e-ISSN: 2349-9745
p-ISSN: 2393-8161

Scientific Journal Impact Factor (SJIF): 1.711

International Journal of Modern Trends in Engineering and Research

www.ijmter.com

A Survey On Ontology Agent Based Distributed Data Mining

Miss Kalyani Sugarwar[1], Mr.Vinay Kapse[2]
[1]Student, M.Tech 3rd sem, VIT, Nagpur
[2]Assistant Professor, VIT, Nagpur

Abstract— With the increased complexity in number of applications and due to large volume of availability of data from heterogeneous sources, there is a need for the development of suitable ontology, which can handle the large data set and present the mined outcomes for evaluation intelligently. In the era of intensive data driven applications distributed data mining can meet the challenges with the support of agents. This paper discusses the underlying principles for effectiveness of modern agent-based systems for distributed data mining.

Keywords—*Distributed data mining, Multi-agent System, Ontology agent.*

I. INTRODUCTION

Relational database systems lead the way of efficient storage and retrieval of data from the organized set of information long back. Knowledge discovery from the databases presents another approach where pattern from the large repository is extracted and analyzed. Data is almost growing in exponential terms and most of the relevant applications are driven by it. One simple data mining technique cannot provide the required solution for it. Main challenges of data mining includes, existence of huge data set located at different sites, intensive computation process with large set and rapid change in input data. The Distributed Data Mining (DDM) mines data sources irrespective of their physical locations. An intelligent agent is an autonomous entity or software program that automatically performs tasks on behalf of the user. An intelligent agent or group of such agents act on data from heterogeneous sources and distributed in proactive manner. A multi-agent system is a system composed of multiple interacting intelligent agents. The development of multi-agent system is increasing with popularity of the internet usage provides the basis for an open environment where agents interact with each other to reach their individual or shared goals. A multi-agent system is a loosely coupled network of problem-solver entities that work together to find solution to problems that are beyond the individual capabilities or knowledge of each entity. Ontology is defined as "an explicit specification of a conceptualization" or "a set of types, properties, and relationship" . Ontology defines concepts in a specific area and relationships; their relationships; however, ontology is more than an agreed-on term. It has a set of well-defined constructs that can be leveraged to build structured knowledge. The agent paradigm is successfully employed in those applications where autonomous, loosely coupled, heterogeneous, and distributed systems need to interoperate in order to achieve a common goal. In a multi-agent system the agents communicate between them in order to fulfill the global goal or their local goals. Also, ontology has established as a powerful tool to enable knowledge sharing, and a growing number of applications have benefited from the use of ontology as a means to achieve semantic interoperability among heterogeneous, distributed systems. Both ontology and agent technologies are central to the semantic web, and their the sharing of heterogeneous, autonomous knowledge sources in a capable, adaptable and extensible manner. This is particularly important for multi-agent systems, where the content of messages exchanged among agents must conform to some ontology in order to be understood. Through the collaborations among different agents, the anticipation

is to achieve a highly efficient, flexible, customizable system that provides better communication, interaction and management among different components of agent based distributed data mining system. Ontology is used throughout the multi-agent system to assist the interactions among different agents as well as to improve the quality of the service provided by each agent. Different systems and models have been developed so far to achieve the advantages of both multi-agent systems and ontology paradigm. Data mining is the extraction of useful patterns and relationships from heterogeneous data sources, such as databases, texts and the web. It uses statistical and pattern matching techniques. Data mining does borrow from statistics, machine learning, databases, information retrieval, data visualization and other fields. Many areas of science, business, and other environments deal with a vast amount of data, which needs to be turned into something meaningful, knowledge. Data Mining provides very interesting information that statistical methods are unable to produce as effectively or properly. The data we have is often vast, and noisy, meaning that it's imprecise and the data structure is complex. This is where a purely statistical technique would not succeed, so data mining is a solution. The issues in data mining are noisy data, missing values, static data, sparse data, dynamic data relevance, interestingness, heterogeneity algorithm efficiency, size and complexity of data. This paper presents a survey of the research related to the ontology agent based distributed data mining. The rest of the paper is organized as follows: Section II reviews research work related to multi-agent distributed data mining .Section III reviews framework for Ontology agent distributed data mining.

II. RELATED WORK

A) **Multi-Agent Based Distributed Data Mining System**:

The Distributed Data Mining (DDM) is a branch of the field of data mining that offers a framework to mine distributed data paying careful attention to the distributed data and computing resources. Usually, data-mining systems are designed to work on a single dataset. On the other hand with the growth of networks, data is increasingly dispersed over many machines in many different geographical locations. Also, even as most practical data-mining algorithms operate over propositional representations are known as first order learning. In existing system, the concept of knowledge is very important in data mining. In order to get the correct knowledge from the data mining system, the user must define the objective and specify the algorithms and its parameters exactly with minimum effort. Even choosing the correct data mining algorithm involves more time for the system. Developing a data mining system that uses specialized agents with the ability to communicate with multiple information sources, as well as with other agents requires a great deal of flexibility. The main objective of this paper titled on "An Analysis on Multi-Agent Based Distributed Data Mining System "describes the knowledge integration, Knowledge Integration in Distributed Data-Mining and Heterogeneous vs. Homogeneous Data-Mining, a literature survey of Multi-Agent Based Distributed Data Mining System, a Model Of Multi –Agent System Based Data Mining, the improving DDM performance by combining distributed data mining and multi-agent system and Data Mining using Multiple Agents.

