IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-6, Issue-9, Sept. 2018 E-ISSN: 2347-2693

A Brief Survey On Ant Based Clustering for Distributed Databases

Dhivya. N[1*] , Sumangala. K[2]

[1]Dept. of Computer Science, Kongunadu Arts and Science Colleg , Coimbatore, India
[2]Dept. of Computer Science (PG), Kongunadu Arts and Science College, Coimbatore, India

Available online at: www.ijcseonline.org

Accepted: 19/Sept/2018, Published: 30/Sept/2018

***Abstract*—** Clustering is a separation of data into collections of parallel objects. Signifying the data by smaller amount of clusters automatically loses certain fine details, but attains simplification. It models data by its clusters. This paper aims to present a brief survey and comparative study on and based clustering theory based on distributed databases in which the goal is to minimize the amount of iterations and cluster sizes is needed to re-optimize the solution when the cluster changes. Number of relative studies namely hybrid, density, Pheromone based ant clustering and cluster analysis. To conclude the discussion, the ant based clustering algorithms are discussed and evaluate the processing time performance on the several distributed datasets. Comparing to these algorithms the efficient Ant based Multiple Pheromone techniques methods outperforms having better performance than other methods.

***Keywords*—** Clustering, partitioning, data mining, Ant clustering, Particle Swarm Optimization.

I. INTRODUCTION

Data mining is the process of finding previously an unknown patterns and trends in databases and using that information to build predictive models. Data mining combines statistical analysis, machine learning and database technology to extract hidden patterns and relationships from large databases.

Clustering is a significant area in data mining. The purpose of clustering is to cluster a great deal of objects into groups according to the rule that is "things of one kind come together". The similarity of objects in a cluster should be as big as possible while the difference of objects in different clusters should be as big as possible too. Clustering has been applied in many domains, such as spatial data analysis, image processing, marketing and pattern recognition etc. Wildly used clustering algorithms are K-means, density based DBSCAN and network based STING etc [2, 3]. The ant based clustering algorithm, which is inspired by the collective behavior of real ants can avoid some shortages like, there is no need to determine the number of clusters artificially in advance in ABC. However, in general algorithms, users have to set the number as an input parameter. This requires rigorous domain knowledge for users, and on the other hand, the final clustering result is every sensitive to the initial number of clusters set in advance.

Dorigo et al. [11-14] have formalized ant colony optimization into a combinatorial optimization metaheuristic. Given a COP, the first step for the application of ACO to its solution consists in defining an adequate model. This is used to define the central component of ACO: the pheromone model. The model of a COP is defined in form of a triplet (P, Ω, f) as follows:

- P is a search space defined over a finite set of discrete decision variables;
- Ω is a set of constraints among the variables;
- $f : P \rightarrow R_0^+$ is an objective function to be minimized

The search space P is defined as follows: Given is a set of discrete variables X_i, $i = 1....n$, with values $v_i^j \in D_j$ $\{v_i^1, v_i^2 \ldots v_i^{|D|}\}$. A solution p ∈ P is a complete assignment in which each decision variable has a value assigned that satisfies all the constraints in the set Ω, is a feasible solution of the given COP.

Ant-based clustering algorithms are based upon the brood sorting behavior of ants. Larval sorting and corpse cleaning by ant was first modeled by [1]. for accomplishing certain tasks in robotics. Their work was actually focused on clustering objects by using group of real world robots. Their model is known as basic model (BM). This model can be described as follows: The data items are randomly scattered into a two-dimensional grid. Initially, each data object that represents a multi-dimensional pattern is randomly distributed over the 2D space. Each ant moves randomly around this grid picking and dropping the data items. The decision to pick up or drop an item is random but is influenced by the data items in the ant's immediate neighborhood. The probability of dropping an item is increased if ants are surrounded with similar data in the neighborhood. In contrast, the probability

of picking an item is increased if a data item is surrounded by dissimilar data, or when there is no data in its neighborhood.

