discovery agent and preprocessing agent. All these systems are normally abstracted in four modules; user module, management module, a processing module and resource module. Figure 2 shows the Intelligent - agent in distributed systems.
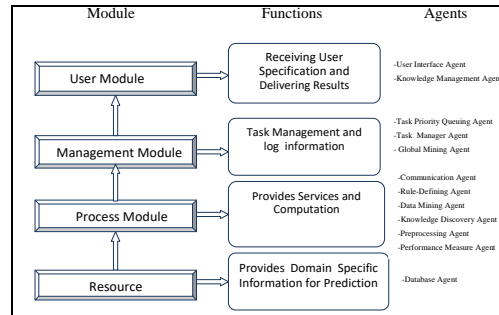
Figure 2 Intelligent-Agent Paradigm

The DDM model has a number of functional components and stages of the decisive making process as below:

- Agent Initialization
- DDF-Distributed Data Fetching
- Pre-processing
- Classification of data for decision making

3.1 Agent Initialization

An Agent initialization is the creation and managing agents. An Agent Management can have the ability to create any number of agents to comply with the query received from User Interface Agent and perform the data mining task.

An agent platform contains one main container and many local containers. The agent manager in the main container can initialize agents to perform the desired task in the local container. When an input or symptoms are received for classification of data, it is placed in the queue, the Agent manager forwards this asynchronously to the data mining agents in all the locations with agent communication language(ACL). Each agent has its own behavior and state of action to perform the desired task and it has its own life cycle.

Pseudo Code for Agent Initialization

Input: Symptoms or attributes k, Distributed Database Locations L

Output: Disease closure prediction Dc

Step 1: enter query with symptoms S as input, the location parameter

**Step 2:** for each queried symptoms task manager agent creates i numbers of task agents. Agent Ai is allocated with DM task DMi.

Step 3: each agent performs its own local data mining task with ACL and generates its own models Mi on the datasets and sends classified mined rule to the main container.

Step 4: establish rule knowledge base; A DMA agent performs its own global data mining (GDM) task with ACL and generates its own models GDM on the classified datasets.

Step 5: stop.

This code initializes agents and performs parallel data mining in the distributed locations and classified models are integrated to obtain the global data mining model. This reduces the processing time, reduces the central storage space and improves the accuracy, reduces the training time.

3.2 DDF-Distributed Data Fetching
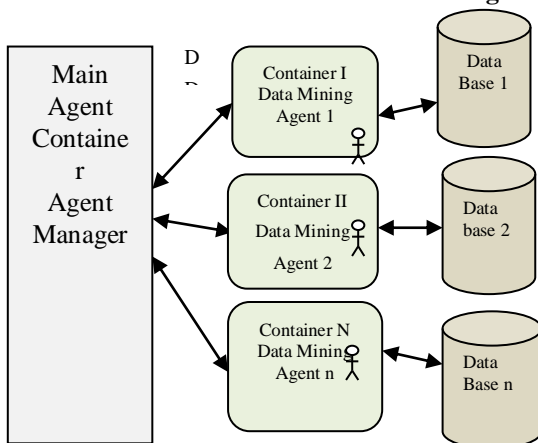
Global Data Mining   Local Data Mining

Figure 3 Block Diagram of Distributed Data Fetching

Preprocessing and Classification of Data for Decision Making. For getting quality data for mining, the retrieved data from distributed sources need to be preprocessed. The Preprocessed data is classified based on the assigned data mining task.

4. Distributed Data Mining Algorithms

Distributed data mining works by analyzing data in a distributed fashion and pays careful attention towards the differences between centralized collection and distributed analysis of data. When the info sets area unit giant scaling up the speed of the info mining task is crucial. Parallel data discovery techniques address this downside by exploitation high performance multicomputer machines. For development of data analysis algorithms that can scale up as we attempt to analyze data sets measured in terabytes and peta bytes on parallel machines with hundreds or thousands of processors. This technology is particularly suitable for applications that typically deal with very large amount of data that cannot be analyzed on traditional machines in acceptable times. Most of the DDM formula designed upon similarity they apply on distributed information. The same formula is applied on totally different web sites manufacturing one native model per site. All native models area unit then mass manufacturing the ultimate model. Each native model represent regionally coherent patterns however lacks the small print required for manufacturing globally meaning

data. The Distributed Data Mining algorithm requires more subset of the nomal data from item to create it & smallest amount data needed for successful DDM algorithm.

4.1 DDM based on Multi Agent System:

Multi Agent System offer architecture for collaborative problem solving in distributed environments. The definion of agents depends on source from different sources. The single Agents in Multi Agent System need to be practical and independent. Agents perceive their environment, dynamically reason out actions based on conditions and interact with each other. Knowledge of MAS is complex and collective, complex in the sense that it is the outcome of data analysis and pre existing domain knowledge. Analysis of information may require information from data mining for identify secreted patterns, construct analytical models and identifying information among others. Collective suggests that every analysis is performed by totally different agents. MAS area unit principally employed in sensing element nodes wherever there's a demand for comparison of knowledge at totally different nodes. Since answer for distributed drawback needs collaboration, semi autonomous behaviour and reasoning, there's an ideal natural process existing between MAS and DDM. Agents are used in this system for the following purposes:

5. INTELLIGENT AGENT DENSITY ESTIMATION BASED FOR LOCATION SELECTION FOR DISTRIBUTED DATA MINING

There are many proposed earlier designs to support decision making and suffers with the problem of accuracy, time complexity and space complexity. The proposed DELS approach is an intelligent technique to choose the set of locations from where the data can be retrieved. The method uses various MD which represents the availability of data in different locations of the network and based on the MD, the method computes the DE measure to choose the most optimal locations by the location selection agent. The number of agent generation is performed according to the DE factor, and based on the patterns generated. The proposed method increases the accuracy by excluding the minimum threshold datasets, and reduces the time, complexity and the space complexity of the system. Figure 4 shows the Density estimation framework.
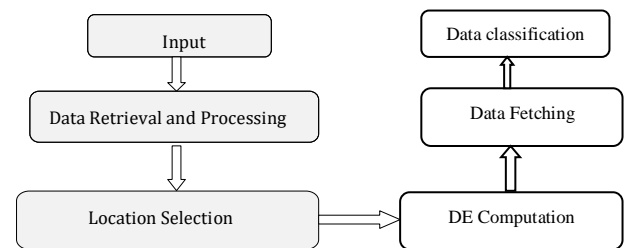
Figure 4 Density Estimation Framework

6. CONCLUSION AND FUTURE WORK

The comparisons of the models is investigate with different datasets. The models is to analyze the intelligent agent for distributed data mining and number of functional components of each stages of the critical making process, accuracy, minimum time to process .In future this method can be enhanced to work with different type of domains and updating the obliging among the services. Improving the efficiency of the model by implementing the agents with a dynamic selection of data mining techniques depending on the situation

REFERENCES

1. Devi, SG.: A Survey on Distributed Data Mining and its Trends, International Journal of Research in Engineering & Technology,2,107-120(2014)

2. Prajapati, RB & Menaria, S.: Multi Agent-Based Distributed Data Mining, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 1,76-81(2012)

3. Liu, X, Huang, J & Zhang, C.: System framework design of automated data mining technology based on intelligent agents, Proceeding of the International Conference on Image Analysis and Signal Processing, 653-655(2011)

4. Rupali Chikhale :Study of Distributed Data Mining Algorithm and Trends IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727 PP 41-47