Journal of Theoretical and Applied Information Technology
31st March 2015. Vol.73 No.3
© 2005 - 2015 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

USING MULTI-AGENTS SYSTEMS IN DISTRIBUTED DATA MINING: A SURVEY

[1]AHMEDAMINE FARIZ, [2]JAAFAR ABOUCHABAKA, [3]NAJAT RAFALIA
LaRIT, Faculty of Sciences, IbnTofail University of Kenitra, Morocco
E-mail: [1]a.amine.fariz@gmail.com, [2]Abouchabaka3@yahoo.fr, [3]arafalia@yahoo.com

ABSTRACT

Extract accurate information from huge databases about petabytes has an unprecedented utility. However, the implementation of a data mining system requires an enormous amount of work and an extremely long run time in some cases of distributed environments. Distributed Artificial Intelligence consists that the entities with certain amount of autonomy must have the capacities of perception and action on their environment, which brings us to the concept of "cooperative agents" and therefore multi-agent systems. These systems are becoming more and more essential in many application fields due to the fact that they can solve the problems of complexity and distribution, especially when it comes to large systems such as data mining. This document is an overview of the integration of multi-agent systems and distributed data mining known as MADM (Multi Agent-based Distributed Data Mining).

Keywords*:* *SMA, DDM, MADM, data mining, agents, Distributed Data mining, Multi-Agent System.*

1. INTRODUCTION

Distributed Data Mining (DDM) is the extraction of knowledge from several databases (Data Mining) regardless of their physical location; it allows the partial analyses of the data extracted from individual distributed sites, and then send the different partial results to other sites to form he final result. The need for such a feature comes from the fact that the data generated locally at each site cannot often be transferred over the network due to the excessive amount of data and privacy issues. Nowadays, DDM becomes increasingly a key element of knowing the systems, because such a decentralized architecture can reach all network-related businesses.
The Multi agent systems (MAS) that deal with complex distributed applications require a distributed resolution of the problems. In many applications, the individual and the collective behaviour of agents depends on the observed data originated from distributed sources. In a typical distributed environment, the analysis of distributed data is a real challenge due to several constraints such as limited bandwidth (wireless networks ...), the sensitivity of confidential data, distributed computing resources. As the MAS are simultaneously distributed systems their combination with the DDM for data-intensive applications is attractive. A number of DDM solutions are available using various techniques like distributed association rules, distributed clustering, Bayesian learning, classification and compression, but only o few of them use intelligent agents.
This paper highlights the possible synergy of multi-agent systems in the distributed data mining to handle their problems.

Other sections of the paper are organized as follows. The 2nd and 3rd section will describe respectively the "Agent" and the "Based on the distributed data mining agents." The 4th section will be devoted to the presentation of "open problems", and in Section V we describe "an overview of the general architecture of the MMDA systems", and finally in Section VI the "open and perceptual issues research”.

2. GENERAL INFORMATION ON DATA MINING

The raw data, despite their quantity that increases exponentially, have almost no value, which is the most important fact, is the knowledge and understanding of the data, but the more we have data the more this process becomes difficult. And

© 2005 - 2015 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

as stated by Shapiro Piatestky "[...] as long as the world keeps Producing data of all kinds [...] at an ever Increasing rate, the demand for data mining continues to grow Will" [1]. Hence Data mining becomes a necessity.

2.1. Definition of Data mining

There are several definitions of DM:

According to P.CABENA and al, The Data Mining is an interdisciplinary field that uses at the same time automatic learning techniques, pattern recognition, statistics of data bases and visualization to identify the ways to extract information from huge data bases [2].

The Data Mining is the analysis of large observational datasets, to discover new relations between them and reformulate these relations to make them more usable by their owners. [3]

2.2. Data mining tasks

The nature of the tasks exercised by Data mining depends on the use of data mining results. Many intellectual, economic or even commercial issues can be expressed as these tasks which are classified as [4]:

1. Exploratory Data Analysis: It is simply exploring the data without any clear ideas of what we are looking for. These techniques are interactive and visual.
2. Descriptive Modelling: It describes all the data, It includes models for overall probability distribution of the data, partitioning of the p-dimensional space into groups and models describing the relationships between the variables.
3. Predictive Modelling: This model permits the value of one variable to be predicted from the known values of other variables.

The first three tasks are examples of Data Mining supervised whose goal is to use the available data to create a model describing a particular variable considered as goal in terms of these data.

4. Discovering Patterns and Rules: It concerns with pattern detection, the aim is spotting fraudulent behaviour by detecting regions of the space defining the different types of transactions where the data points significantly different from the rest.
5. Retrieval by Content: It is finding pattern similar to the pattern of interest in the data set.
   This task is most commonly used for text and image data sets.

2.3. Process steps of Data mining

As discussed in [5] [6] [7] [8] [9], we can define data mining process steps as follow:

1. Data collection: the combination of multiple data sources, often heterogeneous, in a database.
2. Data cleaning (normalization): the elimination of noise (attributes with invalid or no values).
3. Data Selection: Select the useful database attributes for a particular data mining task.
4. Data transformation: the transformation process of attributes structures to be adequate to the information extraction procedure.
5. Extracting information (Data Mining): the application of some Data Mining algorithms on the generated data by the previous step (Knowledge Discovery in Databases, or KDD).
6. Data Visualization: Using visualization techniques (histogram, camembert, tree, 3D visualization) for interactive data exploration (discovery data models).
7. Evaluation of models: identifying strictly interesting models based on data measurements.

