Research on Multi-Agent Experiment in Clustering

Mohammed Ali Shaik, T. Sampath Kumar, P. Praveen, R. Vijayaprakash

Abstract*—Clustering is the process of classification of data in any of the emerging fields in present day scenario. In this paper we have proposed a framework for "multi-agent based clustering (MAS)" which is based on the working methodology of individual agents that initiate to dissolve a cluster. The major identification in this paper is once a cluster is created by an agent distinct types of agents will communicate and send messages for performing a task given to them, this process continues till the dissolving of cluster. An agent has the capability to either create a cluster or kill the cluster. And agents can even perform Inter cluster communication (ICC) between various clusters that are created as per requirement or necessity. We used K-means and KNN algorithms and shown that ICC can improve a clustering environment.

Keywords*—Multi-Agent Data Mining, Clustering, Inter cluster communication.

I. INTRODUCTION

In the field of computers "Data Mining and Multi-Agent Systems (MAS)" are emerging technologies as almost all the present and future applications need data to be classified and provided on demand. One of the major issue in data mining is the ever generating of data in the forms of streams, this data stream is to be stored then analyzed and presented as a report for performing decision making. And in order to do this multi agent architecture is the only best option we have, because a agent also has a capability to create another agent, if more number of agents are created and controlled then data can be processed easily and accurately[2]. MAS provides greater computational power[5].

The major approach to do this is by implementing distributed [6] or parallel [13] processing such that the problem can be addressed effectively to resolve the problem in a specific manner where the course assumes to be appropriate when cloud comprises of effective and latest processors to process the on demand data. On the same hand at a similar level the disadvantage is it posses the centralized control that tends to lack generality. The distributed and parallel data mining techniques are assumed to be a master process that allows to perform data mining task effectively as the control is centralized to perform the master process as the existing system lacks robustness at possible levels.

Another specific issue is we cannot generalize the system that is implemented in either distributed or parallel system as the approach that is used to perform a specific data mining task follows a generalized process to perform the assigned task. This issue is handled very effectively by MAS as the process implemented in MAS is centralized whose basic feature is to control the systems effectively. Due to this feature the system is considered to be more robust when implemented.

An Agent is a software process or a program or a software entity which is capable of communicating with others to perform a specific task by performing communication with software entities (agents) that are intended to cooperate in some or the other manner so as to undertake some sort of processing task as all the agents perform tasks in a request/notify manner.

"Multi-agent Data Mining (MADM) or Agent Assisted Data Mining (AADM)" allows to distribute data over data warehouse that effectively mine the data, as MADAM also supports the creation of frameworks that allows agents to communicate in framework using set of communication protocols designed on specific basis. Agents communicate using "Agent Communication Languages (ACLs)" over various applications.

In this paper we describes the MADM framework where agents communicate using various protocols that directs unsupervised learning in clustering. And the proposed framework comprises of four general classification of agents: 1) user agents 2)data agents 3)clustering agents 4)validation agents 5) housekeeping agents. These agents communicate with one another to perform a task based on request-notify basis.

II. PREVIOUS WORK

This section represents a review that is being conducted to relate the work done by most of the researchers who have done the research somewhat similar to MADM framework and then tries to continue with a brief review of distributed and parallel work on multi agent based clustering techniques that highlight or identifies the difference between proposed MAS clustering approach and existing work.

Majorly there exists two main paradigms for performing interaction and to integrate the transmission of data between agents to perform data mining [4]. Primarily data mining derives agents that support various abilities for performing mining of agents to perform adaptation or coordination or learning or to perform reasoning and secondly in agent driven data mining approach which is also known as the

Revised Manuscript Received on 11 February, 2019.

