Study of agent and its type in distributed environment

Saju Mathew

Department of IT (Information Technology), Al Khawarizmi International College, Abu Dhabi, U.A.E

Abstract: Data mining (DM) is a process of extraction and identifying of non-trivial, implicit information or patterns from data in a large database. Data mining is usually focuses on the process of knowledge discovery (KD). Data mining technology normally adopts data integration method to generate data warehouse, where the data is gathered into a central site and then runs various algorithms to extract the data and make useful evaluation. Distributed data mining (DDM) is an active and growing area since it is very widely used in all the applications such as industrial, scientific etc, to analyze the large data sets which are geographically distributed. Due to massive size of data warehouse and abundance of database and unstructured data sets, the most critical factor affecting the organization is in attempting the control over their data sets. Current DDM techniques regard the distributed data sets as a single valued table if the data's are combined or centralized. The agent and agent mining is a new area of research which is driven by challenges faced by both data mining and distributed data mining and need to develop more advance intelligence information processing and system.
This paper focuses on an overview study of data mining, distributed data mining, challenges faced by them, agents their various dimensions, its types and application.

Keywords: Data Mining, Knowledge discovery, Distributed Data Mining, Agent, Agent Mining.

Introduction - Data Mining

In recent years there have been tremendous development in the growth of technology, increase in size of data and storage capacity, high computational power, which has totally changed and automated all the business, manufacturing, production and all the scientific research processes [18].
Data mining algorithm deals with simple data formats typically flat files with an increasing amount of focus in mining complex and advance data types. One of the aspects of data mining systems is to move from stand alone system using centralized and local computational resources, supporting increasing levels of distribution. In today's informative age, the information is globally distributed and organization has to perform required data mining on distributed data resources either homogenous or heterogeneous to obtain a cohesive and integrated knowledge from this data sets. The conventional technology that is used by the centralized data mining is no longer suitable and it faces many challenges [1].

The main challenges are;

1. Data mining has to deal with huge amount of data located at different sites.
2. Data mining has to process, partition and distribute the data for parallel processing to achieve acceptable time and space performance.
3. Many times data can change rapidly in such cases the knowledge has to be mined fast and efficiently in order to be useable and updated.
4. Data mining lacks security.

Since the data mining environment consists of user, data, hardware and mining software, and have a centralized data repository such an approach is inappropriate for distributed data mining. Due to long response time, lack of proper use of distributed resources and the fundamental characteristic of centralized data mining algorithms do not work well in distributed environment.

Distributed Data Mining

An increase in the number of database and data streams which are made online and changing frequently needs to be monitored regularly. The distributed approach to analyze this data is more scalable and practical when the application involves a large number of data sites.

Distributed data mining (DDM) is a branch of data mining that offers a framework to mine distributed data and computed the distributed resources. Its primary objective to distribute data mining to address the specific issues associated with the application of data mining in distributed computing environment. The following factors which make the distributed data mining significant from centralized mining are as follows [3];

    a. The need to mine distributed subsets of data, the integration of which is non-trivial and expensive.
    b. The performance and scalability bottlenecks of data mining.
    c. Distributed data mining provides a framework for scalability which allows the splitting up of large datasets into smaller datasets.
    d. It focuses on data mining process and the resources that are distributed over network so that it can save time and resources.

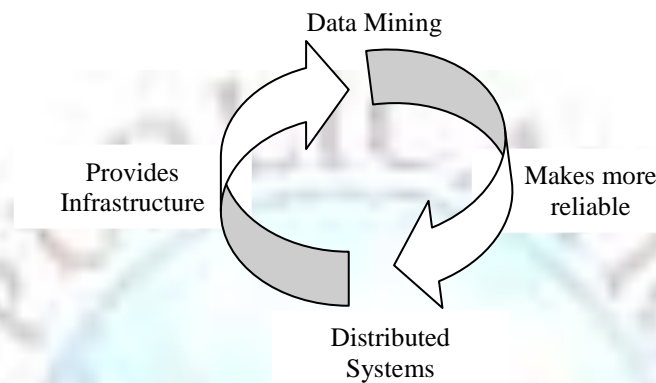The diagram shows data mining and distributed systems [4].

Fig 1: Data mining and distributed systems

The main part of distributed data mining system is the scalability, as the system configuration may be altered time to time.

The challenges faced by distributed data mining are;

    a. Data is distributed over many heterogeneous sources either in tight or loose manner.
    b. It is difficult to conduct the data integration and matching.
    c. It is not possible to process the data in centralized manner.
    d. Due to privacy local data can be restricted.
    e. Distributed data spread across global storage systems is often associated with time difference.
    f. Lack of infrastructure and architecture in the existing distributed data mining system.

Agent Mining

The challenges faced by data mining and distributed data mining have resulted in the new area of research called as 'agent mining'. Agents are defined as software or hardware entities that perform some set of tasks on behalf of users with some degree of autonomy [15].