2.4. Data Mining techniques

To perform the tasks of Data Mining there are several techniques from different scientific disciplines (statistics, artificial intelligence, databases) to show hidden correlations in data repositories to build models from these data.

Below the most mentioned data mining techniques in different documents:

Neural networks, the decision trees, Genetic algorithms, the association rules, the algorithm of

© 2005 - 2015 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

thek-Nearest neighbours, thek-means algorithm (K-Means), clustering algorithm K-Means

The data mining techniques presented above represent some of the existing techniques for data mining, and the reason why there are so many techniques is that these tasks don't have the same object, and none of them can be optimal in all cases, they complement each other when they are combined intelligently (obtaining very significant performance gains building the so-called meta-models or models).

2.5. Categorization of data mining systems

Data mining systems can be categorized according to several criteria. Among the existing categorizations we note:

1. Classification based on the type of data to be explored: in this classification data mining systems are grouped according to the type of data they handle such as spatial data, time series data, textual data and the World Wide Web, etc.
2. Classification by advanced data models: This classification categorizes data mining systems on the basis of advanced data models such as relational databases, the object-oriented databases, data warehouses, transactional databases, etc.
3. Classification by type of knowledge to discover: this classification categorizes data mining systems based on the type of knowledge to be discovered or data mining tasks such as classification, estimation, prediction, etc.
4. Classification by the exploration techniques used: this classification categorizes data mining systems following the data analysis using the pattern recognition approach, neural networks, genetic algorithms, statistics, visualization, database-oriented or-oriented data warehouse, etc. [5].

2.6. Data Mining application fields

The Data Mining technology has a great significance thanks to the possibilities that it offers to optimize the management of human and material resources.

Banking, Bio-informatics and Biotechnology, Direct Marketing and Fundraising, Fraud detection

The management of scientific data, The insurance sector, Telecommunications, Medicine and Pharmacy, Retail trade, E-commerce and the World Wide Web, The stock market and investment, Analysis of supply chain

2.7. Data mining software solutions

There are many statistical and data mining software products which can sort based on their specialties and techniques below:

Neural Networks: (Start miner edited by Grimmer soft, Predict edited by Neural Ware, Neuro One, edited by NETRAL, 4Thought, edited by Cognos)

Decision Tree: (Alice edited by Isoft, Know Knowledge SEEKET edited by Angoss, CART edited by Salford systems, Microsoft Analysis Services edited by Microsoft)

Associations: (Wizwhy edited by Wizsoft)

We can also sort data mining software products based on their architectures below:

Single-user software solutions:

-WEKA (Waikato Environment for Knowledge Analysis), an open source project that contains several data mining techniques, derived from the automatic learning community.
- TANAGRA, a free experimentation platform for teaching and research

Client/Server software solutions:

- Clementine SPSS (Statistical Package for the Social Sciences), used to forecast market share, fraud detection, market segmentation and implementation of outlets
- Enterprise Miner SAS used for cost reduction, retention and prospecting
-IBM Intelligent Miner
-Data Miner Statisca

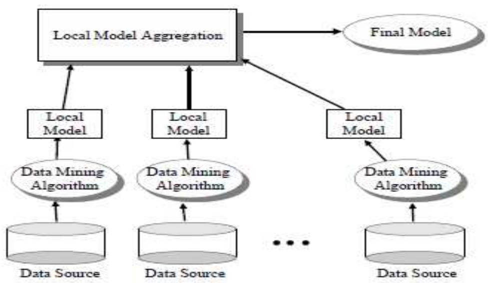
Journal of Theoretical and Applied Information Technology
31st March 2015. Vol.73 No.3
© 2005 - 2015 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3. DISTRIBUTED DATA MINING

DDM is derived from data mining field that focuses on the distribution of resources on the network. In a distributed scenario where the DDM is applied, the system consists of several independent sites, these sites of data and calculation communicate by sending messages, and this communication is often expensive in terms of power autonomy, privacy and security.

3.1. - Architecture and Features of a DDM system

DDM is planned to perform partial analysis of data at individual sites and then send the partial results to other sites where it may need to be aggregated to the overall result. Thus individual nodes communicate with a rich centralized node resources and neighbouring nodes via messages on an asynchronous network to accomplish their tasks. Hence, the majority of DDM methods manage more abstract architecture which includes several Sites with an important storage capacity and independent computing power. The local computation is performed on each site and a central site that communicates with each location distributed to calculate global models using a peer-to-peer architecture.



Two hypotheses are usually adopted on how data is distributed on sites:

- Homogeneous distribution (horizontally partitioned)
  - the global table is divided horizontally
  - The tables in each site are subsets of the global table that have exactly the same attributes as them.
- Heterogeneous distribution (vertically partitioned)
  - The table is divided vertically.
  - Each site contains a collection of columns
  - The sites do not have the same characteristics.

It is important to accentuate that the perspective of the global table is strictly conceptual.

3.2. DDM methods

Several systems have been developed for distributed data mining. These systems can be classified according to their strategy in three types, Central Learning, Meta-Learning and Hybrid Learning.

