January-February 2020
ISSN: 0193 - 4120 Page No. 10924 - 10930

An Exercise on Multi Data Agent with Distributed Data Mining and Related Sequence Modeling

K.D. Gupta
Assistant Professor, Department of Computer Science
Apex University, Jaipur
kdevgupta@gmail.com

Article Info
Volume 82
Page Number: 10924 - 10930
Publication Issue:
January-February 2020

Article History
Article Received*: 18 May 2019*
Revised*: 14 July 2019*
Accepted*: 22 December 2019*
Publication*: 19 February 2020*

Abstract

The data mining technology has become a way to identify patterns and trends in large amounts of data. Mining technology commonly adopted method of data integration generate data stores, which collect all data in one central location, and then run a algorithm against the data to extract the module prediction and evaluation of useful knowledge. However, a mining technique alone was not tested for every area and data set. Technical data mining involved in such a complex environment must meet high dynamic due to changes in the system can affect the overall system performance. Agent team whose goal is to address complex systems has revealed opportunities to improve distributed data mining systems in a number of ways. The multi-agent systems (MAS) often deal complex applications that require troubleshooting distributed. In many applications the individual and the collective behavior of agents depends on the observed data from distributed sources.

Keywords**:** Multi Agent, Knowledge Discovery, Distributing Mining, Sequence Modeling

1. Introduction

Distributed data mining is the need of mining in distributed data sources. The field of Distributed Data Mining (DDM) addresses these challenges in the analysis of data distributed and offers many different algorithmic solutions for operations data analysis and mining of a fundamentally distributed manner that pays attention special resource constraints. Since multi-agent systems are often distributed and proactive and reactive agents have features that are very useful for knowledge management systems, the combination of DDM with MAS for data-intensive applications is attractive. This document describes the integration of multi-agent and exploration of distributed data system, also known as multiple data mining based distributed agents in terms of importance, system overview, systems existing search and trends.

Agents and multi agent systems need to have a significant influence to achieve a vision of global service network and information rich to support the interaction with the dynamic discovery of digital business, emerging technology. Since the agent claimed that the first strong platform was the next stage of software development, the important work on the multi-agent system has already been over 10 years, multi agent and innovative e - business application. The remaining section of this report is to briefly outline the agent system, multi agent system. After observation, this research explains some of the most common multi agent systems on the development platform.

A native of **knowledge discovery from databases** (KDD), also known as data mining (DM), extraction of data sources distributed data (DDM) mines, regardless of their physical location. The need for such a feature results from the fact that data produced locally at each site often cannot be transferred through the network due to the excessive amount of data and confidentiality. Recently, DDM has become an essential component of knowledge-

January-February 2020
ISSN: 0193 - 4120 Page No. 10924 - 10930

based systems, due to its decentralized architecture reaches all networking companies. Mining still poses many challenges for the research community.

2. Motivation

Data Preprocessing is one of the significant factors that affects on the success of Data Mining and Machine Learning approaches. Generally, data preprocessing stage represents the quality of data. This should not present incorrect and incomplete data, the result is unreliable. However, data preprocessing task is time consuming because it includes many phases like data cleaning, data integrating, data transforming and data reducing. However, the best performance of the data pre-processing algorithms is relied on the nature of each data set. Hence, it would be nice, if the research have interested methodology to adapt for choosing the best performance of the data preprocessing algorithms for each data set.

In Current process, every researcher of data mining (DM), to be included the existence of many problems: accuracy, efficiency and elective subjects, privacy and security, and scalability. Accuracy is particularly important in the context of classification. Effectiveness issues and elections will keep DM discipline. The security center on confidentiality and legal issues and the desire of the owner of many data holds that data their copyright. Because the scalability problem is particularly important, the amount of data currently available for DM Scope and year by year rapidly increasing year. A potential solution to the problem of capacity scalar parallel or distributed DM, although it often requires significant communication costs.