There are so many existing studies are focused on clustering based framework in distributed framework. Many clustering algorithms have been proposed in literature. But due to changing domain of applications, clustering is still a critical research issue. Swarm Intelligence is the property of a system whereby the collective behaviors of (unsophisticated) agents interacting locally with their environment cause coherent functional global patterns to emerge. It provides a basis with which it is possible to explore collective problem solving locally without provision of a global model. The unsupervised nature of clustering makes an analogy with nature inspired metaheuristics which work on self-organization principle without central control. The self-organization guides clustering to organize objects in some meaningful manner and decentralized control provide unsupervised learning.

This survey paper explores the Probabilistic Ant based Clustering for Distributed Databases to process sequences of distributed databases facilitates learning of new concepts that characterize common features of, and differences between, datasets.

II. RELATED WORK

***J. Handl and B. Meyer* (2002)** [4] discussed a sorting and clustering methods inspired by the behavior of real ants are among the earliest methods in ant-based meta-heuristics. They revisited these methods in the context of a concrete application and introduce some modifications that yield significant improvements in terms of both quality and efficiency. Firstly, re-examine their capability to simultaneously perform a combination of clustering and multi-dimensional scaling. In contrast to the assumptions made in earlier literature, the results suggest that these algorithms perform scaling only to a very limited degree. They showed how to improve on this by some modifications of the algorithm and hybridization with a simple pre-processing phase. Secondly, to discuss how the time-complexity of these algorithms can be improved. To further analyze the performance, to measure Pearson correlation on several levels: (1) Overall correlation of all elements; (2) Inter-cluster correlations, where the weighted average of all cluster elements is used as cluster center; (3) Intra-cluster correlations, which are observed by computing the correlations for the clusters individually.

***J. Handl, J. Knowles, and M. Dorigo* (2003)** [5] applied variously, from problems arising in commerce, to circuit design, to text-mining, all with some promise. However, although early results were broadly encouraging, there has been very limited analytical evaluation of ant based clustering. Toward this end, they first proposed a scheme that enables unbiased interpretation of the clustering solutions obtained, and then use this to conduct a full evaluation of the algorithm. To analysis uses three sets each of real and artificial data, and four distinct analytical measures. These results are compared with those obtained using established clustering techniques and to find evidence that ant-based clustering is a robust and viable alternative.

***I. El-Feghi, M. Errateeb, M. Ahmadi and M.A. Sid-Ahmed* (2009)** [6] presented a new Adaptive Ant-based Clustering Algorithm (AACA) for clustering data sets. The algorithm takes into account the properties of aggregation pheromone and perception of the environment together with other modifications to the standard parameters that improves its convergence. The performance of AACA is studied and compared to other methods using various patterns and data sets. It is also compared to standard clustering using a set of analytical evaluation functions and a range of synthetic and real data collection. Experimental results have shown that the proposed modifications improve the performance of ant-colony clustering algorithm in term of quality and run time.

***J.B. Brown and M. Huber* (2010)** [7] addressed the issues most ant-based clustering approaches suffer from inefficiencies due to large numbers of unproductive ant movements and inefficient cluster merging, leading them to produce too many clusters and to converge too slowly. So authors presented a new ant-based clustering algorithm in which ants are organized in a loose two-level hierarchy with worker ants maintaining movement zone boundaries around each cluster and organizing its internal structure while a single queen ant in each cluster is responsible for moving items between clusters by directly handing them to other queens. This provides an infrastructure that avoids excessive ant movements between cluster regions while allowing for efficient long distance cluster merging. Comparison of this approach with traditional ant-based clustering shows its promise to significantly improve performance and scalability.

***L. Li, W-C Wu and Q-M Rong* (2010)** [8] presented a hybrid clustering algorithm based on density and ant colony algorithm, that to determine the initial cluster centers according to cluster objects distribution density method, and then use the swarm intelligence and randomness of ant colony algorithm to find that arbitrary shape of clusters, to avoid falling into local convergence, to get a relatively stable global optimal solution. Theoretical analysis and experimental results show that the improved algorithm can achieve better clustering results.