Mohammed Ali Shaik, PhD Research Scholor, APJ Abdul Kalam University, Indore, (M.P.), India (niharali@gmail.com), Assistant Professor, S R Engineering College, Department of Computer Science & Engineering

T. Sampath Kumar, Assistant Professor, S R Engineering College, Department of Computer Science & Engineering

Dr. P. Praveen, Assistant Professor, S R Engineering College, Department of Computer Science & Engineering

Dr. R. Vijayaprakash, Professor, S R Engineering College, Department of Computer Science & Engineering

"Multi-Agent Data Mining (MADM)" is specifically used to collect data from various agents to perform data mining tasks by using the surveys of agent-based distributed data mining as represented in PADMA [14] and PAPYRUS [2] where both reported multi-agent clustering systems.

These systems aims to achieve the integration process of knowledge discovered from different sites with a minimum amount of network that tends to communicate with the maximum amount of local processing of data using PADMA [14] as it facilitates the coordinator agent and forces it to perform the direct interaction between various mining agents which and clusters that are based on a centralized architecture which is used to access the local data for performing analysis and to generate the global clusters.

PADMA [14] has proposed the generation process of various hierarchical clusters in the context of document categorization that is formed over PADMA in a distributed clustering system PAPYRUS [12] has adopted a Peer-to-Peer model that tends to transmit the data between all the existing agents that are based on the MAS strategies. "Multi-agent clustering system is KDEC [16]" is a purely based on the "distributed density-based clustering algorithms" that are implemented in a Peer-to-Peer model to estimate KDEC density through the sample data that is transmitted or moved instead of actual data values. This process preserve data privacy in order to minimize communication gap between various sites or web sources.

In a multi-agent clustering system various documents will serve or propose the objective to improve accuracy and relevancy to retrieve information from the cluster through clustering agents. These agents deal with distinct data streams for distributing dynamic environments for obtaining the input data sets and to measure decision criteria which can be changed at runtime as the clustering results are available at any hook of time and are continuously revised as per the requirements in order to achieve the global clustering methodology.

III. THE PROPOSED FRAMEWORK (AGENT-MADM)

The proposed framework comprises of five distinct categories of agents:

User agents:

Primary one is the agents are the interface between distinct types of users such as end users in a MADM environment where a user agent is responsible for getting the input from the user then generating or planning for creation process of clusters by clustering agents based on results attained.

Data agents:

Data agents are also called as the owners of various data sources using peer to peer relation within data sources and their data agents where the data agents will be only allowed to access data in MADM framework.

Clustering agents:

Clustering agents are also called as the owners of clusters or cluster groups where every agent will represent a cluster that is generated by the clustering algorithm as clustering agents are configured to do so. Each clustering agent will select the data records from data sets based on the requirement of user agents.

Validation agents:

Validation agents are desired to perform validation on the data generated by data agents for performing various operations on the results obtained by clusters or cluster groups. These validations are the rules imposed either by user or by the environment.

Housekeeping agents:

For performing the clustering task all the existing derived clustering results can be added through MADM agents as housekeeping.

The proposed Agent-MADM framework is represented in the figure1 that includes various types of agents such as a User Agent then a collection agent then a Clustering Agent and then a Data Agent then a Validation Agent and some housekeeping agent these agents communicate in the framework to complete a task given by an user which is represented by directed edges or lines as the communication process can be bidirectional or even some times unidirectional where a data agent is required to communicate with all the other agents as per the requirement.

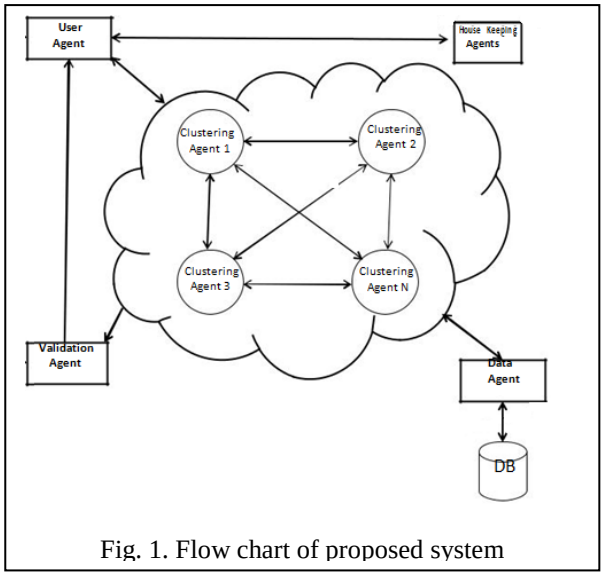

Fig. 1. Flow chart of proposed system

The proposed framework is being implemented using "PHP and Java Agent Development Environment (JADE) along with the AMS (Agent Management System) agent and the DF (Directory Facilitator) agent", where the AMS is the agent which is responsible for controlling or implementing the agent lifecycle.