3. Multi-Agent Systems

To solve the problem, the resultant system is a multi agent system (MAS) by combining multiple agents into a single system. These systems consist of medicines to solve the simple problem of an overall problem individually. They can communicate and help each other to achieve greater goals and more complicated. In this way, the problem that software developers were previously considered too complex can now be solved by placing the problem Resolution. For example, MAS is used to predict the stock market, industrial automation and e-learning.

In general, MAS conforms to the following three characteristics. First, the MAS need to specify the appropriate interaction with the communication protocol. Although the agent is an element of problem solving architecture, individual problems cannot be solved if general communication fields and intact protocols exist, Secondly, the MAS needs to be open and decentralized without prior knowledge, for example, the number of participants and actions. In MAS running, new medicines are often unpredictable. They can participate at any time to comply with communication protocols, to act selectively. Finally, the MAS may be made up of foreign substances that are scattered around the environment and are likely to work alone or jointly.

4. Strategy Planned

This work presents the concepts related to MADM vision. A study of the part that avoids the use of the Meta language format and focused on the core features of the framework has agreed on a framework that supports wrappers. Suggestion Investigate the nature of the drug framework / platform used to evaluate and evaluate ideas on the operation of MADM and establishment of multi agent framework. The purpose of these studies is to identify and evaluate the problem of this outcome in context MADM to assess.

5. Analysis of Activity Sequence Control

The field of E-Learning Data Mining is a wide (EDM) and data mining that combines aspects and questions from various fields (P. Example: E-learning / Distance learning, etc. Machine learning adaptive system, the authors [Srivastava *et al*., 2000] (a) classify work by searching educational data of statistical and visualization (b) can search the web Cluster, abnormal value and classification detection, extraction of association rule and sequential extraction pattern, and text timing. The decompression web further integrates classification and place in offline web search, to find models and other information to support education to learn online web and validate models of quest model discovered can be online equipped with "intelligent" systems that can help students of their learning efforts.

6. Monitoring of Activity Sequence Modeling

Monitoring and interpretation of sequential activity learning can improve adaptation and customization in the educational environment. This research can explore new semantically meaningful information to learners, to model the learner's problem modeling solution, to use the model at the target, ultimately to automate clustering we present an approach based on. Approach is applicable to different levels to detect resolution of problem of predefined style. To identify the problem of resolution style by analyzing learner's behavior according to the known dimension of learning, we will establish to discover how semi-automatic learning dimensions and concrete model of problem solving. In this report, the approach itself explains the feasibility of applying to the real world data and address aspects of the approach that can be adjusted to different learning contexts. Finally, Thesis reported the adaptation cycle to the appropriate intervention in the dialogue process, the incorporation of the proposed approach in the data collection system.

7. Tools and Techniques

Agent technology is considered one of the most innovative technologies for the development of decentralized software systems. Although this is not yet

the dominant approach in software engineering, a lot of research has been done and many applications have been developed and presented. Some software applications originally developed for research purposes are in the industry For example, JADE, Agent Builder, FIPA-OS, Jack and ZEUS. In this research we will review the most popular toolkits JADE, JACK and MASDK. Reviewing all of MAS toolkits is out of the scope of this research. A more complete list of more than 100 agent toolkits can be found in (Hamburg 2009) and at Agentlink.com.

8. Functional and Description

A functional model which is similar to activity model or process model is graphical representation of an enterprise's function within a defined scope. The purpose of functional model is to describe the functions and processes, assist with discovery of information needs, help identify opportunities, and establish a basis for determining product and service costs.

9. Flow Charts

A flowchart is a type of diagram that represents an algorithm or process, showing the steps as boxes of various kinds, and their order by connecting them with arrows. This diagrammatic representation illustrates a solution to a given problem. The flowchart in the Figure depicts the flow of the proposed system.

