Distributed Data mining using Multi Agent data

Aliyi Huessein Nure
Lecturer,Bule Hora University-Ethiopia

Abstract: *Multi-Agent Systems (MAS) offers architecture for distributed problem solving. Distributed Data Mining (DDM) algorithms focus on distributed problem solving tasks. In this paper we are discussing about how Multi-Agent Distributed Data Mining approach works on security issues and also discussed about the connection between distributed data mining (DDM) and multi agent system (MAS). Distributed data mining situate from the need of mining on decentralized data sources. Agent based computing aim is to deal with complex data systems has revealed opportunities to improve distributed data mining systems in a number of ways. In many applications the individual and collective behavior of the agents depends on the observed data from distributed sources.*

Keywords: Distributed data mining, Agent mining, KDD, Multi agent system.

I. Introduction:

The Knowledge Discovery (KDD) process is a set of series focused on the discovery of knowledge within databases. The data mining is an application of a number of artificial intelligence, machine learning and statistics techniques to data. Data mining applied with huge amounts of data located at different sites, the amount of data can easily exceed the terabyte limit.

Distributed Data Mining (DDM) is an emerging technology to speed performance and security issues because DDM avoids the transference across the network of very large volumes of data and the security issues occurs from network transferences. Distributed data mining (DDM) mines the data sources regardless of their physical locations. Distributed Data Mining

(DDM) focus on recognizing patterns from distributed heterogeneous data bases in order to compose them within a distributed knowledge base and use for the purposes of decision making.

A data mining agent is a pseudo-intelligent computer program designed to find out specific types of data, along with identifying patterns among those data types. These agents are typically used to detect trends in data, alerting organizations to paradigm shifts so effective strategies can be implemented to either take

advantage of or minimize the damage from alterations in trends. In addition to reading patterns, data mining agents can also "pull" or "retrieve" relevant data from databases, alerting end-users to the presence of selected information.

This paper deals the possible synergy between Multi Agent System (MAS) and Distributed Data Mining (DDM) technology. It particularly focuses on distributed agents, a problem finding increasing number of applications in networks, distributed information retrieval, and many other domains.

Distributed data mining, in particular, for the purposes of distributed classification, has attended active research [1]. However, in this research the prime attention is to date paid to the algorithmic aspects of distributed data mining and combining decisions [2]. In this issue concerning cooperation protocols of distributed software components both in DDM and distributed classification as well as use of new technologies like multi-agent one is paid smaller attention.

I. Distributed Data mining:

Data mining [3] [9] discuss with the problem of analyzing data in scalable manner. DDM is a part of the field of data mining that offers a framework to mine distributed data paying careful attention to the distributed data and computing resources.

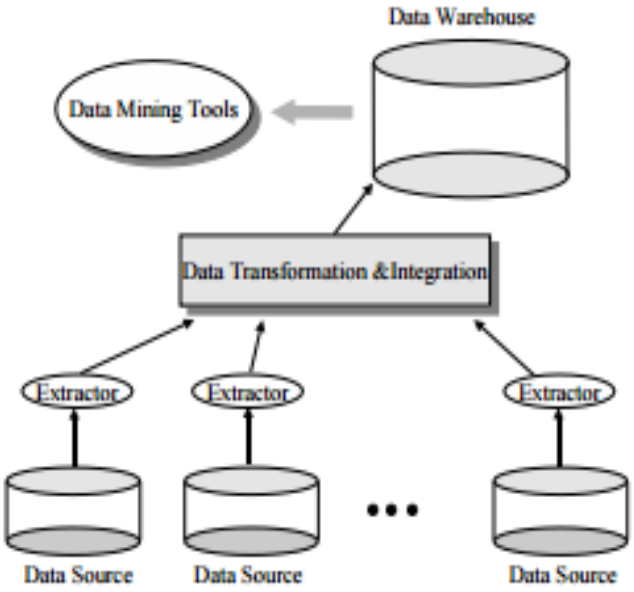

Figure 1: Data warehouse Architecture

Distributed data mining (DDM) considers data mining in this broader context. In figure (1), objective of DDM is to perform the data mining operations based on the type and availability of the distributed resources. It may choose to download the data sets to a single site and perform the data mining operations at a central location.