3.2.1. Central Learning Strategy

Data can be collected at a central site, and therefore one model can be made. The only requirement is to be able to move data to a central location to merge and then apply sequential DM algorithms. This strategy is used when the distributed data are geographically limited; it is usually very expensive but also more accurate. In general, the process of data collection is not just a melting step, it depends on the original distribution. For example, the individual logs are placed in different locations, different attributes of the same records are distributed over different locations or different tables can be placed at different locations, so during the data collection, it is necessary to adopt a proper fusion strategy. However, this strategy is generally impractical [10].

3.2.2. Meta-Learning Strategy

It provides a way to exploit the classifiers from distributed data homogeneously. Meta-Learning follows three main stages. The first is to generate classifiers based on each site using the classifiers learning algorithms. The second step is to collect the classifiers based on a central site, and produce meta-data levels from a separate set of validation and the predictions generated by the base classifier. The third step is to generate the final classifier (meta-classifier) from the meta-data level via a combiner or an arbitrator. Copies of classification agent will exist or will be deployed on the nodes of the network used. The more mature agent systems based on Meta-

Journal of Theoretical and Applied Information Technology
31st March 2015. Vol.73 No.3
© 2005 - 2015 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Learning systems are: JAM system and BODHI [11].

3.2.3. Hybrid Learning Strategy

Is a technique that combines local learning and centralized for building models [12]. For example, Papyrus [13] is designed to support learning strategies. Unlike Jam BODHI Papyrus can not only move the site models site, but can also move the data when this strategy is desired? Papyrus is a specialized system that is designed for clusters while JAM BODHI and are designed for data classification.

3.3. Issues and Challenges of DDM

The main criticism of these systems is that it is not always possible to obtain an accurate final result, knowing that the global knowledge model obtained may be different from the one obtained by applying a model approach (if possible) to the same data.

The approximate results are not always a major concern, but it is important to be aware of this. In addition, the use of material resources of these systems is not optimized. In a distributed environment analysing distributed data is a significant problem because of many constraints such as:

3.3.1. Communication:

- The Limit of the overall performance of the system, since it is supposed to be carried out by message passing
- The communication between sites is expensive and it is performed only by message
- The cost of transferring large blocks of data can be prohibitive and lead to very inefficient implementations.
- The limited bandwidth (eg wireless networks)
- The main objective of a many research and DDM methods is to minimize the number of sent messages).

3.3.2. Resources and distributed computing nodes.

-The system consists of several data sites and independent calculation.
- The sites have resource constraints
• (eg the battery power) => some methods are also trying to balance the load between sites to prevent the performance to be dominated by time and space utilization of an individual site.

3.3.3. Confidentiality

-These issues of privacy and autonomy become especially important in business application scenarios where, for example, different companies (often contradictory) may want to collaborate for the detection of fraud, but not to share data of their individual customers or to disclose it to third parties.

- For example, when the data can be found on the data warehouse from several different perspectives and at different levels of abstraction, it can threaten the objective to protect personal data and guard against the invasion of life private.

3.3.4. Scalability:

The core of the DDM systems is scalability, where system configuration can be changed from time to time, in which scalability is a critical issue in a distributed system. To inform each unit in the system on the configuration of the system update, such as a new data site is adhered to the system requires additional human intervention or a complex mechanism where a decrease in performance can occur.

3.4. Solutions :

These and some other peculiarities require the development of new approaches and technologies of data mining to identify patterns in distributed data. Distributed data mining (DDM), in particular, Peer- to-Peer (P2P) data mining, and multi-agent technology are two responses to the above challenges. [14]
For these reasons, the characteristics of the agents are important for DDM systems.

3.5. Application fields of DDM

The distributed Data Mining applications include detecting fraud credit card system, intrusion detection system, health insurance, safety-related applications, the distributed clustering, segmentation, sensor networks the customer profiling, assessment of retail promotions, analysis of credit risk, etc.

Journal of Theoretical and Applied Information Technology
31st March 2015. Vol.73 No.3
© 2005 - 2015 JATIT & LLS. All rights reserved.
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4. MULTI-AGENT SYSTEMS:

Before starting the contribution of multi-agents in the field of data mining systems, it would be interesting to understand a few concepts like, distributed artificial intelligence, the evolution of the individual aspect (the behaviour of a single agent) towards the collective aspect (behaviour of agents in a society).

4.1. Distributed Artificial Intelligence

Artificial intelligence is a science dedicated to solving the problems that cannot be solved by traditional computing; it brings together many sciences such as computer science, psychology, and philosophy to produce machines (computers) that can be described as intelligent. Over time, the Artificial Intelligence (AI) Classic showed is limited when it comes to solving complex problems. In order to address these limitations the researchers felt the need to move from individual behaviour to collective behaviour and the need to distribute intelligence across multiple entities. All studies covering collective behaviour constitute the field of Distributed Artificial Intelligence (DAI). Technically, the IAD is a branch of AI, which proposes to replace the centralized software bysoftware based on the interaction of basic software components. It is based on the principle of 'divide and rule' that facilitates the development and test of problem-solving systems and allow better reusability of components, and it is structured around three axes:

1. Parallel Artificial Intelligence (IPA), develops languages and parallel algorithms designed to improve the performance of AI systems.
2. Distributed Resolution of Problems (RDP), focuses on how to divide a problem into a distributed set of entities and how to share the knowledge of a problem in order to obtain the solution.
3. Multi-Agent Systems (MAS), which favour a decentralized approach to model and focus on the collective aspects of the systems. [15]

4.2. Agent concept :

From a purely IT perspective, an agent can be defined as an object (as defined object languages) whose behaviour is described by a "script" (principal function main), with its own calculation ways, and can move from place to place (a place that can be a remote computer site from the original site of the agent) to communicate with other agents.

