International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 11, ISSN No. 2455-2143, Pages 11-16
Published Online March 2018 in IJEAST (http://www.ijeast.com)

DISTRIBUTED DATA MINING USING MULTI AGENT DATA

Shahid Hussain Wani
Department of Computer Science and Engineering,
Desh Bhagat University, Punjab India.

Abreen Rashid
Department of Computer Science and Engineering,
Desh Bhagat University, Punjab India.

Mohammad Ilyas Malik
Department of Computer Science and Engineering,
Desh Bhagat University, Punjab India

Abstract: **A multi-agent system (MAS or "self-organized system") is a computerized system composed of multiple interacting intelligent agents. Multi-agent systems can solve problems that are difficult or impossible for an individual agent or a monolithic system to solve. Distributed Data Mining (DDM) algorithms mainly focus on distributed problem-solving tasks. In this paper, we will examine how Multi-Agent Distributed Data Mining approach deals with security issues and furthermore talk about the association between distributed data mining (DDM) and multi-agent system (MAS). Distributed data mining having a site from the need of mining on decentralized information sources. Agent based computing aim is to manage complex information frameworks has uncovered chances to enhance disseminated information mining frameworks in various ways. In numerous applications, the individual and aggregate conduct of the specialists relies upon the watched information from dispersed sources.**

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

I. INTRODUCTION

Knowledge discovery in databases (KDD) refers to the overall process of discovering useful knowledge from data [37]. **KDD** is an integration of multiple technologies for data management such as database management and data warehousing, statistic machine learning, decision support, and others such as visualization and parallel computing [37].

KDD refers to the overall process of discovering useful knowledge from data, and data mining refers to a particular step in this process [38]. The basic problem addressed by the **KDD** process is one of mapping low-level data into other forms that might be more compact, more abstract, or more useful [38]. **KDD** focuses on the overall process of knowledge discovery from data, including how the data are

stored and accessed, how algorithms can be scaled to massive data sets ultimate and still run efficiently, how results can be interpreted and visualized, and how the overall man-machine interaction can usefully be modeled and supported [38].

Distributed computing plays an important part in the Data Mining process for a few reasons. To start with, Data Mining frequently requires enormous measures of storage space and computation time. To make system adaptable, it is important to develop the mechanism that disseminates the workload among a few locales adaptably or in a flexible way. Second, information is regularly inherently dispersed into a few databases, making a unified preparing of this information extremely wasteful and inclined to security dangers. Distributed Data Mining investigates methods of how to apply Data Mining in a non-concentrated manner. At long last, numerous information mining assignments require associating heterogeneous resources, for example, information sources, preparing hubs, and end-client applications.

A data mining agent [1] is a pseudo-keen computer program [1] that are intended to seek out the particular kinds of information, alongside recognizing designs among those information composes. These specialists are normally used to identify trends in information, alarming associations to perspective changes so powerful procedures can be executed to either exploit or limit the harm from adjustments in patterns. Notwithstanding perusing designs, information mining operators can likewise "pull" or "retrieve" important information from databases, cautioning end-clients to the nearness of chose data.

This paper deals with the conceivable joined activities or tasks between Multi-Agent System (MAS) and Distributed Data Mining (DDM) technology. It especially centers on appropriated operators an issue finding the expanding number of uses in systems dispersed data recovery and numerous different spaces

Distributed data mining, specifically, for the reasons for appropriated grouping, has attended active research [2]. However in this research, the prime consideration is to date paid to the algorithmic parts of distributed information mining and consolidating choices [3]. In this issue concerning participation conventions of conveyed

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 11, ISSN No. 2455-2143, Pages 11-16
Published Online March 2018 in IJEAST (http://www.ijeast.com)

programming segments both in DDM and appropriated arrangement, and also the utilization of new innovations like multi-agent one, is given careful consideration.

II. DISTRIBUTED DATA MINING

Data mining innovation has risen as a method for recognizing examples and patterns from substantial amounts of information [17]. Data mining [16] [9] examine the issue by analyzing data in a versatile way. DDM is a part of the field of data mining that offers a system to mine appropriated information giving careful consideration to the distributed data and computing resources.