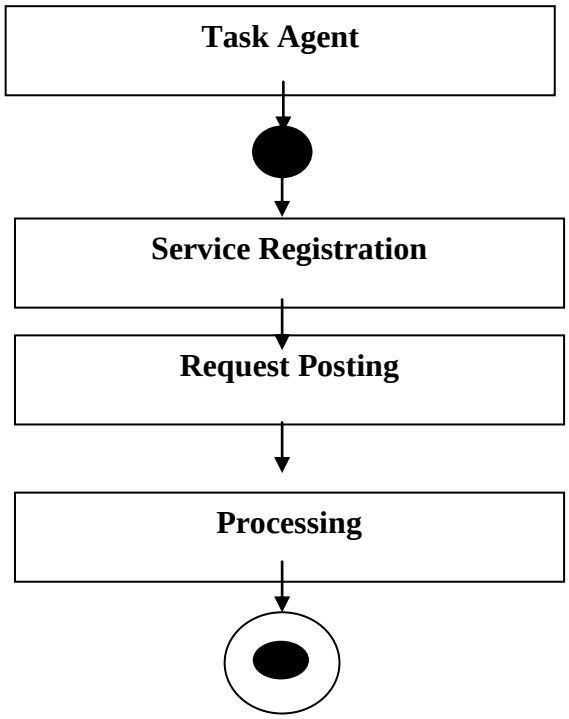

Figure 1: Flow Chart

10. Activity Diagram

Activity diagrams are graphical representations of workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control. Use cases show what your system should do. Activity diagrams allow you to specify how your system will accomplish its goals. Activity diagrams show high-level actions chained together to represent a process occurring in your system. An activity diagram is essentially a flowchart, showing owe of control from activity to activity. Unlike a traditional flowchart, an activity diagram shows concurrency as well as branches of control.

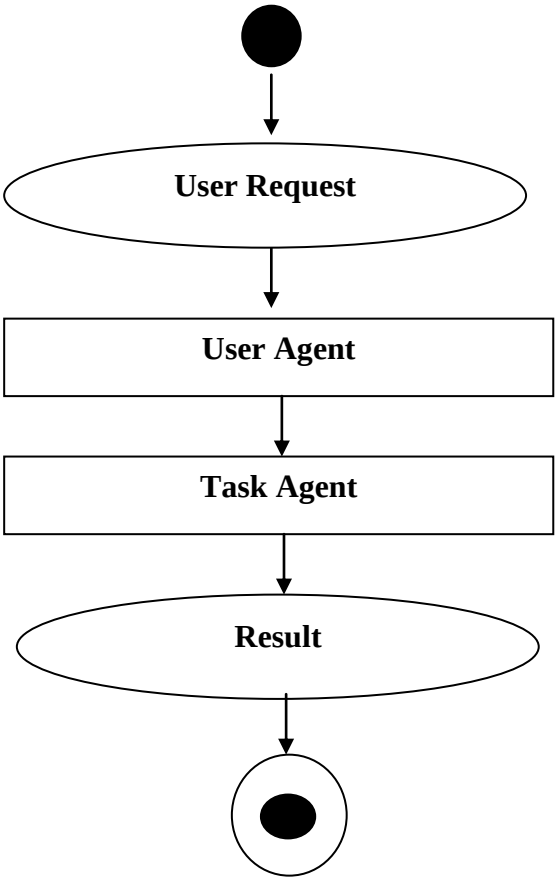

Figure 2: Activity Diagram

The result and discussion describe the impact on performance in terms of efficiency and accuracy. Preliminary results for overall Indian data of type II agents have been reported.