With its "script", the agent is able to follow a life behaviour that will be instilled at the time of implementation and that will allow him to have as main feature to be fully autonomous.

One of the discriminate characteristics of the agents is the representation and reasoning on the environment (the external world and other agents), based on this feature, we find two different classes.

4.2.1. Cognitive agents

A cognitive agent is an agent that has an explicit representation of its purpose and its environment. The actions it performs to achieve its goal are the result from a reasoning on the state of the environment. Usually a cognitive system includes a small number of agents; each is similar to a more or less complex expert system. In this case we speak of high granularity agent.

4.2.2. Reactive agents

A reactive agent is an agent whose behaviour responds only to the stimulus/ share law, the stimulus is an element of the environment (action, message, location, etc.). Typically a reactive system has a large number of low granularity agents. These agents do not necessarily have an explicit goal to obtain. By cons, they can implement a complex reasoning on their internal state to perform their actions. [16]

Journal of Theoretical and Applied Information Technology
31st March 2015. Vol.73 No.3
© 2005 - 2015 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.3. Mobile Agent

A mobile agent is a running program that can move from host to host in a network which created a new paradigm for data exchange and resource sharing in rapidly growing and continually changing computer network.

It is capable of migrating autonomously and intelligently in various target nodes through network to perform computation in response to changing conditions in the network environment. The agent's dispatcher purposes are to achieve and fulfil user's objective on behalf of user. It is used for information retrieval, searching information, filtering, and intrusion recognition in networks. Mobile agent suspends its execution, Transport itself from one host connected to the network to another, and continue its execution on the new host.. [17] [biblio2]

4.4. Multi-Agents system

4.4.1. Definition

- A multi-agent system is a system with a set of agents that interact with the communication protocols that are capable of acting on their surroundings. Different agents have different spheres of influence, because they control (or at least can influence) on different parts of the environment. These spheres of influence may overlap in some cases; the fact that they coincide may cause dependencies reports between agents.

The MAS can be used in many application areas such as electronic commerce, economic systems, distributed information systems, organizations. [18] [biblio3]

4.4.2. When using a MAS?

When the problem is too complex but can be divided.

- When there is no general solution or when it is too expensive in CPU.
- For modelling purposes (populations, molecular structures, sand piles ...)
- When we can parallelize the problem (saving time).
- When we want certain robustness (redundancy).
- When the expert comes from different sources.
- When the data, controls, resources are distributed.
- When the system needs to be adaptive. [19]

4.4.3. Conclusion

The combination of multi-agent systems with diverse applications in various fields was a natural result thanks to the advantage they give in soling complex problems based on the basic principles: collaboration between agents and parallelism. Our project is a case like this; it is interested in the contribution of SMAs in data mining systems which were the cause to devote this chapter to determine the possibilities of this contribution.

5. MULTI-AGENT SYSTEMS BASED ON DISTRIBUTED DATA MINING

The technology of multi-agent systems has generated much excitement in recent years thanks to the promises that it gives committing a new design paradigm and implementation of software systems. These promises are particularly attractive for the creation of software running in distributed environments. Multi-agent systems (MAS) often treat complex applications requiring a distributed resolution of problems. In many applications, the individual and collective behaviour of agents depends on observed data from distributed sources.

Since SMA is also distributed systems, the combination of DDM with MAS for data-intensive applications is attractive.

5.1. Why combine the DDM and the MAS?

We can identify the following arguments for and against their use.

5.1.1. Autonomy of data sources:

DM agent can be considered as a modular extension of a data management system to deliberatively manage the access to the source data underlying according to the constraints specified on the required autonomy of the system.

An autonomous agent can be treated as a computing unit that performs multiple tasks based on dynamic configuration. The agent interprets the configuration and generates an implementation plan to multitask. [20], [21], [22], [23], [24] and [25] discuss the benefits of deploying agents in the DDM systems.

This is in full compliance with the paradigm of cooperative information systems.

5.1.2. Interactive DDM:

© 2005 - 2015 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Proactive auxiliaries agents can significantly reduce the number of human users who need to monitor and interfere with the Data mining process running, for example, DM agents can anticipate potential individual limits of the large search space and appropriate intermediate results.

5.1.3. Dynamic selection of sources and data collection:

In the multi-source open environments the DM agents can be applied to select adaptively data sources, according to the criteria given: such as the expected amount, type and quality of data expected in the reporting source, real network and the load of the DM server. [40]

5.1.4. DM Scalability of distributed massive data:

A set of DM agents allows to make an approach of "divide and rule" for the local execution of operational tasks to each of the data locations. DM Agents aggregate pre-selected data related to their origin server for further processing and they can evaluate the best strategy between remote working or the migration of data sources. Experiments using mobile agents for information filtering in distributed data environments are encouraging [26].

To the concern of the scalability, collaborative learning agents [26] [27]are able to share information, in this case, on changes in system configuration, and spread from one agent to another, in order to adapt the system to accurate the individual agent. In addition, mobile agents As discussed above can help to reduce the network load and the DM application server as in the state of the art [28][29]systems.