B) **Improving performance of Ontology-Agent Based Distributed Data Mining System**:

Autonomous agents and multi-agent systems (or agents) and knowledge discovery (or data mining) are two of the most active areas in information technology. Ongoing research has revealed number of intrinsic challenges and problems facing each area, which can't be addressed solely within the confines of the respective discipline. A profound insight of bringing these two communities together has unveiled a tremendous potential for new opportunities and wider applications through the synergy of agents and data mining. With increasing interest in this synergy, agent mining is emerging as a new research field studying the interaction and integration of agents and data mining.

In this paper, we give an overall perspective of the driving forces, theoretical underpinnings, main research issues, and application domains of this field, while addressing the state-of-the-art of agent mining research and development.

III. PROPOSED WORK ON ONTOLOGY AGENT DISTRIBUTED DATA MINING

Many types of system have been developed for distributed data mining. These systems can be classified according to their strategy. Central learning, Meta-learning, and Hybrid learning are the three types of strategies made for this purpose. In the central learning strategy all the data can be gathered at a central site and a single model can be build. The key requirement for the same is to be able to move the data to a central location in order to merge them and then apply sequential Data Mining algorithms. This strategy is used when the geographically distributed data is small. The strategy is generally very expansive but also more accurate. The process of gathering data in general is not simply a merging step; it depends on the original distribution. For example, different records are placed in different sites, different attributes of the same records are distributed across different sites, or different tables can be placed at different sites, therefore when gathering data it is necessary to adopt the proper merging strategy. However, as pointed before this strategy in general is unfeasible. Agent technology is not very preferred in such strategy.Meta-learning strategy offers a way to mine classifiers from homogeneously distributed data. The first step of meta-learning is to generate base classifiers at each site using a classifier learning algorithms. The second step is to collect the base classifiers at a central site, and produce meta-level data from a separate validation set and predictions generated by the base classifier on it. The third step is to generate the final classifier (meta-classifier) from Meta level data via a combiner or an arbiter. Copies of classifier agent will exist or deployed on nodes in the network being used. The most mature systems of agent-based Meta learning systems are: JAM system, and BODHI. Hybrid learning strategy is a technique that combines local and centralized learning for model building. Papyrus is designed to support both learning strategies. Papyrus cannot only move models from site to site, but can also move data when that strategy is desired. Papyrus is a specialized system, which is designed for clusters while JAM and BODHI are designed for data classification.

On looking analytically we find the following facts: It is not always possible to obtain an exact final result, i.e. the global knowledge model obtained may be different from the one obtained by applying the one model approach (if possible) to the same data. In these systems hardware resource usage is not optimized. If the heavy computational part is always executed locally to data, when the same data is accessed concurrently, the benefits coming from the distributed environment might vanish due to the possible strong performance degradation. One more drawback is that occasionally, these models are induced from databases that have different schemas and hence are incompatible. Ontology Agent: It maintains and provides overall knowledge of ontology and answers queries about the ontology. It may simply store the ontology as given, or it may be as advanced as to be able to use semantic reasoning to determining the applicability of a domain.

IV. CONCLUSION

The existing approach of multi-agent data mining system in distributed environment. The analytical survey and literature view emphasize the gradual development and significance of this field. The refinement in existing approach may be observed with the development of suitable ontology, communication interface, and development of multi-agent system on the right back. Ontology remove the the drawback of models are induced from databases that have different schemas and hence

are incompatible, It maintains and provides overall knowledge of ontology and answers queries about the ontology.

REFERENCES

[1]. Sung W. Baik, Jerzy W. Bala, and Ju S. Cho. "Agent based distributed data mining -- Lecture Notes in Computer Science", 3320:42-45, 2004.
[2]. 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.
[3]. A. Bordet sky. "Agent-based Support for Collaborative Data Miningin Systems Management". In Proceedings of The Annual Hawaii International Conference On System Sciences, page 68, 2001.
[4]. R. Bose and V. Sugumaran. IDM: an intelligent software agent based data mining environment. 1998 IEEE International Conference on Systems, Man, and Cybernetics, 3, 1998.
[5]. DE Loach, S., M. Wood and C. Sparkman, "Multiage Systems Engineering", The International Journal of Software Engineering and Knowledge Engineering, Vol 11 no. 3, (June 2001)
[6]. Dumitru Roman, Digital Enterprise Research Institute "Agents Communication Languages (ACL)"