January-February 2020
ISSN: 0193 - 4120 Page No. 10924 - 10930

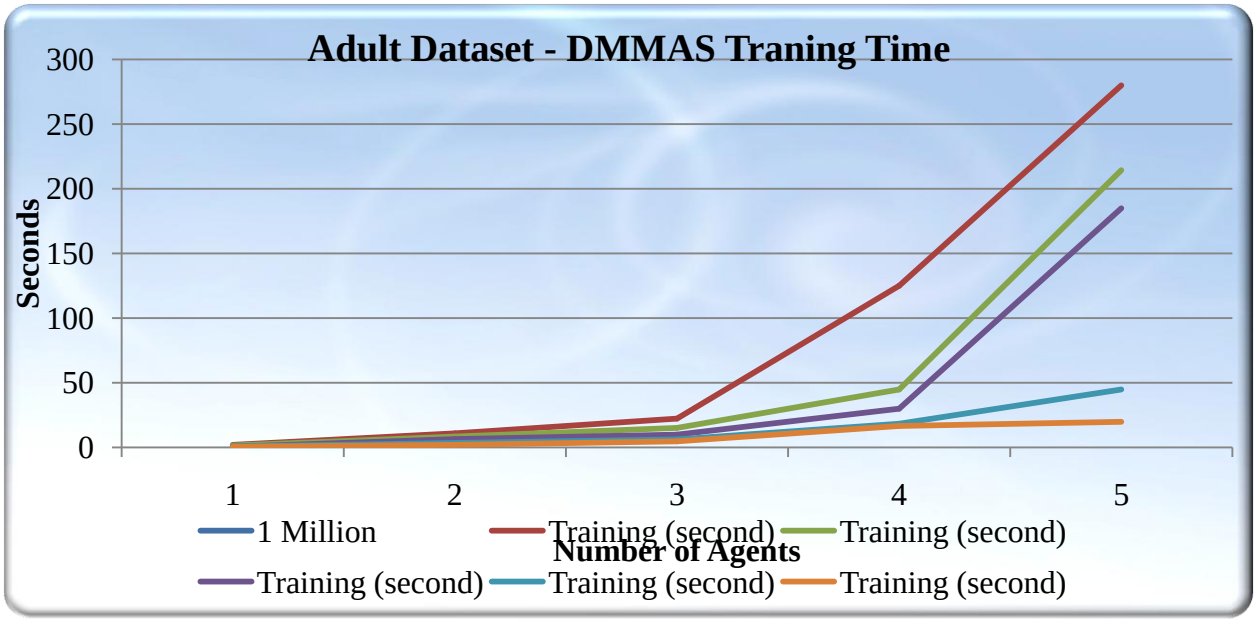

Figure 3: DMMAS Training Time

11. Agents Training

Training set up using the Training tab (As shown in Figure 4.10). The user has the function of specifying whether the test needs to be executed after the end of the training. Currently only providing test mechanism is to divide the ratio, divide the data into a set of training and testing. This means that the scores assigned by the agent 66.6% are reserved for training and the remaining 33.3% will be examined.