***L.M. Li and M-M Shen* (2010)** [9] To solve the problems of the excessive clustering time consumption and the redundant numbers of the resulting clusters, commonly encountered with the ant-based clustering algorithms, an improved ant colony clustering algorithm based on dynamic neighborhood

is proposed. The algorithm seeks for pure neighborhoods by performing auto adaptive adjustments of dynamic neighborhood, and enhances ant's memory by additionally storing the sizes of the pure neighborhoods. The ant can exchange information with other ants, load multiple similar objects at once, and merge the similar neighborhoods to form the final clusters efficiently.

S. Rana, S. Jasola and R. Kumar **(2011)** [10] discussed a data clustering is one of the most popular techniques in data mining. It is a method of grouping data into clusters, in which each cluster must have data of great similarity and high dissimilarity with other cluster data. The most popular clustering algorithm K-mean and other classical algorithms suffer from disadvantages of initial centroid selection, local optima, low convergence rate problem etc. Particle Swarm Optimization (PSO) is a population based globalized search algorithm that mimics the capability (cognitive and social behavior) of swarms. PSO produces better results in complicated and multi-peak problems. They presented a literature survey on the PSO application in data clustering. PSO variants are also described. An attempt is made to provide a guide for the researchers who are working in the area of PSO and data clustering.

Saroj Bala, S.I. Ahson and R.P. Agarwal **(2012)** [12] illustrated the swarm intelligence is a collective effort of simple agents working locally but resulting in wonderful patterns. Labor division, decentralized control, stigmergy and self organization are the major components of swarm intelligence. Ant based clustering is inspired by brood sorting in ant colonies, an example of decentralized and self organized work. Stigmergy in ant colonies is via a chemical pheromone. The authors proposed a pheromone based ant clustering algorithm. In that method, ant agents use pheromone to communicate the place to make the clusters, not the path as in real ants. The pheromone concentration helps to decide pick and drop of objects with less parameters and calculations.

Saroj Bala, S.I. Ahson and R.P. Agarwal **(2012)** [13] discussed grouping different objects possessing inherent similarities in clusters have been addressed as the clustering problem among researchers. The development of new metaheuristics has given another direction to data clustering research. Swarm intelligence technique using ant colony optimization provides clustering solutions based on brood sorting. After basic ant model of clustering, number of improvements has been proposed. But the ant clustering still suffers with low convergence. So the authors presented a novel model of intelligent movement of ants including the negative pheromone and direction selection. Negative pheromone plays a role of barrier in the direction of empty area and direction selection avoids the calculations not contributing to the clustering process. Simulations have shown good results.

Saroj Bala, S.I. Ahson and R.P. Agarwal **(2012)** [14] discussed Clustering is a data mining technique for the analysis of data in various areas such as pattern recognition, image processing, information science, bioinformatics etc. Hierarchical clustering techniques form the clusters based on top-down and bottom-up approaches. Hierarchical agglomerative clustering is a bottom-up clustering method. Ant based clustering methods form clusters by picking and dropping the objects according to surroundings. The authors proposed an agglomerative clustering algorithm, AGG_ANTS based on ant colonies. AGG_ANTS clusters the objects by moving ants on the grid and merging their loads according to similarity resulting in bigger clusters. It avoids the calculation of similarity in the surrounding and pick/drop of objects again and again resulting in a more efficient algorithm.

J. Chircop and C.D. Buckingham **(2013)** [15] discussed the Ant colony optimisation algorithms (ACO) work via a process called stigmergy in which ants deposit pheromone traces in order to influence foraging patterns. Pheromone traces are picked up and followed by other ants but they evaporate over time. Paths with more pheromone will survive longer and have a higher chance of getting followed and reinforced whilst weaker traces simply fade away. The premise behind the proposed Multiple Pheromone Algorithm for Cluster Analysis (MPACA) is that ants detect individual features of objects in space and deposit pheromone traces that guide towards these features. Each ant starts off by looking for a particular feature but they can combine with ants looking for other features if the match of their paths is above a given threshold. This enables ants to detect and deposit pheromone corresponding to feature combinations and provides the colony with more powerful cluster analysis and classification tools. The basic elements of MPACA are that: (i) at the start of the learning process, every object has at least one ant assigned to it for each feature; (ii) each ant searches for other objects with a matching feature value; (iii) a pheromone is laid down whenever an ant has found an object with a matching feature; (iv) if ants detecting different features find their paths are matching above a certain level, they will combine and start looking for the conjunction of features; and (v) ants become members of the same colony when the population density of ants in the area is above a threshold value.