5.1.5. Multi-strategy DDM:

Agents DM can learn in due time of their deliberative action by combining several DM techniques to choose according to the pursued exploratory tasks and the type of data taken from different sites. The learning of the multi-strategy selection DM methods is similar to the adaptive selection of coordination strategies in multi-agent system as proposed in, for example, [30] .

5.1.6. Security

Any failure to implement the least privilege to a data source could cause unwanted access to the sensitive data of Mining-Agent. The agent code and data integrity is a crucial issue in the secure DDM. Reversing it or diverting DM agent gives more reliability to the software (mobile), therefore any sensitive data made or transmitted by the agent is under the control of an intruder.

If the DM agents are even allowed to migrate the methods of computing environments remotely, to ensure authentication and confidentiality of a mobile agent, thus safety precautions should be applied.

The agents can be programmed in a manner where light agents, can be transmitted through the network instead of the data that can be larger. Be capable of transmitting agents from one to the other host allows a dynamic organization of the system. As in [33], the authors present a framework in which mobile agents move in the network system by allowing it to maintain the confidentiality of data.

Finally, thanks to the characteristics of reproduction and agent self-organizing which release the system to transfer the data across the network, and can also help to prevent malicious hosts to block or destroy the temporarily residents DM agents, and increases data security.[31][32]

5.1.7. Reliability:

DM agents may deduce sensitive information even from partial integration to some degree and with some probability. This problem, known as the so-called inference problem, particularly happens in brackets where officers can access to data sources across confidence borders which enable them to integrate implicit knowledge from different sources. The inference problem is always studied in research as an independent thread of existing DDM systems based on agent or not.

5.2. The most important agents for MDD?

Considering the most important and representative agent based DDM based systems to date: BODHI, Padma, JAM, and Papyrus (details in [34] [20]).

Journal of Theoretical and Applied Information Technology
31st March 2015. Vol.73 No.3
© 2005 - 2015 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5.3. MADDM system

The chapter highlights the possible synergy between the MAS and DDM. Generally the construction of an ADDM system involves three essential features: interoperability, dynamic system configuration and performance aspects, discussed as follows.

5.3.1. Interoperability

The concerns about interoperability are not only the cooperation of agents in the system, but also the external interaction that allows new agents to enter into the system seamlessly. The system architecture must be open and flexible in order to support interaction, including communication protocol, integration policy, and the service directory. The communication protocol includes the coding of messages, encryption and transport between agents, however, they are normalized by the Foundation for Intelligent Physical Agents (FIPA) 1 and are publicly available. The majority of agent platforms as Jade2 and Jack3 are FIPA and are possible. The integration policy specifies how a system behaves when an external component, such as an agent or data site required to enter or to exit.

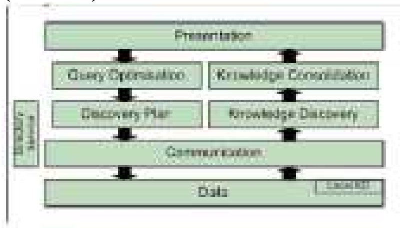
5.3.2. Dynamic System Configuration

The issue is further discussed in relation to the characteristic of interoperability, which tends to manage the dynamic system configuration; it is a matter of challenge because of the complexity of planning and exploration algorithms. a Mining group can have multiple data sources and agents which are configured to equip an algorithm and process the provided data sets. The change in the data affects the Mining task, as an agent that may still be running the algorithm.

5.3.3. Performance

Finally, the performance can be improved or altered because the distribution of data is a major constraint. In a distributed environment, tasks can be executed in parallel and in exchange the competition issues arise. The performance of service quality control for DM perspectives and data system is desired, but it may be derived from both data mining and field agents.

5.4. Agent of distributed data mining (ADDM)



We can generalize the system activities in requests and responses, where which one has a different set of components. The basic elements of ADDM system areas are as follows [40]:

5.4.1. DATA:

This layer is the founder of our interest. In a distributed environment, the data can be housed in various forms, such as relational data bases online data stream, web pages, etc.in which the data objects varied.

5.4.2. Communication:

The system selects the related resources from the service directory that maintains a list of data sources, the extraction algorithms, data schemas and data types, etc. The communication protocols may vary depending on the implementation of the system, such as client-server and peer-to-peer, etc.

5.4.3. Presentation: (UI)

The user interface (UI) interacts with the user to receive and respond, it simplifies the complex distributed systems in a friendly message as network diagrams, visual reporting tools, etc.
On the other hand, when a user requests for data retrieval via the user interface, the following components are involved:

5.4.4. Query optimization :

A query optimizer analyses the request to determine the type of exploration tasks and selects the appropriate resources for the application. It also determines the ability to parallelize tasks, since the data is distributed and can be operated in parallel.

5.4.5. Discovery Plan :

A scheduler allocates sub-tasks with related resources. At this stage, the mediator agents play an important role in coordinating multiple calculation units, since mining sub-tasks are performed asynchronously, and the results of these tasks. Furthermore, when a task extraction is complete, the following items are taken up.

Journal of Theoretical and Applied Information Technology
31st March 2015. Vol.73 No.3
© 2005 - 2015 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5.4.6. Local Knowledge Discovery (KD) :
In order to transform the data into patterns which adequately represent the data that can be transferred over the network, on each data site, the mining process may take place locally in accordance with the individual implementation.