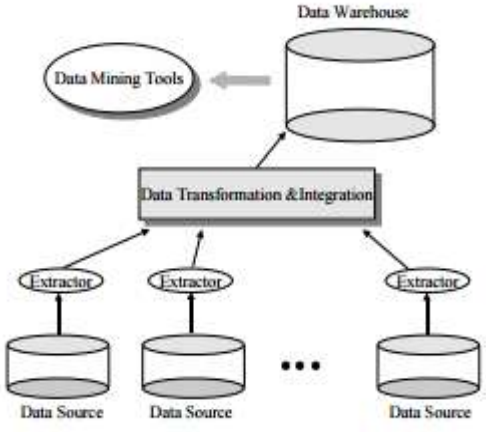

Figure 1: Data warehouse Architecture

Distributed data mining (DDM) considers data mining in this more extensive context. In figure (1), a target of DDM is to play out the data mining activities in light of the sort and accessibility of the distributed resources. It might download the informational collections to a solitary website and perform the data mining tasks at a focal area or central location.

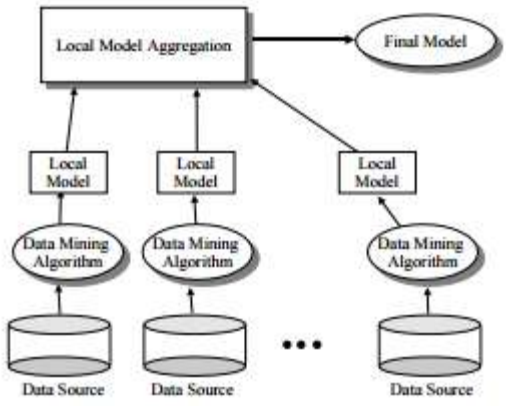

Figure 2: Distributed Data mining Architecture

A distributed architecture for data mining is likely meant to decrease the correspondence load and furthermore to diminish the battery control all the more uniformly over the diverse hubs (different nodes) in the sensor network which demonstrated is Figure (2). One can easily imagine [18] similar needs for distributed computation of data mining primitives in ad hoc wireless networks of mobile devices like PDAs, PDAs, and wearable PCs. The wireless domain isn't the only illustration. Indeed, a large portion of the applications that arrangement with time-basic distributed data are probably going to profit by giving careful consideration to the distributed resources for computation, storage, and the cost of communication. A dispersed way to deal with break down this information [19] is probably going to be more versatile and commonsense especially when the application includes a substantial number of information destinations. Thus, for this situation we require data mining designs that give careful consideration to the distribution of data, computing and communication, with a specific end goal to access and utilize them in a close ideal manner.

In the DDM literature, one of two suspicions [22] is normally received in the matter of how information is distributed across sites: homogeneously (horizontally partitioned) and heterogeneously (vertically partitioned) [22]. The two perspectives [23] embrace the reasonable perspective that the data tables at each site are partitions of a single global table. In the homogeneous case [23], the global table is partitions of a single global table. The tables at each site are subsets of the global table; they have the very same attributes. In the heterogeneous case [24], the table is vertically parceled, each site contains a gathering of columns (sites do not have the same attributes). Be that as it may, each tuple at each site is accepted to contain a one of a kind identifier to encourage coordinating. Stress that the global table perspective is entirely calculated. It isn't really expected that such a table was physically acknowledged and partitioned to shape the tables at each site. The improvement of data mining algorithms [25] that function admirably under the imperatives forced by distributed datasets has gotten huge subterranean insect consideration from the data mining group in recent years. The field of DDM has risen as a dynamic territory of study. The main part of DDM strategies in the writing work over a unique engineering which incorporates different destinations having free figuring power and storage capability. Nearby calculation is done on every one of the sites and either a central site communicates with each distributed site to process the global models or a distributed engineering is utilized. In the latter case, singular hubs may communicate with an asset rich concentrated hub or node, however they perform the majority of the tasks by communicating with neighboring hubs or nodes by message passing over an asynchronous network. For example, the sites may represent to

autonomous sensor hubs which associate with each other in a specially appointed manner.

A few highlights of a circulated situation where DDM is applicable are as follows.

1. The system comprises of various autonomous destinations of information and calculation which convey just through message passing.

2. Communication between the sites is expensive.

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

4. Sites have privacy concerns.

The privacy issue is assuming an undeniably vital part in the developing information mining applications [20]. For example, let us assume a consortium of various banks working together to detects frauds. On the off chance that a brought together arrangement was embraced, every one of the information from each bank ought to be gathered in a solitary area, to be handled by a data mining system [21]. Nevertheless, in such a case a distributed data mining system ought to be the normal technological decision: both it can take in models from conveyed information without trading the crude information between various repository, and it permits discovery of extortion by protecting the security of each bank's client exchange information. For what concerns procedures and design, it merits seeing that numerous few different fields impact Distributed Data Mining systems ideas.