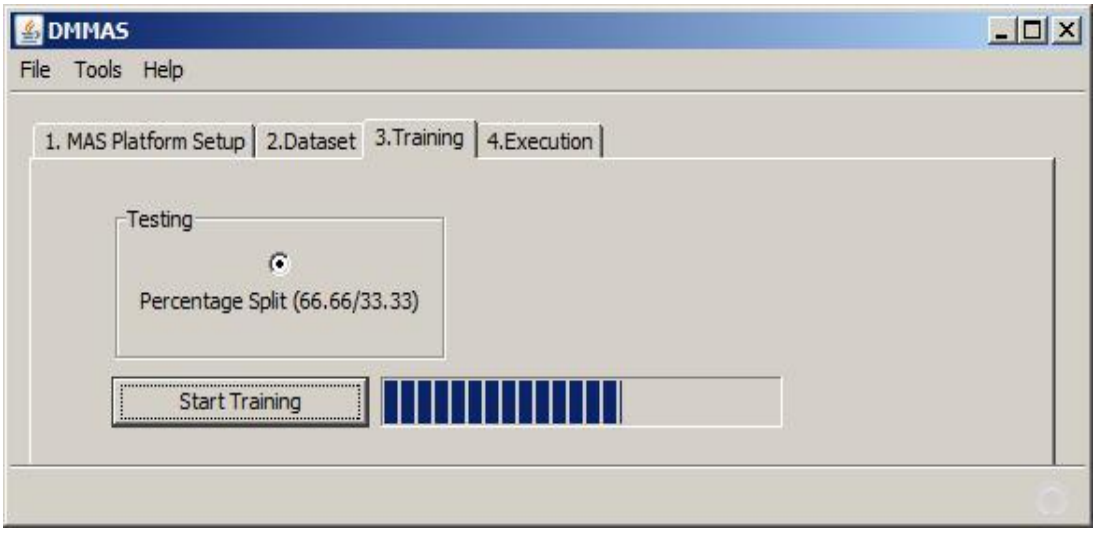

Figure 4: DMMAS Training Time

The system user interface allows the user to effectively form unequal clusters in WSN environment and route the data to the BS.

January-February 2020
ISSN: 0193 - 4120 Page No. 10924 - 10930

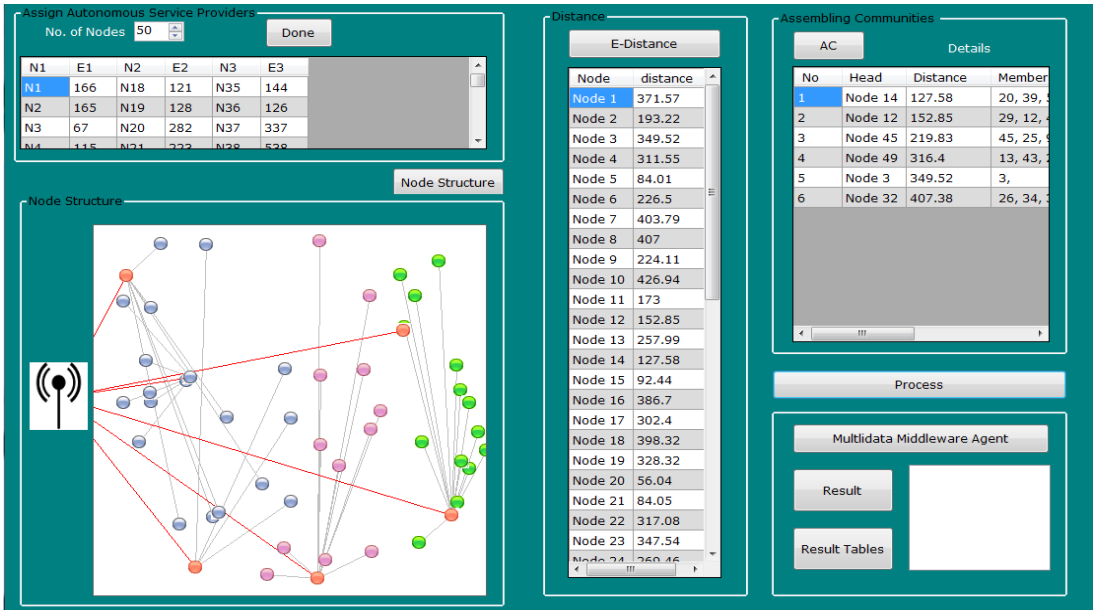

Figure 5: changing node GUI

In figure-5, there is number of nodes and it can change these nodes as per our requirement. Suppose it have multiple data agents and all are working in database and it can increase number of node from hare as a plus and minus. After this as every node has some weight or every page have some data for forward and send. When click on done button then energy/data assigned against every node. After assigned energy, it will be placed virtually on GUI. All nodes are placed graphically in node structure area. N1 has weight/energy 166 that is given in above figure. When it take mouse pointer over node in graphical area, it display x and y coordinate of particular node and same would be display on server. Now server and node(suppose n1), it have to make equally distance of both because when we want make multi data agent then equally distance is required between server and node.