5.4.7. Knowledge Discovery :
Also known as mining, it runs the algorithm as required by the obtaining knowledge task from the specified data source.

5.4.8. Knowledge Consolidation:
To present a compact and efficient operating income to the user, it is necessary to standardize the knowledge obtained from various sources. The component involves complex methodologies to combine knowledge/schemas from distributed sites. Strengthening homogeneous knowledge/schemas is promising but it still difficult for the heterogeneous case.

5.5. Architecture of MADDM
Most MADM Frameworks adapt to similar architectures (see Fig .2.) And provide common structural components [35], [36], [37]. They use KQML or FIPA-ALC, which are standard agent communication languages that facilitate interactions between agents.

Each agent in multi-agent system generally contains interface module, process module and knowledge module. The interface module is responsible for communicating with other agent or with the environment. The knowledge module provides necessary knowledge to be proactive or reactive in various scenarios. The process module does necessary processing to make decisions. Different types of agents in multi-agent based spatiotemporal data mining system are Interface agent, Facilitator agent, Broker agent, Data agent, Pre-processing agent, Data mining agent, Result validation agent. These agents are briefly described below.

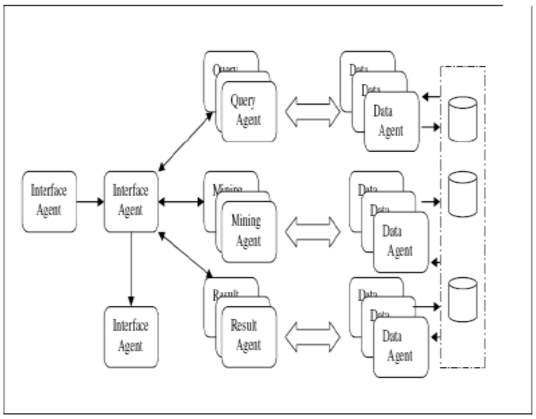


Figure .2

The following is a definition for the most common agents used in the MADM, the names may be different, but they share the same features in most cases. [38]

5.5.1. Agent Interface :
This agent communicates to user. It accepts user requirements and provides data mining results to him. The interface module is responsible for getting input from the user as well as inter-agent communication. Methods in the process module capture user input and communicate it to the facilitator agent. The knowledge module manages history of user's interaction and their profiles [38].

5.5.2. Agent animator (or Agent Manager):
This agent takes the responsibility of activation and synchronization of various agents. It receives a request from the interface agent and produce a work plan. It elaborates various tasks to be completed in the work plan and ensures the work plan is completed. It communicates the results to the interface agent [38]. [biblio4]

5.5.3. Agent des ressources (data agent):
Actively maintains the metadata information on each of the data sources. It also provides "ad hoc" predefined extraction capabilities. It is responsible for collecting the necessary data sets required by the data mining agent for a specific data extraction operation. It takes into account the heterogeneity of databases, and the resolution of conflicts in the definition and representation of data. Its inter face module supports inter-agent communication as well as interface to existing

© 2005 - 2015 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

data sources. The processing module provides facilities for ad hoc and recovery (Extraction) predefined data. Based on user demand, appropriate queries are generated and executed against the data base and the results are communicated to the Host or other agents.

5.5.4. Mining Agent:
It implements specific data mining techniques and algorithms. The interface module support sinter-agent communication. The process module contains methods to launch and implement data mining activities, capturing the results of the data mining, and subsequently communicate the result of agent or facilitate or agent. The knowledge module contains meta-knowledge on data mining methods, i.e., what is the appropriate method for this type of problem, the entry requirements for each of the extraction methods, format of input data, etc. This knowledge is used by the processing module in the initiation and execution of a particular algorithm DM for the problem at hand.

5.5.5. Result Agent :
Result Validation Agent: This agent gets the results from the spatiotemporal data mining agents. It performs the validation operations on the data mining results. The agent is able to process the results to fulfil various presentation and visual representation software. It maintains details about visualization primitives and report templates that are used to present the results.

5.5.6. Broker Agent (or Match maker):
Broker Agent: This agent maintains names, ontology and capabilities of all the agents which are registered with it to become a part of the multi-agent based data mining system. The broker agent receives a request from the facilitator agent and responds with names of appropriate agents which have the capabilities requested.

5.5.7. Query Agent:
Request agent is generated for each user request. The knowledge module contains metadata information, including local schemas and a global schema. These schemas are used to generate the queries for retrieving data.

5.5.8. Ontology Agent :
The ontology agent maintains and provides comprehensive knowledge ontologies and answers questions on ontologies. It can simply store the ontology as given, or it can be so advanced can use the semantic reasoning to determine the applicability of a domain to any request for special data mining (DM).

5.5.9. Mobile agent :
Some systems use the agent mobility feature. A mobile agent travels around the network. On each site, it processes the data and sends the results back to the main host, instead of expensive transferring large amount of data across the network. This has the advantage of low network traffic because the agents do data processing locally. However, it provokes a major security issues. As an organization receiving a mobile agent for execution at your local machine require strong assurances about the agent's attentions. There is also the requirement of installing agent platform at each site.

5.5.10. Local Task Agent:
In the majority of systems the data agent is a local agent located on the local site. It can submit information to the facilitator agent and also answer data mining requests mining agents. A local agent can recover its local database, and perform calculations and returns the results to the system.