The Agents;
    1. Should have certain amount of intelligence so that the agent can make the course of action, plan, communicate, adapt changes in the environment and learn from the past experience and increase its knowledge discovery capability [15].
    2. An agent interprets the configuration and generates an execution plan to complete the task. [7][8][10]
    3. Agents can be programmed as compact as possible and can be transmitted across the network.
    4. Agents can work in parallel and share the information they have gathered.
    5. Agents can exchange information and data in the form of message and they termed as multi-agent systems.
    6. Agent's intelligence can range from basic monitoring and data reporting, to more advance forms of decision making.
    7. The data mining agents specializes in performing some activity in the domain of interest.

Evolution of Agent Mining

The evolution process and the characteristic of agent mining are as follows [5]

a. From one way interaction to two way interaction: It was originally initiated by incorporating data mining into agent to enhance agent learning [22].
b. From single need driven to multi need driven: Initially research started on the single need to integrate one into other, but now it is driven by both needs from both parties [23].
c. Intrinsic associations and utilities: This drives the research on mutual issues and the synergetic research and system coupling both technologies in a more advance form [9][21].
d. Application Drives: Application request can be handled by the agents and the mining techniques.
e. Major research groups and researchers: They are trying to link together agent-based distributed learning, agent-based data mining infrastructure, data mining driven agent intelligence enhancement so that they can solve the problems [6].

Dimensions of Agent Mining

As an emerging field of research the study of agent mining focuses on various methodologies, principles, types, techniques and its applications. Based on the type of agent and data mining, the agent mining develops a strong relationship between them which is based on different dimensions.

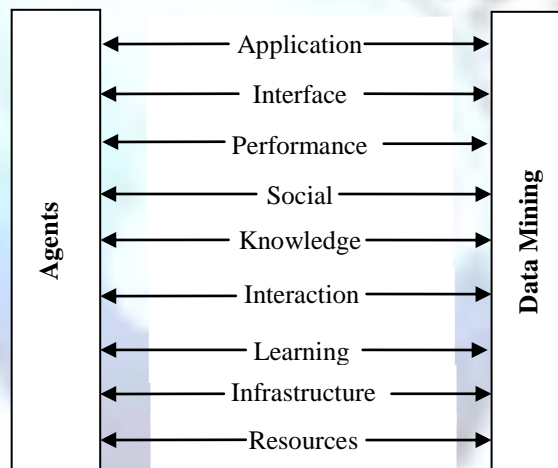The figure shows the multi dimensional agent mining with various layers [5].

Fig 2: Multi-Dimensional agent mining

a. Application Layer: Interaction can be carried on a particular application or a domain problem.
b. Interface Layer: Interaction may be with human, user model or design.
c. Performance Layer: Interaction may be with performance evaluation with various technologies used.
d. Social Layer: Interaction may be on social or organizational factors.
e. Knowledge Layer: Interaction may on the level of the knowledge or extracted or discovered from the resources.
f. Interaction Layer: Interaction may be on co-ordination, co-operation or communication perspective.
g. Learning Layer: Interaction may be on learning methods, capabilities.
h. Infrastructure Layer: Interaction may be on infrastructure, architecture or on processes used.
i. Resource Layer: Interaction may be on the data, or the level of information used.

Types of Agents

Since in distributed data mining, data are distributed geographically much attention is paid towards the cost and accuracy of the gathered data. As defined above agents play a very important role in mining.
The various types of agents are as follows;

1. Collaborative Agent
2. Interface Agent
3. Mobile Agent
4. Facilitator Agent
5. Information Agent
6. Resource Agent
7. Result Agent
8. Reactive Agent
9. Query Agent
10. Local Agent
11. Pre-Processing Agent
12. Post-Processing Agent

1. Collaborative Agent: It emphasizes autonomy and co-operation with other agents in order to perform tasks for their owners.
2. Interface Agent: They are also called as user agent, who interacts with the user and ask the user to provide his requirements and gives the user the visualized results.
3. Mobile Agent: They are the agents capable of roaming in the network, gathering the information on behalf of their owners and perform the duties set by its user.
4. Facilitator Agent: They are also called as manager agent, who is responsible for managing various agents and elaborates the work plan and ensuring that the plan is carried out properly.
5. Information Agent: They are also called as Internet agent, whose responsibility is to manage, manipulate and collate all the information currently available from the distributed resources.
6. Resource Agent: They are also called as data agent, who maintains all the meta-data information and their responsibility is to retrieve the necessary data sets when requested by the data mining agent for a specific mining operation.
7. Result Agent: They obtain the results from the various mining agents; integrate them with the facilitator agent to show the result to the user.
8. Reactive Agent: They represent a special type of agent, which do not possess internal symbolic models of their environment, instead they act and response in a stimulus response manner to the present state of the environment in which they are embedded.
9. Query Agent: It is generated on the demand of the user for the data retrieval from meta data information, which is held by the resource agent.
10. Local Agent: in most of the system data agent is a local agent since most of the sites are local. The local agent can retrieve its local database, performs calculations and returns its result to the system.
11. Pre-Processing Agent: Its responsibility is to cleanse the data before mining.
12. Post-Processing Agent: It evaluates the performance and the accuracy of all the agents.