III. COMPARISON ANALYSIS

This survey paper aims to collect and consider papers that deal with Probabilistic Ant based Clustering for Distributed Databases techniques. The aim is not to assume a constraint reviews, but quite to provide a broad state-of-the-art view on these related fields. Several previous methods have been projected to assist ant based clustering, particle swarm optimization algorithms, which has mentioned in a body of literature that is spread over a wide variety of applications.

Table 1: SUMMARY TABLE FOR COMPARISON OF PROBABILISTIC ANT BASED CLUSTERING FOR DISTRIBUTED DATABASE TECHNIQUES

TITLE	ALGORITHM	KEY-IDEA	TECHNIQUES	PERFORMANCE
An Adaptive Ant-Based Clustering Algorithm with Improved Environment Perception (2009) [6]	Adaptive Ant-based Clustering Algorithm (AACA).	Data unsupervised clustering and data exploratory analysis.	Data preprocessing and similarity scaling factor.	Ant algorithm reliably identifies the correct number of clusters.
Research on Hybrid Clustering Based on Density and Ant Colony Algorithm (2010) [8]	Hybrid clustering algorithm based on density and ant colony algorithm.	To determine the initial cluster centers according to cluster objects distribution density method.	Clustering, Density, Pheromone; Ant colony algorithm.	It performs 85.5463% accuracy.
A review on particle swarm optimization algorithms and their applications to data clustering (2011) [10]	Particle Swarm Optimization (PSO) and K-mean clustering.	Data clustering results into better cluster formation.	Data mining, Data clustering, K-mean clustering and Particle swarm optimization.	PSO gives good results and accuracy for single objective optimization.
A Pheromone Based Model for Ant Based Clustering (2012) [12]	Pheromone based ant clustering algorithm.	Ant agents use pheromone to communicate the place to make the clusters, not the path as in real ants.	Ant colony optimization, Clustering, Stigmergy, Pheromone.	Ants update their memory continuously and then calculate the probability accordingly.
An Improved Model for Ant based Clustering (2012) [13]	Ant colony clustering algorithm.	A model of intelligent movement of ants including the negative pheromone and direction selection.	Clustering, Ant colony Optimization, Pheromone and Distance method.	Improved results with lesser number of iterations and comparatively less calculations during one iteration.
A multiple pheromone algorithm for cluster analysis (2013) [15]	Multiple Pheromone Algorithm for Cluster Analysis (MPAC)	To detect individual features or combinations of features that link similar objects together in multidimensional space.	Ant Colony Algorithms, Artificial Life, Emergent Behavior, Cluster Analysis, and Self-Organization	The results illustrate the potential to learn clusters and to convert them into operational classification rules.

IV. CONCLUSION

This paper presents a brief survey about Probabilistic Ant based Clustering for Distributed Databases discussed with the different techniques. This survey can be classified into hybrid, density, Pheromone based ant clustering and cluster analysis. To conclude the discussion, the ant based clustering algorithm has a number of features that make it an interesting study of cluster analysis. It has the ability of automatically discovering the number of clusters. It linearly scales against the dimensionality of data. The nature of the algorithm makes it fairly robust to the effects of outliers within the data.

The further work enhanced and expanded for the Probabilistic Ant based Clustering for Distributed Databases technique in Fuzzy-Collective Intelligence of Colonies that leverages a parallel execution algorithm.