5.5.11. KDD system agents:
Certain MADM systems contain other agents to maintain all the knowledge discovery process on data that includes preparation and evolution of data. These agents are:

5.5.11.1.Pre-processing Agent:
It prepares the data for mining. It is responsible for performing the necessary data cleansing before using the data set, data mining. The processing module contains methods for data cleaning and preparation of the data needed for specific data mining algorithms

5.5.11.2.Post data mining Agent:
It evaluates the performance and the accuracy, etc., data mining agents.

© 2005 - 2015 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

6. DM TASK PLANNING AND MOVEMENT OF SYSTEM OPERATIONS:

DM task scheduling is performed through negotiations between the facilitator agent and mining agents via message passing mechanism.
Suppose a user agent sends a request to the facilitator agent to inform that it would like to do data mining with other agents in the organization. The user agent also needs to give information of model definition (dependent and independent attributes, attribute type (numeric or categorical), model type (linear or nonlinear) with its request. When the facilitator receives the request from the user agent, it negotiates with the broker agent to determine which agents to launch for this task. For example, if the user wants all possible rules meeting minimum support and confidence levels, across all available data sources, then the mining agent must ask every data agent possible for all association rules. If the user wants to find all items Y which have statistical significance for a given X, the mining agent must ask only the agents that have information about X. It would be time consuming and wasteful to ask the agents that have access to data that doesn't contain X, therefore no rules would be generated .Finally, if the user specifies an X and a Y and asks for the level of support and confidence between the two, the agent must only ask the agents that have information about both X and Y. The mining agent is then responsible for completing the task, while the facilitator agent continues to plan future DM requests. When the mining agent completes its work it returns the results and the facilitator agent passes them onto the user agent. [39]

7. CONCLUSION AND PERSPECTIVE

Multi-agent systems and data mining are among other domains that are the most active areas in the field of information technology. Ongoing research has shown a number of challenges and inherent limitations faced by each area. However, the synergy between the two technologies offers great potential and opportunities for more sophisticated applications.

The growing interest in this synergy allows the "Mining Agent" to become a new field of research for the current studies.

In this paper, we gave an overview regarding: Driving forces, theoretical foundations, major research issues and areas of application of this combination, taking into account the state of the art research development Data mining and multi-agent. Another perspective is that we are currently studying the definition of a new process of MADM system based on game theory, which is the formal study of decision making which several players (agents in our study) must make choices that may affect the interests of other players. Which promises to achieve high quality performances of DDM systems using any domain previously cited application.

REFERENCES:

[1] G. Piatetsky-Shapiro, "Data Mining and Knowledge Discovery 1996 to 2005", 15(1), 99-105.
[2] P. Cabena, P. Hadjinian, R. Stadler, J. Verhees and A. Zanasi, "Discovering Data Mining: From Concept to Implementation", Prentice Hall, Upper Saddle River, Nj, 1998.
[3] D. Hand, H. Mannila and P. Smyth, "Principles of Data Mining", Mit Press, Cambridge, Ma, 2001.
[4] S. P. Deshpande1 And V. M. Thakare International Journal of Distributed and Parallel Systems (IJDPS) Vol.1, No.1, September 2010
[5] M. J. Berry, G. S. Linoff, "Data Mining Techniques for Marketing, Sales, and Customer Relationship, Management", Second Edition, 2004.
[6] O. R. Zaïane, "Principles of Knowledge Discovery in Databases", Cmput690, University Of Alberta, 1999.
[7] Ph. Preux, « Fouille De Données : Notes De Cours », Université De Lille 3, 9 Octobre 2008.
[8] G. Dong, J. Pei, 'Sequence Data Mining', Springer Edition, 2007.
[9] S. Prabhu, N. Venkatesan, "Data Mining And Warehousing", New Age International (P) Ltd., Publishers, New Delhi, 2007.
[10] J. Dasilva, C. Giannella, R. Bhargava, H. Kargupta, and M. Klusch. "Distributed Data Mining and Agents. Engineering Applications of Artificial Intelligence", 18(7):791–807, October 2005.
[11] S. Datta, K. Bhaduri, C. Giannella, R. Wolff, and H. Kargupta. "Distributed Data Mining In Peer-To-Peer Networks'. Internet Computing, IEEE, 10(4):18–26, 2006.

© 2005 - 2015 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[12] W. Davies and P. Edwards. "Distributed Learning: An Agent-Based Approach To Data-Mining". In Proceedings Of Machine Learning 95 Workshops On Agents That Learn From Other Agents, 1995.

[13] U. Fayyad, R. Uthurusamy, and Others. "Data Mining and Knowledge Discovery In Databases". Communications of the ACM, 39(11):24–26, 1996.

[14] Vuda Srinivasa Rao Et. Al. / International Journal Of Engineering Science And Technology Vol. 2(6), 2010, 1758-1766

[15] K. P. Sycara, Multi agent Systems, American Association for Artificial Intelligence, 1998.

[16] J. Ferber, « Les Systèmes Multi Agents: Vers Une Intelligence Collective »,1995.