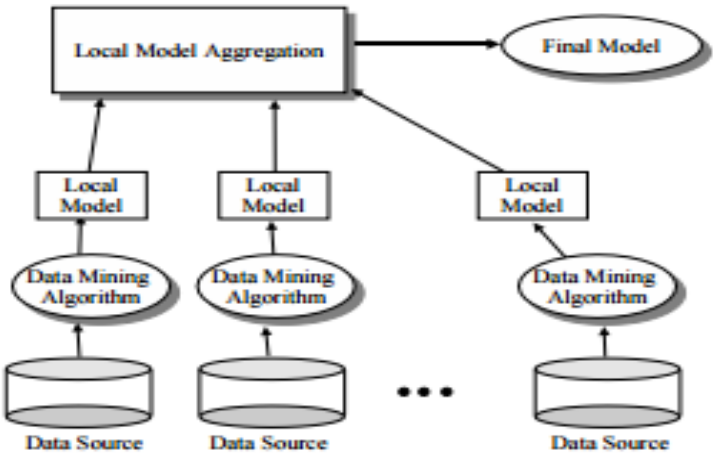

Figure 2: Distributed Data mining Architecture

A distributed architecture for data mining is likely aimed to reduce the communication load and also to reduce the battery power more evenly across the different nodes in the sensor network which shown is Figure (2). One can easily imagine similar needs for distributed computation of data mining primitives in ad hoc wireless networks of mobile devices like PDAs, cell phones, and wearable computers. The wireless domain is not the only example. In fact, most of the applications that deal with time-critical distributed data are likely to benefit by paying careful attention to the distributed resources for computation, storage, and the cost of communication. A distributed approach to analyze this data is likely to be more scalable and practical particularly when the application involves a large number of data sites. Hence, in this case we need data mining architectures that pay careful attention to the distribution of data, computing and communication, in order to access and use them in a near optimal fashion.

Some features of a distributed scenario where DDM is applicable are as follows.

1. The system consists of multiple independent sites of data and computation which communicate only through message passing.

2. Communication between the sites is expensive.

3. Sites have resource constraints e.g. battery power.

4. Sites have privacy concerns.

The privacy issue is playing an increasingly important role in the emerging data mining applications. For example, let us suppose a consortium of different banks collaborating for detecting frauds. If a centralized solution was adopted, all the data from every bank should be collected in a single location, to be processed by a data mining system. Nevertheless, in such a case a distributed data mining system should be the natural technological choice: both it is able to learn models from distributed data without exchanging the raw data between different repository, and it allows detection of fraud by preserving the privacy of every bank's customer transaction data. For what concerns techniques and architecture, it is worth noticing that many several other fields influence Distributed Data Mining systems concepts.

II. Need of Agents:

In Data mining perspective, an agent can be defined as an object whose behavior is described by a "script", with its own calculation ways, and can move from place to place to communicate with other agents. With its "script", the agent is able to follow a life behavior 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, which are Cognitive agents, Reactive 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.

According to J. Ferber [4], A reactive agent is an agent whose behavior responds only to the stimulus or share law, the stimulus is an element of the environment. 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.

III. Challenges in Distributed Data mining and Mining Agents:

Agents can enhance data mining through involving agent intelligence in data mining systems, while an agent system can benefit from data mining via extending agents' knowledge discovery capability [12]. Nevertheless, the agent mining interaction symbiosis cannot be established if mutual issues are not solved [13]. These mutual issues involve fundamental challenges hidden on both sides and particularly within the interaction and integration. Issues in agent-mining interaction highlighting the existence of mutual issues. Mutual issues constraining agent-mining interaction and integration consist of many aspects such as architecture and infrastructure, constraint and environment, domain intelligence, human intelligence, knowledge engineering and management, and nonfunctional requirements.

Architecture and infrastructure Data mining always faces a problem in how to implement a system that can support those brilliant functions and algorithms studied in academia.

Nonfunctional requirements Nonfunctional requests are essential in real-world mining and agent systems. The agent-mining simians may more or less address nonfunctional requirements such as efficiency, effectiveness, action ability, and user and business friendliness.

Constraint and environment both agent and mining systems need to interact with the environment, and tackle the constraints surrounding a system [15]. In agent communities, environment could present characters such as openness, accessibility, uncertainty, diversity, temporality, spatiality, and/or evolutionary and dynamic processes. These factors form varying constraints on agents and agent systems.

Human intelligence both agent and mining need to consider the roles and components of human intelligence. Many roles may be better played by humans in agent-mining interaction.

IV. Using Multi Agent data in Distributed Data mining:

Multi-agent system has revealed opportunities to improve distributed data mining in a number of ways. However, a single data mining technique has not been proven appropriate for every domain and data set [5]. An agent is a computer system that is capable of autonomous action on behalf of its user or owner. An agent is capable to figure out what it is required to be done, rather than just been told what to do [7]. A multi-agent system is one that consists of a number of agents, which interact with one another. We discuss a mining task that involves a number of agents and data sources. Agents are configured to choose an algorithm and deal with given data sets[6]. The performance can be improved because mining tasks can be executed in parallel.