III. NEED OF AGENTS

In Data mining viewpoint, an agent [26] can be characterized as an object whose behavior is described by a "script", with its own particular computation ways, and can move from place to place to speak with different specialists or agents. With its "script", the specialist or agent can take after an existence behavior that will be ingrained at the time of usage and that will enable him to have the fundamental component to be completely autonomous. One of the segregate attributes of the agents [27] is the representation and reasoning on the environment (the outer world and different specialists), in light of this element, we discover two distinct classes [28], which are Cognitive agents, Reactive agents.

A psychological agent is an agent that has an explicit representation of its motivation and its condition. The activities it performs to accomplish its objective [29] are the result of a thinking on the condition of the earth. More often than not, a psychological framework [30] incorporates few specialists; each is like a pretty much complex master framework or system. For this situation, we talk about high granularity specialist.

According to J. Ferber [4], A reactive agent is an agent whose behavior reacts just to the stimulus or share law, the stimulus is a component of the earth. Ordinarily a reactive system has an expansive number of low granularity agents. These agents don't really have an unequivocal or explicit objective to get. By cons, they can implement an intricate thinking on their inner state to play out their activities.

IV. CHALLENGES IN DISTRIBUTED DATA MINING AND MINING AGENTS

Agents can upgrade data mining by including agent intelligence in data mining systems, while an agent system can benefit from data mining by means of broadening agent's knowledge discovery capability [12]. Nevertheless, the agent mining association beneficial interaction can't be set up if shared issues are not unraveled [13]. These shared issues include essential difficulties covered up on the two sides and especially inside the communication and joining. Issues in agent mining communication featuring the presence of shared issues. Shared issues compelling agent mining connection and joining comprise of numerous perspectives, for example, design and framework, limitation and condition, space insight, human insight, learning building and administration, and nonfunctional prerequisites.

Architecture and infrastructure Data mining dependably faces an issue in how to implement a framework or system that can support those splendid capacities and calculations contemplated in the scholarly community.

Nonfunctional prerequisites Nonfunctional solicitations are essential in real-world mining and agent systems. The agent-mining simians may pretty much address nonfunctional necessities, for example, efficiency, efficiency, acting capacity, and client and business neighborliness.

Constraint and environment both agent and mining systems need to associate with the environment and handle the requirements encompassing a framework [15]. In agent groups, nature could show characters, for example, receptiveness, openness, vulnerability, assorted variety, fleetingness, spatiality, or potentially transformative and dynamic procedures. These elements frame shifting imperatives on agents and agent systems.

Human insight both agent and mining need to think about the parts and segments of human intelligence. Numerous parts might be better played by people in agent mining interaction.

International Journal of Engineering Applied Sciences and Technology, 2018
Vol. 2, Issue 11, ISSN No. 2455-2143, Pages 11-16
Published Online March 2018 in IJEAST (http://www.ijeast.com)

V. USING MULTI AGENT DATA IN DISTRIBUTED DATA MINING

Multi-agent system has uncovered chances to enhance distributed data mining in various ways [31]. However, a solitary data mining technique [32] has not been demonstrated proper for each domain and dataset [5]. An agent is a computer system that is equipped for self-ruling activity in the interest of its user or owner. An agent is skilled to make sense of what it is required to be done, as opposed to simply be instructed [7]. A multi-agent system is one that comprises of various agents, which cooperate with each other [33]. We talk about a mining undertaking that includes various agents and data sources. Agents are configured to pick an algorithm and manage given data sets [6]. The execution can be enhanced on the grounds [pp] that mining assignments can be executed in parallel.

The accompanying structure depicts a performance agent which, as per the status built up from negotiation and statistics, it can decide the system to implement the algorithms through clustering agents running on parallel. Figure (3) demonstrates the Multi-Agent System for Distributed Data Mining Framework.