[17] S. Sathia kumar Et Al. (Eds.), "Proceedings Of International Conference On Internet Computing And Information Communications, Advances In Intelligent Systems And Computing" 216, Doi: 10.1007/978-81-322-1299-7_39, _ Springer India 2014

[18] S. Maalal; M. addou (IJACSA) International Journal Of Advanced Computer Science And Applications, Vol. 2, No. 11, 2011

[19] F. Sandakly, « Contribution À La Mise En Œuvre D'une Architecture À Base De Connaissances Pour L'interprétation De Scènes » 2d Et 3d, Thèse Doctorat, Université De Nice-Sophia Antipolis, 1995.

[20] J. P. Briot Et Y. Demazeau, « Introduction Aux Agents », 2001.

[21] Sung W. Baik, Jerzy W. Bala, And Ju S. Cho. "Agent Based Distributed Data Mining. Lecture Notes in Computer Science", 3320:42–45, 2004.

[22] C. Giannella, R. Bhargava, and H. Kargupta. "Multi-Agent Systems And Distributed Data Mining". Lecture Notes in Computer Science, Pages 1–15, 2004.

[23] Vladimir Gorodetsky, Oleg Karsaev, and Vladimir Samoilov. "Multi-Agent Technology for Distributed Data Mining and Classification". In IAT, Pages 438–441. IEEE Computer Society, 2003.

[24] Sven A. Brueckner H. Van Dyke Parunak. "Engineering Swarming Systems. Methodologies and Software Engineering For Agent Systems", Pages 341–376, 2004.

[25] M. Klusch, S. Lodi, and G. Moro. "The Role of Agents in Distributed Data Mining: Issues and Benefits". In Proceedings Of The 2003 Ieee/Wic International Conference On Intelligent Agent technology (IAT 2003), Pages 211–217, 2003.

[26] Matthias Klusch, Stefano Lodi, and Gianluca Moro. "Issues of Agent-Based Distributed Data Mining. In Proceedings of the Second International Joint Conference on Autonomous Agents and Multi agent Systems" (Aamas-03), Pages 1034–1035. ACM, 2003.

[27] A. Bordetsky. "Agent-Based Support for Collaborative Data Mining In Systems Management. In Proceedings of the Annual Hawaii International Conference on System Sciences", Page 68, 2001.

[28] Jan Tozicka, Michal Jakob, and Michal Pechoucek. Market-Inspired approach to Collaborative Learning. Lecture Notes In Computer Science, 4149:213–227, 2006.

[29] S. Bailey, R. Grossman, H. Siva Kumar, and A. Turinsky. Papyrus: A System For Data Mining Over Local And Wide Area Clusters And Super-Clusters. In Supercomputing '99: Proceedings of the 1999 ACM/IEEE Conference on Supercomputing (CDROM), Page 63, New York, Ny, Usa, 1999. ACM.

[30] Hillol kargupta, Byung-Hoon Park, Daryl Hershberger, and Erik Johnson. Collective Data Mining: A New Perspective toward Distributed Data Mining. In Advances in Distributed and Parallel knowledge discovery, Pages 133–184, 1999.

[31] F. Bergenti, M. P. Gleizes, and F. Zambonelli. Methodologies and Software Engineering For Agent Systems: The Agent oriented Software Engineering Handbook. Kluwer Academic publishers, 2004.

[32] V. Gorodetsky and I. Kotenko. The Multi agent Systems for Computer Network Security Assurance: Frameworks and Case Studies. In IEEE International Conference On Artificial Intelligence Systems, 2002, Pages 297–302, 2002.

[33] Frank E. Walter, Stefano Battiston, and Frank Schweitzer. A Model of a Trust-Based Recommendation System On A Social Network, Nov 2006.

[34] Xining Li and Jingbo Ni. Deploying Mobile Agents in Distributed Data Mining. Lecture Notes in Computer Science, 4819:322–331, 2007.

[35] M. Klusch, S. Lodi, and G. Moro. "Agent Based Distributed Data Mining: The KDEC Scheme". Agent link, Springer Lecture Notes In Computer Science, 2586, 2003.

© 2005 - 2015 JATIT & LLS. All rights reserved·
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[36] H.Kargupta, I. Hamza oglu and B. Stafford "Scalable, Distributed Data Mining An Agent Based Application". Proceedings Of Knowledge Discovery And Data Mining, August, 1997.

[37] Kargupta, H., Hamzaoglu, I., Stafford, B.: "Scalable, Distributed Data Mining Using an Agent-Based Architecture". In D. Heckerman, H. Mannila, D.Pregibon, R.Uthurusamy, Eds.: Proc. 3rd International Conference On Knowledge Discovery And Data Mining, Newport Beach, California, Usa, Aaai Press (1997) 211–214

[38] I.L. Narasimha Rao1, A. Govardhan2, And K. Venkates wararao department Of Cse, Aurora's Technological And Research Institute, Uppal, Hyderabad-98school Of It, JNTU, Hyderabad department Of Cse, Cvr College Of Engineering, Ibrahim patnam, Rr Dt., A.P, India, Springer International Publishing Switzerland 2014

[39] V.Sreenivasa. Rao: "Multi Agent-Based Distributed Data Mining: An Over View". JNT University, Hyderabad. Andhra Pradesh, India, 2010

[40] Triloknath pandey, Niranjan Panda And Pravat Kumar Sahu : "Improving Performance Of Distributed Data Mining (DDM) With Multi-Agent System" IJCSI International Journal Of Computer Science Issues, Vol. 9, Issue 2, No 3, March 2012