The following framework describes a performance agent which, according to the status established from negotiation and statistics, it is able to determine the strategy to implement the algorithms through clustering agents running on parallel. Figure (3) shows the Multi-Agent System for Distributed Data Mining Framework

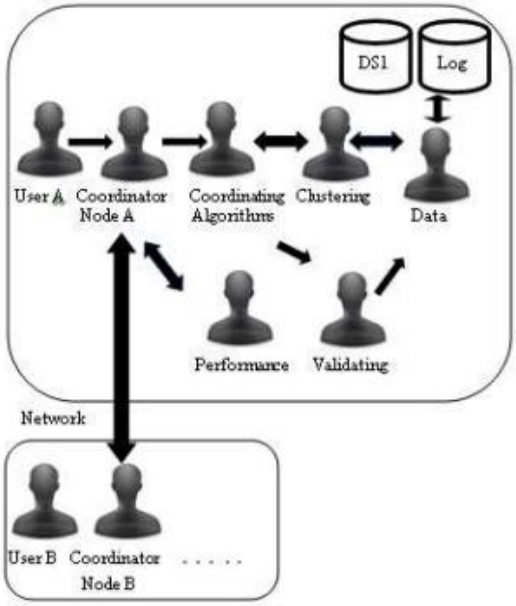

Figure 3: Multi Agent System for Distributed Data mining

The frame work above mentioned, work proposes the implementation of the Multi-Agent System for Distributed Data Mining framework described in previous section [14]. We have developed a web platform through Agent-Oriented Programming paradigm (AOP).

In order to allow inter-agents communication, agents must share the same language, rules, protocols and vocabulary. In order to achieve so, we have followed the recommendations of the standard Foundation for Intelligent, Physical Agents (FIPA)[10]. However, one must define specific ontology's, with its own vocabulary and semantics of the content of the messages exchanged by the agents. We have developed our proposed framework with Java Agent Development (JADE), which integrates a library called "jade.gateway" [8] for the agent programming within a web interface.

The Agent Communication Language may be modified according to system requirements. Message Transport Service (MTS) is a service provided to transport FIPA-ACL messages between agents in any given agent platform and between agents on different agent platforms [11]. The Agent Management System is responsible for managing the operation of an agent platform, such as the creation, deletion, status, overseeing and migration of agents. The Directory Facilitator provides yellow pages services to other agents, maintaining a list of agents and providing the most current information about agents in its directory to all authorized agents.

V. Conclusion:

Distributed data mining and Agent integration has emerged as a prominent and promising area in recent years. Multi-agent systems are fundamentally designed for collaborative problem solving in distributed environments. 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. In this paper, we gave an overview regarding challenges of agent data in distributed data base, 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.

References:

[1] F.Provost and D.Hennessy. Scaling up: Distributed machine learning with cooperation. Working Notes of IMLM-96, 1996, 107–112.

[2] L.Todorovski and S.Dzeroski. Combining classifiers with meta decision trees. Proceedings of 4th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD- 00), Springer Verlag, 2000, 54-64.

[3] Han J. and Kamber M. Data Mining: Concepts and Techniques. Morgan Kaufman Publishers, San Francisco, CA, 2001.

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

[5] M. Wooldridge, "An introduction to multiAgent systems", 2nd ed., John Wiley & Sons, ISBN-10: 0470519460.

[6] S. Chaimontree, K. Atkinson, and F. Coenen, "A multi-agent approach to clustering: Harnessing The Power of Agents," Springer-Verlag, 2012, pp. 16-29.

[7] S. R. Vuda, "Multi agent-based distributed data mining, an overview," International Journal of Reviews in Computing, pp. 83-92, ISSN: 2076-3328, E-ISSN: 2076-3336.

[8] F. Bellifemine, F. Bergenti, G. Caire, and A. Poggi, "JADE: a java agent development framework," Multi-agent Programming: Languages, Platforms, and Applications, Springer-Verlag, 2005, p. 295.

[9] Witten I. and Frank E. Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufman Publishers, San Fransisco, 1999.

[10] FIPA: Communicative Act Library Specification. Tech. Rep. XC00037H, Foundation for Intelligent Physical Agents, 2013.

[11] N.P. Trilok, P. Niranjan, and K.S.Pravat, "Improving performance of distributed data mining (DDM) with multiagent system," International Journal of Computer Science, vol. 9, no. 2& 3, 2012, pp. 74-82, ISSN:1694-0814.

[12] Cao, L., Gorodetsky, V. and Mitkas, P. Agent Mining: The Synergy of Agents and Data Mining. IEEE Intelligent Systems (2009).

[13] Cao, L. Integrating Agent, Service and Organizational Computing. International Journal of Software Engineering and Knowledge Engineering, 18(5): 573-596 (2008).

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

[15] Benny Pinkas. Cryptographic techniques for privacy-preserving data mining. ACM SIGKDD Explorations Newsletter, 4(2):12–19, 2002.

ALiyi Huseein Nure is working as a Lecturer in the Dept of Information Technology functioning under Faculty of Engineering Technology since from 4 years. His field of interest includes Database Systems, Information Retrieval systems, Storage Area Networks and Data mining and Data Warehousing.