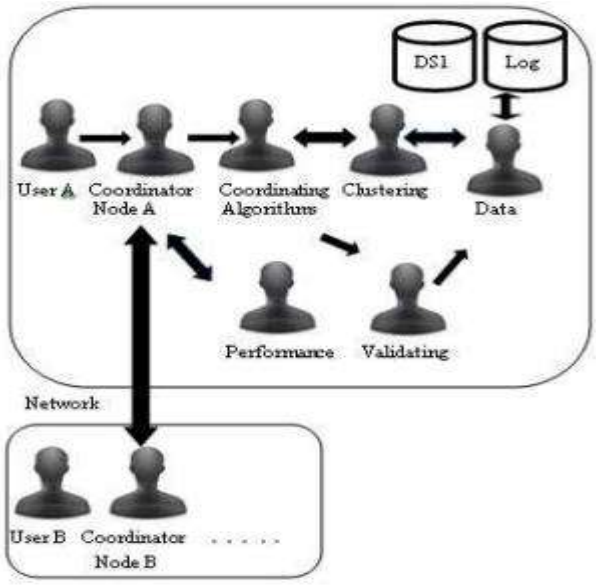

Figure 3: Multi Agent System for Distributed Data mining

The structure above mentioned, work proposes the usage of the Multi-Agent System for Distributed Data Mining structure described in past segment [14]. We have built up a web platform [34] through Agent-Oriented Programming worldview (AOP).

With a specific end goal to permit inter-agents communication, agents must have a similar language, guidelines, conventions, and vocabulary. With a specific end goal to accomplish thus, we have taken after the proposals of the standard Foundation for Intelligent [36], Physical Agents (FIPA) [10]. In any case, one must define particular ontology's [35], with its own particular vocabulary and semantics of the content of the messages exchanged by the agents. We have built up our proposed structure with Java Agent Development (JADE), which integrates a library called "jade.gateway" [8] for the agent programming inside a web interface.

The Agent Communication Language might be adjusted by system requirements. Message Transport Service (MTS) is an administration gave to transport FIPA-ACL messages between agents in any given agent platform and between agents on various agent platforms [11]. The Agent Management System is in charge of dealing with the task of an agent platforms, for example, the creation, deletion, status, directing and migration of agents. The Directory Facilitator gives business index administrations to different agents, maintaining a list of agents and giving the most current data about agents in its directory to all authorized agents.

VI. CONCLUSION AND FUTURE ENHANCEMENT

Multi-agent systems are fundamentally designed for collaborative problem solving in distributed environments. Distributed data mining and Agent integration have emerged as a noticeable and promising area in recent years. Multi-agent systems are fundamentally designed for collaborative problem-solving in distributed environments. Continuous research has demonstrated various difficulties and inherent impediments faced by every region. However, the synergy between the two technologies offers great potential and open doors for more refined applications. In this paper, we gave an overview with respect to difficulties of specialist information in the distributed database, main thrusts, hypothetical establishments, real research issues and regions of utilization of this blend, considering the cutting edge explore advancement Data mining and multi-agent. This paper suggests that traditional centralized data mining techniques mining procedures may not function admirably in numerous distributed environments where data centralization might be troublesome as a result of constrained transmission capacity, security issues or potentially the request on response time. This paper pointed out that distributed data mining algorithms may offer a better solutions since they are intended to work in an appropriated domain by giving careful consideration to the computing and communication resources.

VII. REFRENCES

[1] Michael Anissimov, Niki Foster, "What is Data Mining", in wiseGEEK, 21 February 2018 Retrieved from http://www.wisegeek.com/what-is-data-mining.htm

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

[3] 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.

[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.

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

[17] S. SulaihaBeevi, K. L. Shunmuganathan, "Multi-agent system technology for distributed data mining and clustering using JADE" January 2015, Retrieved from https://www.researchgate.net/publication/283862190_Multi_agent_system_technology_for_distributed_data_mining_and_clustering_using_JADE

[18] R. Chen, K. Sivakumar, H. Kargupta, "Collective Mining of Bayesian Networks from Distributed Heterogeneous Data" in Knowledge and Information Systems, March 2004, Volume 6, Issue 2, pp 164–187, Retrieved from https://rd.springer.com/article/10.1007%2Fs10115-003-0107-8

[19] Cisco solutions, "The easy guide to data and voice networking", Retrieved from https://www.scribd.com/document/88217552/The-Easy-Guide-to-Data-and-Voice-Networking

[20] Sanjay Singh Solanki, Naina Songira, "Big Data: The Key to Economic Growth & Development", Vishleshan-International Journal of Engineering and Management (VIJEM) Volume 2, Issue 4 (July-Sep, 2017) ISSN (Online): 2455 - 524X, Retrieved from http://www.vijem.com/Issue/Vol2Iss4/Papers/VIJEM_242_2017.pdf