IV. AGENT COMMUNICATION PROCESS WITHIN THE FRAMEWORK

The agents within has to communicate with one another in the proposed framework where JADE provides a API for performing communication mechanism among distinct

agents by making use of FIPA ACL [9] programming language, though it has limited applicability when performing validation. This issue is overcome in our proposed framework by generation notifications and allowing agents to perform dialogues that tend to exchange or communicate agents for obtaining clusters or group of clusters. In our proposed framework agents communicate by sending dialogue for creating of clusters, dividing clusters, merging clusters, destroying clusters and so on.

The validating agent must meet all the preconditions that are prewritten in the framework protocols section must be automatically verified as a part of protocol that allows exchange of information between various types of agents based on their semantics related to programming communication language.

A housekeeping agent opens dialogue that consists of a request generated by any of the user agents for performing mining of data based on the biz-rules where either clusters are partitioned or joined or merged based on the dialogue generated by validation agent that avoid cohesion and performs coupling of clusters by performing accept view or by reject view based on the cloud configuration to handle the mining task with respect to processing and storage capacities. Otherwise the cloud has to be reconfigured as per the requirements specification of MADM architecture.

Clustering Agents are invoked by User Agent based on the clustering request either by suing K-means strategy or the KNN strategy, where the K-means strategy specifies allowable number of clusters represented by constant value K in case of K Clustering Agents that are decided by MADM architecture in two phase format: 1) bidding phase and 2) Refinement phase. Where the Clustering Agents compete by performing bidding of records to perform auction phase as per the prewritten protocol: “Data Agent act”.

During refinement phase each Clustering Agents may sometimes even pass unwanted records that are no longer required in the cluster may also exist. These clusters are to be removed by performing auction by each of the Clustering Agents based on the protocol act implemented by a local auctioneer who either donates or sells the unwanted data to other clustering agents, this phase is called as the refinement phase.

Algorithm: Bidding of K-means Spawning strategy

Step 1: Start

Step 2: User Agent scans K Clustering Agents {f:C1, C2,…., Ck }

Step 3: Each of the Clustering Agent sends a data request to Data Agent.

Step 4: Data Agent sends K records {D1, D2,….., Dk} to the K Clustering Agents{C1, C2,…}.

Step 5: Cluster centroid is calculated by each Clustering Agents.

Step 6: di = D (i = K + 1 to n)

Step 7: cj = C (j = 1 to K)

Step 8: bidDistance = di * centroid(cj);

Step 9: allocate minDistance(di, cj)

Step 10: Stop.

Biding Phase founded on the K-means Spawning Strategy

The operation that is performed for implementing bidding process using K-means strategy that is presented in the above algorithm for scanning clustering agents to represent the clusters where each Clustering Agent is allocated for retrieving a single record for identifying Data Agent for representing the centroid of each cluster. The nearest neighbor defines the distance of cluster where the cloud is configured based on K records that are selected in initial clusters.

Biding Phase founded on the KNN Spawning Strategy

The biding phase is implemented using the KNN scanning strategy that commences with a single Clustering Agent represented by Ci. where the bidding operation is performed by identifying the nearest neighbor based on its threshold value. If the generated threshold value is lesser than the bidding value the data record is stored in the nearest cluster and if the threshold value is greater than the bidding value it is stored in the distance based cluster.