REFERENCES

[1] J.-L. Deneubourg, S. Gross, N. Franks, A. Sendova-Franks, C. Detrain and L. Chretien, "The dynamics of collective sorting: Robot-like ants and ant-like robots", In Proceedings of the First International Conference on Simulation of Adaptive Behavior: From Animals to Animats, Cambridge, MA, MIT Press, 1991, pp. 356-363.

[2] Jain A. K., Murty M. N., Flynn P. J., "Data Clustering: A Review". ACM Computing Surveys, Vol. 31, No. 3, pp 264-323, 1999.

[3] Han J., Kamber M. Data Mining: Concepts and Techniques, Beijing: Higher Education Press, 2001.

[4] J. Handl and B. Meyer. "Improved ant-based clustering and sorting in a document retrieval interface",In Proceedings of the Seventh International Conference on Parallel Problem Solvingfrom Nature (PPSN VII), volume 2439 of LNCS, pages 913-923. Springer-Verlag, Berlin, Germany, 2002.

[5] J. Handl, J. Knowles, and M. Dorigo. "Ant-based clustering: a comparative study of its relative performance with respect to .-means, average link and Id-som." Technical Report TR/IRIDIA/2003-24, IRIDIA, Universit 'e Libre de Bruxelles, July 2003.

[6] I. El-Feghi, M. Errateeb, M. Ahmadi and M.A. Sid-Ahmed, "An Adaptive Ant-Based Clustering Algorithm with Improved Environment Perception", International Conference on Systems, Man, and Cybernetics, pp. 1431-1438, 2009.

[7] J.B. Brown and M. Huber, "Pseudo-hierarchical ant-based clustering using Automatic Boundary Formation and a Heterogeneous Agent Hierarchy to Improve Ant-Based Performance", International Conference on Systems Man and Cybernetics, pp. 2016-2024, 2010.

[8] L. Li, W-C Wu and Q-M Rong, "Research on Hybrid Clustering Based on Density and Ant Colony Algorithm", Second International Workshop on Education Technology and Computer Science, pp. 222-225, 2010

[9] L.M. Li and M-M Shen, "An improved ant colony clustering algorithm based on dynamic neighborhood", International Conference on Intelligent Computing and Intelligent Systems, Vol. 1, pp. 730-734, 2010.

[10] S. Rana, S. Jasola and R. Kumar, "A review on particle swarm optimization algorithms and their applications to data clustering", Journal Artificial Intelligence Review, Vol. 35, pp. 211-222, 2011.

[11] Saroj Bala and R.P. Agarwal,"Hybridization of Ant based Clustering with Particle Swarms" (communicated)

[12] Saroj Bala, S.I. Ahson and R.P. Agarwal, "A Pheromone Based Model for Ant Based Clustering", International Journal of Advanced Computer Science and Applications, Vol. 3, No. 11, pp. 180-183, 2012.

[13] Saroj Bala, S.I. Ahson and R.P. Agarwal, "An improved Model for Ant Based Clustering", International Journal of Computer Applications, Vol. 59, No. 20, pp. 9-12, December 2012.

[14] Saroj Bala, S.I. Ahson and R.P. Agarwal, "Agglomerative Ants for Data Clustering", International Journal of Computer Applications, Vol. 47, No. 21, pp. 1-4, June 2012.

[15] J. Chircop and C.D. Buckingham, "A multiple pheromone algorithm for cluster analysis", Nature Inspired Cooperative Strategies for Optimization, Studies in Computational Intelligence, Vol. 512, pp. 13-27, 2013.

Authors Profile

Dhivya.N M.sc, Pursuing Mphil Research Degree in Kongunadu Arts and Science College at Coimbatore. She did her UG in Sri Ramakrishna Engineering College and PG degree in Kongunadu Arts and Science College at Coimbatore.

Sumangala.K has done her Master of Comoputer Applications at Madras University and Master of Philosophy at Mother Terasa University and pursuing her Ph.D in Bharathiar University. She currently works as Associate Professor in Department of Computer Science PG, Kongunadu Arts and Science College. His research work focuses on Intelligent Datamining and Nature Inspired algorithms. He has 23years of teaching experiences and 13 years of research experiences.