[21] Dr. Masoud Yaghini, "Data Mining Functionalities", 2009, Retrieved from http://webpages.iust.ac.ir/yaghini/Courses/Data_Mining_881/DM_01_02_Data%20Mining%20Functionalities.pdf

[22] Josenildo C.da Silvaa, Chris Giannellab, Ruchita Bhargava Hillol Kargupta, Matthias Klusch, "Distributed data mining and agents" in Engineering Applications of Artificial Intelligence, Volume 18, Issue 7, October 2005, Pages 791-807, Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S095219760500076X

[23] MikeRayMSFT, Carl Rabeler, Sven Aelterman, Craig Guyer, "Partitioned Tables and Indexes" in Microsoft, 01/20/2016, Retrieved from https://docs.microsoft.com/en-us/sql/relational-databases/partitions/partitioned-tables-and-indexes

[24] Trilok Nath Pandey, Niranjan Panda, Pravat Kumar Sahu, "Improving performance of distributed data mining (DDM) with multi-agent system", in IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 2, No 3,

March 2012 ISSN (Online): 1694-0814 , Retrieved from http://www.ijcsi.org/papers/IJCSI-9-2-3-74-82.pdf

[25] https://www.tutorialspoint.com/data_mining/dm_quick_guide.htm

[26] Aliyi Huessein Nure, "Distributed Data mining using Multi Agent data" in International Research Journal of Engineering and Technology (IRJET) , e-ISSN: 2395 -0056, Volume: 04 Issue: 3 | Mar -2017, www.irjet.net, p-ISSN: 2395-0072, Retrieved from https://www.irjet.net/archives/V4/i3/IRJET-V4I3133.pdf

[27] https://www.igi-global.com/dictionary/agent/837

[28] https://en.wikipedia.org/wiki/Talk:Intelligent_agent

[29] https://www.revolvy.com/main/index.php?s=Management&item_type=topic

[30] Irene Vasilachis de Gialdino, "Ontological and Epistemological Foundations of Qualitative Research" in Forum Qualitative Social Research Sozialforschung, Volume 10, No. 2, Art. 30 – May 2009, Retrieved from

http://www.qualitative-research.net/index.php/fqs/article/view/1299/3163

[31] S. SulaihaBeevi, K. L. Shunmuganathan, "Multi-agent system technology for distributed data mining and clustering using JADE" January 2015, Retrieved from https://www.researchgate.net/publication/283862190_Multi_agent_system_technology_for_distributed_data_mining_and_clustering_using_JADE

[32] M.Govindarajan, RM.Chandrasekaran, "Intrusion detection using neural based hybrid classification methods" in Computer Networks, Volume 55, Issue 8, 1 June 2011, Pages 1662-1671, Retrieved from https://www.sciencedirect.com/science/article/pii/S1389128610003750

[33] Pradeep Peiris, Mark Weiser, "How IoT Is Strengthening Ubiquitous Computing (Part 2)" in / IoT Zone, · Apr. 24, 17 · IoT Zone · Opinion, Retrieved from https://dzone.com/articles/how-iot-is-strengthening-ubiquitous-computing-part-2

[34] Eug´enio Oliveira, Paulo Martins, "Agent-based approach for Ship Damage Control" , Retrieved from https://web.fe.up.pt/~niadr/PUBLICATIONS/2016/WASA.pdf

[35] Aliyi Huessein Nure," Distributed Data mining using Multi Agent data", in International Research Journal of Engineering and Technology (IRJET), e-ISSN: 2395 -0056, Volume: 04 Issue: 3 | Mar -2017, www.irjet.net, p-ISSN: 2395-0072 , Retrieved from https://www.irjet.net/archives/V4/i3/IRJET-V4I3133.pdf

[36] http://www.fipa.org/specs/fipa00001/SC00001L.html

[37] M. De Martino, A. Bertone, R. Albertoni, H. Hauska, U. Demsar, M. Dunkars. Technical Report of Data Mining, INVISIP IST-2000-29640, Information Visualisation for Site Planning, WP No2: Technology Analysis, D2.2, 28.2.2002

[38] U. Fayyad, G. P.-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. AI Magazine, 17(3):37-54, Fall 1996.
http://citeseer.ist.psu.edu/fayyad96from.html