Negotiation Phase

Based on the generated cohesion values negotiation or refinement process is implemented where individual and overall cohesion values of a cluster are compared with the configuration separation value, a acceptable scenario is separation value must be more than that of the cohesion value. Then it is said to be acceptable and it is performed on each and every agent available in the system based on the centroid value.

Cluster Configuration phase

The cluster configuration phase generates the cluster configuration matrix. This matrix is generated based on the bidding value based on the existence of record, or existence of non required data record or the required which is missing in the cluster are represented in the form of matrix as shown in below figure 2.

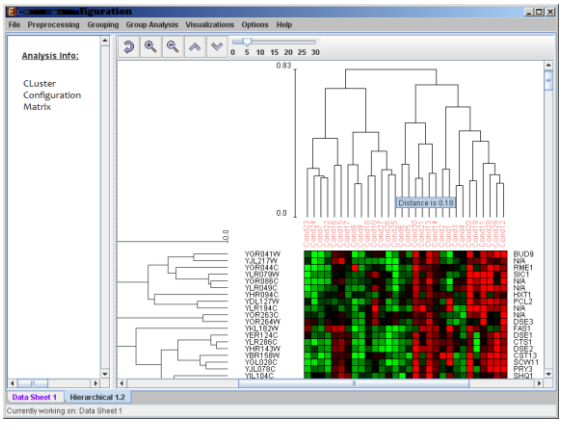

Fig. 2. cluster configuration matrix

The bidding of agents is represented in above matrix based on the attained cluster values that are implemented by validation agent and adopted by user agent. metric must be

adopted. Matrix also represent the resource utilized that is measured by normalizing the data to represent the goodness or level of non-goodness in a cluster data or the data being made available in the cluster.

V. RESULTS

The proposed MADM architecture id implemented and the results attained are based on the inputs provided with variant inputs are:

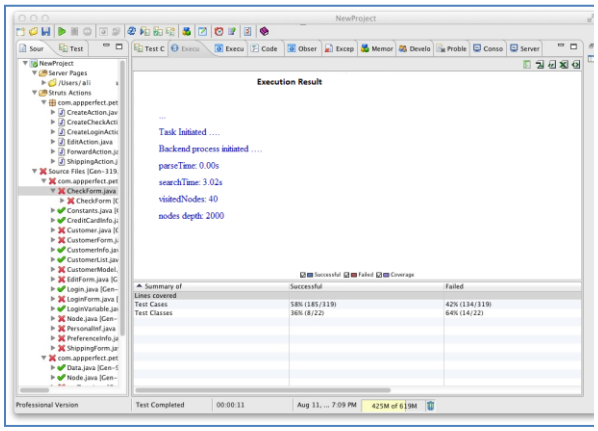

Fig. 3. Evaluation results attained

The figure comprises of clusters and sub clusters data that will scan the data from clusters or cluster groups using MADM architecture.

VI. CONCLUSION

A MADM framework is proposed in this paper to achieve multi-agent based clustering where the main feature of the framework is that it allows agents to negotiate or communicate for improving the initial clustering. The framework can implemented effectively in almost all the areas in possible number of ways. As our proposed methodology uses K-means and KNN algorithms. And the results are generated using a number of datasets and clusters that are taken.