Application of Agents

The research area of agent mining can be implanted in a variety of applications. Some of the applications are;

1. Business Intelligence
2. Auctions
3. Text Mining
4. Web Mining
5. Peer-to-peer computing
6. Artificial and electronic markets
7. E- Commerce
8. Artificial immune system
9. Health care
10. Customer Management
11. Parallel computing
12. Financial data mining

13. Fraud detection
14. Grid Computing
15. Distributed Data mining

Conclusion/Results

In the recent years the interaction between distributed data mining and agents is one of the hottest research topics. This paper tries to bring out the various challenges faced by the data mining and distributed data mining, which can be overcome by using agents. The author has carried out an overview study of agents, agent mining and various types of agents which are useful in mining the information in the distributed environment. Further research can be made on various new innovative and super intelligent techniques used by the agents in the distributed environment.

References

[1]. Vuda Sreenivasa Rao. Multi-Agent-Based Distributed Data Mining: An Overview. International Journal of Reviews in Computing, 2009-2010. ISSN: 2076-3328, pages 83-92.

[2]. 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

[3]. Dr (Mrs). Sujni Paul. Parallel and Distributed Data Mining. New Fundamental Technologies in Data Mining. ISBN 978-953-307-547-1

[4]. Michael Mock, Dennis Wegener. A data mining based approach to reliable distributed systems.

[5]. Vuda Sreenivasa Rao, S. Vidyavathi, G. Ramaswamy. Distributed Data Mining and Agent Mining Interaction and Integration: A Novel Approach. IJRRAS 4 (4), September 2010, pages 388-398.

[6]. Hyacinth S. Nwana. Software Agents: An Overview. Knowledge Engineering Review, Vol. 11, No 3, pp. 205-244,October/November 1996.

[7]. Sung W. Baik, Jerzy W. Bala, and Ju S. Cho. Agent based distributed data mining. Lecture Notes in Computer Science, 3320:42–45, 2004.

[8]. C. Giannella, R. Bhargava, and H. Kargupta. Multi-agent Systems and Distributed Data Mining. Lecture Notes in Computer Science, pages 1–15, 2004.

[9]. Cao.L. Integration of Agents and Data Mining. Technical Report, 25 June 2005.

[10]. Sven A. Brueckner H. Van Dyke Parunak. Engineering swarming systems. Methodologies and Software Engineering for Agent Systems, pages 341–376, 2004.

[11]. M. Klusch, S. Lodi, and G. Moro. The role of agents in distributed data mining: Issues and benefits. In Proceedings of the 2003IEEE/WIC International Conference on Intelligent AgentTechnology (IAT 2003), pages 211–217, 2003.

[12]. 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

[13]. W. Davies and P. Edwards. Distributed Learning: An Agent-Based Approach to Data-Mining. In Proceedings of Machine Learning 95 Workshop on Agents that Learn from Other Agents, 1995.

[14]. M. Klusch, S. Lodi, and G. Moro. Agent based distributed data mining: The KDEC scheme. AgentLink, Springer Lecture Notes in Computer Science, 2586, 2003.

[15]. S. Russell, P. Norvig. Artifical Intelligence: A Modern Approach, Prentice –Hall, 1995.

[16]. Han.J, and M. Kamber. Data Mining: Concepts and Techniques, Morfan Kaufmann Publishers, 2001.

[17]. H.S. Nwana, M. Wooldridge. Software Agent Technologies, Software Agents and Soft Computing: Towards Enhanced Machine Intelligence, Lecture Notes in Artificial Intelligence 1198, pp 59-77, 1997.

[18]. Cristian Aflori, Florin Leon. Efficient Distributed Data Mining Using Intelligent Agents. In proceedings of 8th International Symposium on Automatic Control and Computer Science, 2004.

[19]. Ayse Yasemin SEYDIM "Intelligent Agents: A Data Mining Perspective", Southern Methodist University, Dallas, 1999.

[20]. Cieslak D.A, Thain.D, Chawla N.V. Short Paper: Troubleshooting Distributed Systems via Data Mining. High Performance Distributed Computing, 2006, 15th IEEE International symposium, pp 309-312, June 19-23,2006.

[21]. Cao.L, Luo. C, Zhang C. Agent-Mining Interaction: An Emerging Area. AIS-ADM 60-73 (2007).

[22]. Brazdil.P, Muggleton.S. Learning to Relate Terms in a Multiple Agent Environment. EWSL 91(1991).

[23]. Cao.L . Agent-Mining Interaction and Interaction- Topics of Research and Development. http://www.agentmining.org/

[24]. Gasser.L, Rosenschein.J.S, Ephrati.E. Introduction to Multi-Agent Systems, Tutorial, Presented at the 1st International Conference on Multi-Agent Systems, San Francisco CA, June 1995.