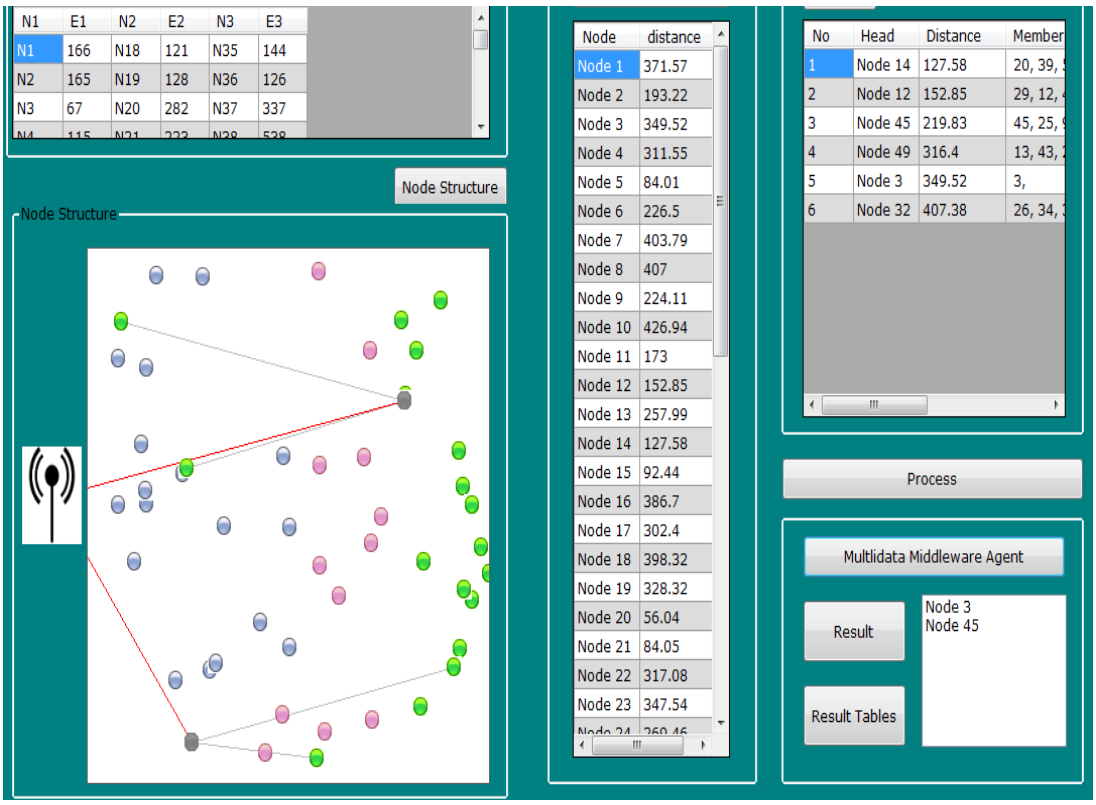

Figure 6: Multiple members against nodes

January-February 2020
ISSN: 0193 - 4120 Page No. 10924 - 10930

In figure-6, After make equally distance, as see node n1 have distance 351.57 means in graphical area node n1(x,y location) distance is 351.57 from server (x,y location). For measurement of this distance a formula was applied. After this, it go communication area, in this make group of node/agent or area and after make the agent of node, passes the data accordingly. As a head of node14 (in above given figure) and in this groups are created automatically. It can see that hare is multiple members and node 14 is the head of multiple members against of node12 and node 49. Likewise created groups and created data agents for everyone. All nodes have multiple members, that are called process, process means for data passes. It represented it on GUI, as node12 is looking in red color. There is multi file against node12 or multiple agents like 29,12,14,39, 8 etc.

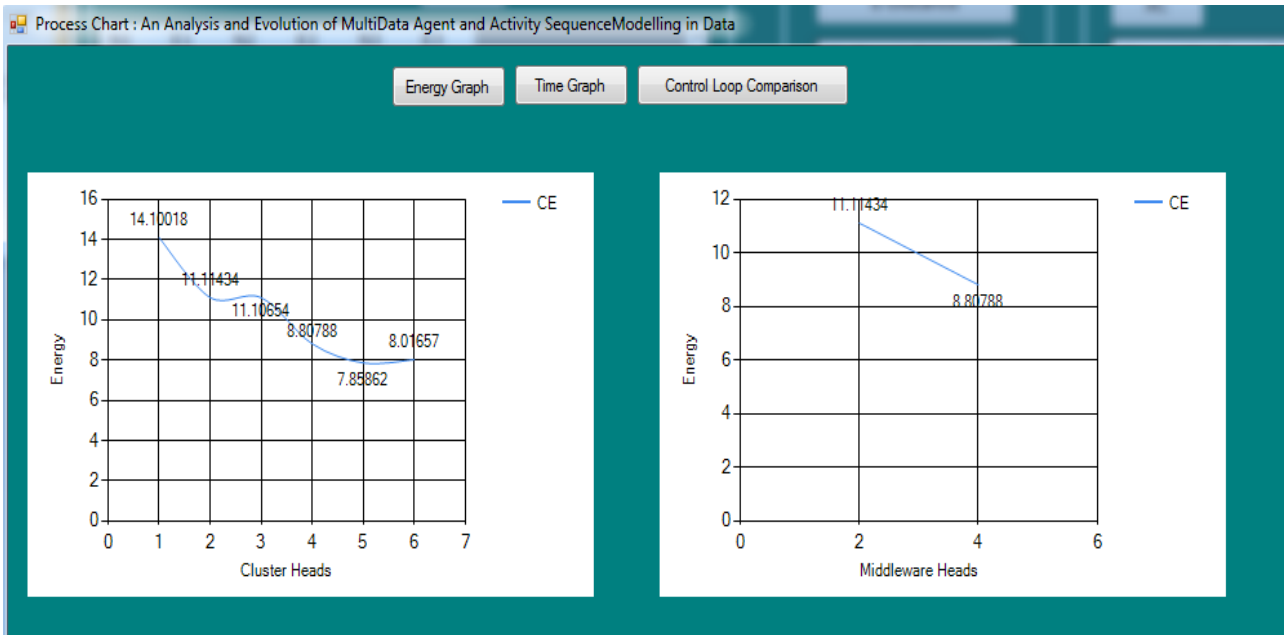

Figure 7: Multiple Cluster Head

In figure-7, these graphs displaying the multiple cluster head and all attached. Now it's available on existing system. Existing system generates multiple data agent and transfers the same to server individually. In previous, process where multiple middleware agents were present, it sorted out all agents and created a head from them. After that any agents that are sub head (Multi data agents) took data called super multi data agents.

This research deal with MADS, two levels, system architecture and component level to address the complexity of scalability and efficiency issues. In system architecture, emphasis is on system components and overall architecture. Assuming that the data mining system contains sites of multiple data, the unique resources, databases and agents MADS and protocols that allow each data site to work efficiently without interfering with it Support number progression. The use of an effective distributed protocol, however, only addresses the problem of scalability in part. The scalability of the system depends greatly on the efficiency of its components (agents).

Analysis of dependencies between agents and their effectiveness compared to DM work constitutes the other half of the scalability problem. With this principle rationale specific to the DM task, this problem is addressing MADS. For example, problem Meta ARM.

References

[1] 17th Australian Joint Conference on Artificial Intelligence, volume 3339, pages 917–922. Springer, 2004.

[2] Ajith, G. Crina editors. Swarm Intelligence in Data Mining, volume 34 of Studies in Computational Intelligence. Springer, 2006.

[3] Akyildiz, W. Su, Y. Sankara, and E. Cayirci. A survey on sensor net-works. In IEEE Communications Magazine, pages (102-114), 2002.

[4] Albashiri A *et al.*, "Data Partitioning and Association Rule Mining Using a Multi-Agent System", International Journal of Engineering Science and Innovative Technology (IJESIT) Volume 2, Issue 5, September 2013.

[5] B. Babcock, S. Babuand M. Datar, R. Motwani, and J. Widom. Models and Issues in Data Stream Systems. In Proceedings of the 21th Association for Computer Machinery (ACM) Special Interest Group on Management of Data (SIGMOD) Symposium on Principles of Database Systems (PODS), 2002.

[6] B. Chandrasekaran and T. Johnson. Generic tasks and task structures: His-tory, critique and

January-February 2020
ISSN: 0193 - 4120 Page No. 10924 - 10930

new directions. In Proceedings of the Second Generation Expert Systems, Springer - Verlag, Berlin, Germany, pages (232-272), 1993.

[7] B. Kegl and G. Lapalme. Performance Evaluation of an Agent Based Dis-tributed Data Mining System. In Proceedings of AI 2005, LNAI 3501, Springer-Verlag Berlin, Heidelberg, Germany, pages (25-32), 2005.

[8] Bansal A, "Comparison of K-means and Backpropagation Data Mining Algorithms", International Journal of Computer Technology and Electronics Engineering (IJCTEE) ISSN 2249-6343 , Volume 2, Issue 2, PP:151-155.

[9] Bee-gent: Bonding and Encapsulation Enhancement Agent Framework for Development of Distributed Systems. In Proceedings of the 6th Asia-Pacific Software Engneering Conference, 1999.

[10] Bhardwaj D., "Rise of Data Mining: Current and Future Application Areas", IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 5, No 1, September 2011 ISSN (Online): 1694-0814, PP:256-260.

[11] Bordetsky *et al.* Agent-based Support for Collaborative Data Mining in Systems Management. In Proceedings Of The Annual Hawaii International Conference On System Sciences, page 68, 2001.

[12] Burnett M. *et al.*, "Mining Problem-Solving Strategies From HCI Data", Oregon State University.

[13] C. Baumer and T. Magedanz. Grass Hopper 2, an intelligent mobile agent platform written in 100% pure Java. In Proceedings of the 3rd Interna-tional Workshop on Intelligent Agents for Telecommunication Applications (IATA), Springer LNAI 1699, Berlin, Germany, pages (19-32), 1999

[14] C. Blake and C. Merz. UCI repository of machine learning databases. Irvine, CA: University of California, Department of Information and Computer Science, 1998. http://archive.ics.uci.edu/ml/.

[15] Chan P.K, "Real Time Data Mining-based Intrusion Detection" research article, Computer Science Department, Columbia University, New York, NY 10027.

[16] Cho Ju S. *et al.* Agent based distributed data mining. Lecture Notes in Computer Science, 3320:42–45, 2004.

[17] D. Cheung and Y. Xiao. Effect of Data Distribution in Parallel Mining of Associations. In Proceedings of the Data Mining and Knowledge Discovery 3(3), pages (291-314), 1999

[18] D. Hand, H. Mannila, and P. Smyth. Principals of Data Mining. MIT press, Cambridge, Mass, 2001.

[19] D. Kerr, D. O'Sullivan, R. Evans, R. Richardson, and F. Somers. Ex-periences using Intelligent Agent Technologies as a Unifying Approach to Network and Service Management. In Proceedings of the Intelligence in Ser-vices and Networks: Technology for Ubiquitous Telecom Services, Antwerp, Belgium, 1430, pages (115-126), 1998.

[20] Dr. Jain Y.K., "Improved Data mining approach to find Frequent Item set Using Support count table", International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume 1, Issue 2, July – August 2012, ISSN 2278-6856 PP:195-201.

[21] Dr. Jain Y. Kumar *et al.*, " Efficient Retrieval of Text for Biomedical Domain using Data Mining Algorithm", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 4, 2011, PP:77-80.