REFERENCES

1. Agogino, A., Tumer, K.: E cient agent-based cluster ensembles. In: Proc. 5th Int. Conf. on Autonomous agents and multiagent systems. pp. 1079{1086. AAMAS '06, ACM, New York, NY, USA (2006)
2. Bailey, S., Grossman, R., Sivakumar, H., Turinsky, A.: Papyrus: A system for data mining over local and wide area clusters and super-clusters. IEEE Supercomputing (1999)
3. Bellifemine, F., Bergenti, F., Caire, G., Poggi, A.: JADE: a java agent development framework. In: Bordini, R.H. (ed.) Multi-agent programming: languages,platforms, and applications, p. 295. New York: Springer (2005)
4. Cao, L., Gorodetsky, V., Mitkas, P.A.: Guest editors' introduction: Agents and data mining. IEEE Intelligent Systems 24(3), 14{15 (2009)
5. Chaimontree, S., Atkinson, K., Coenen, F.: Best clustering con guration metrics: Towards multiagent based clustering. In: Proc 6th Int. Conf. Advanced Data Min-ing and Applications (ADMA'10). pp. 48{59. Springer LNAI 6440 (2010)
6. Coenen, F., Leng, P., Ahmed, S.: T-trees, vertical partitioning and distributed association rule mining. In: Proc. 3rd IEEE Int. Conf. on Data Mining. pp. 513{ 516. ICDM '03, IEEE Computer Society, Washington, DC, USA (2003)
7. Dasarathy, B.V.: Nearest neighbor (NN) norms: NN pattern classi cation tech-niques. IEEE Computer Society Press, Las Alamitos, California (1991)
8. P. Praveen, B. Rama and T. Sampath Kumar, "An efficient clustering algorithm of minimum Spanning Tree," 2017 Third International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), Chennai, 2017, pp. 131-135.doi: 10.1109/AEEICB.2017.7972398
9. FIPA: Communicative Act Library Speci cation. Tech. Rep. XC00037H, Founda-tion for Intelligent Physical Agents (2001), http://www.fipa.org
10. Forman, G., Zhang, B.: Distributed data clustering can be e cient and exact. ACM SIGKDD Explorations Newsletter 2, 34{38 (2000)
11. Frank, A., Asuncion, A.: UCI machine learning repository (2010), http:// archive.ics.uci.edu/ml
12. Giannella, C., Bhargava, R., Kargupta, H.: Multi-agent systems and distributed data mining. In: Klusch, M., Ossowski, S., Kashyap, V., Unland, R. (eds.) Cooper-ative Information Agents VIII. Lecture Notes in Computer Science, vol. 3191, pp. 1{15. Springer Berlin / Heidelberg (2004)
13. Praveen P., Rama B. (2018) A Novel Approach to Improve the Performance of Divisive Clustering-BST. In: Satapathy S., Bhateja V., Raju K., Janakiramaiah B. (eds) Data Engineering and Intelligent Computing. Advances in Intelligent Systems and Computing, vol 542. Springer, Singapore
14. P.Praveen, B.Rama, "An Efficient Smart Search Using R Tree on Spatial Data", Journal of Advanced Research in Dynamical and Control Systems, Issue 4,ISSN:1943-023x.
15. Kiselev, I., Alhajj, R.: A self-organizing multi-agent system for online unsupervise learning in complex dynamic environments. In: Proc. 23rd AAAI Conference on Arti cial Intelligence. pp. 1808{1809. AAAI Press (2008)
16. Klusch, M., Lodi, S., Moro, G.: Agent-based distributed data mining: The KDE: scheme. In: Lecture Notes in Arti cial Intelligence (Subseries of Lecture Notes in Computer Science). vol. 2586, pp. 104{122 (2003)
17. Klusch, M., Lodi, S., Moro, G.: The role of agents in distributed data mining: Issues and bene ts. In: IAT '03: Proceedings of the IEEE/WIC International Conference on Intelligent Agent Technology. p. 211. IEEE Computer Society, Washington, DC, USA (2003)
18. MacQueen, J.B.: Some methods for classi cation and analysis of multivariate observations. In: Proc. 5th Berkeley Symposium on Mathematical Statistics and Prob-ability. pp. 281{297 (1967)
19. McBurney, P., Parsons, S., Wooldridge, M.: Desiderata for agent argumentation protocols. In: Castelfranchi, C., Johnson, W.L. (eds.) Proc. 1st Int. Conf. on Au-tonomous Agents and Multi-Agent Systems (AAMAS 2002). pp. 402{409. ACM Press: New York, USA, Bologna, Italy (2002)
20. P. Praveen and B. Rama, "An empirical comparison of Clustering using hierarchical methods and K-means," 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB), Chennai, 2016, pp. 445-449. doi: 10.1109/AEEICB.2016.7538328

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication