ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

A Document of Ant Clustering By the Result of Topographic Mapping Intention

TV.V.Satyanarayana
2nd year M.Tech,
Aditya Institute of Technology and Management,
Tekkali.

Gannavarapu Nagendra Kumar
Sr. Assistant Professor,
Aditya Institute of Technology and Management,
Tekkali.

ABSTRACT:

The clustering and topographic mapping is a motivated model for explaining two types of developing behavior observed in real ant colonies. Existing work demonstrated some promising characteristics of the heuristic but did not extend to a rigorous investigation of its capabilities. We are using an Improved Technique called ATTA includes adaptive, heterogeneous ants, a time-dependent transporting activity, and a method that transforms the spatial embedding produced by the algorithm into an clear partitioning. ATTA is then subjected to the most painstaking experimental evaluation of an ant-based clustering and sorting algorithm. In the case of ant-based clustering and sorting, two related types of natural ant behavior are modeled. When clustering, ants gather items to form heaps. Its compare the performance to standard techniques for clustering and topographic mapping using a set of analytical evaluation functions and a range of synthetic and real data collections. It Return an explicit partitioning of data, which would allow clustering solutions to be evaluated without human intervention. It express the ability of ant-based clustering and sorting to automatically identify the number of clusters inherent to a data collection, and to produce high quality solutions .it is particularly robust for clusters of differing sizes , and overlapping clusters.

KEYWORDS:
Ant Based Heuristic, Ant Based Clustering and Sorting, Clustering, Topographic Mapping, Swarm Intelligence.

I.INTRODUCTION:

The large volumes of data arising today in fields such as bioinformatics and electronic document retrieval exemplify a trend that is occurring throughout our ‘knowledge economy’. Novel technologies (such as the Internet in document retrieval, micro-array experiments in bioinformatics, physical simulations in sci¬entific computing and many more) give rise to enormous warehouses of data, which can only be handled and processed by means of computers. While the ever-increasing storage and speed of computers facilitates this trend, it nonethe¬less relies upon a continuous development in database technology and informa¬tion processing techniques. The automatic analysis of data and their lucid presentation to humans is crucial in this context, as it is only when the data are interpreted, that they become meaningful and can provide new knowledge and insight.

The research field addressing these major challenges is generally referred to as data-mining. In the following, we focus on two sub problems encountered in data-mining, namely cluster analysis and topographic mapping. Cluster analysis is concerned with the division of data into homogeneous groups such that data items within one cluster are similar to each other, and those within different clusters are dissimilar. Clustering methods have been studied for many years, but they continue to be the subject of active research. Due to this, many clustering methods are available nowadays, differing not only in the principles of the underlying algorithm (which determine runtime behavior and scalability) but also in other characteristics, such as the types of attributes handled or the shapes of identifiable clusters.

To date, the four main classes of clustering algorithms are partitioning methods, hierarchical methods, density-based clustering and grid-based clustering. For an extensive survey the reader is referred to [15]. Algorithms performing topographic mapping go one step further than those for mere clustering. They are not limited to the detection of homogeneous groups within the data but aim to capture neighborhood relations within a two-dimensional (or possibly three- or higher-dimensional) visualization of a high-dimensional data-space. This is realized by pinpointing additional information in the form of (1) relationships between individual clusters and (2) relationships between data items belonging to the same cluster. In this sense, a kind of 'sorting' of the data items in two (or more) dimensions is obtained.

In an effort to close this gap, it is the main goal of this work to evaluate thoroughly the performance of ant-based clustering and sorting, and to com¬pare it to a number of standard techniques for both clustering and topographic mapping. We make this evaluation not on the basic algorithm, but on a much-improved version, called ATTA, which we present in detail here. This algorithm remains faithful to the underlying principles of ant-based clustering and sorting but introduces a number of modifications that are crucial to enable its experi-mental evaluation. In particular, the changes improve the spatial separation of the clusters, and enable the algorithm to be run robustly over a range of data sets.

To obtain a thorough understanding of all aspects of the capabilities of ant-based clustering and sorting, we employ a range of artificial and real test data. These enable us to, respectively, control different properties of the data sets and to verify the overall performance on realistic benchmarks. In each case, we employ a number of selected experimental evaluation functions to give a precise picture of the performance achieved by ATTA and by the algorithms against which we compare it, on these data sets. Some of the results on the clustering performance of our algorithm have been previously reported in [11, 12] in a less complete form.

2.Ant-based clustering and sorting:

Ant-based clustering and sorting was first introduced by Deneubourg et al. [4] to explain different types of naturally-occurring emergent phenomena. It is an instance of the broad category of ant algorithms [5], that is, algorithms that model 'some behavior' observed in real ants.[1] In the case of ant-based clustering and sorting, two related types of natural ant behavior are modeled [2]. When clustering, ants gather items to form heaps; an example of this being the cemetery formation (i.e., the clustering of dead corpses) observed in the species Pheidole pallidula. And when sorting, ants discriminate between different kinds of items and spatially arrange them according to their properties; a type of activity that can, for example, be observed in nests of Leptothorax unifasciatus where larvae are arranged as a function of their size.

In their paper [4], Deneubourg et al. proposed a continuous model to de-scribe these behaviors. From this, a discrete Monte Carlo model was derived, which was experimentally validated. In the computer simulation ants were rep-resented as simple agents, which randomly moved in their environment, a square grid with periodic boundary conditions. Items that were scattered within this environment could be picked up, transported and dropped by the agents. These operations were biased by the distribution of items within the agents' local neighborhoods, such that items that were either isolated or surrounded by dissimilar ones were more likely to be picked up, and then tended to be dropped again in the vicinity of similar ones. As a result, a clustering and sorting of the items on the grid was obtained.

Deneubourg work [4] mainly aimed at deriving a model applicable to collective robotics, but it was soon applied to data analysis. Lumer and Faieta [22] introduced a number of modifications to the model that both enabled the handling of numerical data and improved solution quality and the algorithms convergence time. It was in that paper that the ants sorting process was for the first time termed a "heuristic mapping of a possibly high-dimensional and sparse data set on a plane, in a way which preserves neighborhood relation-ships as much as possible”, that is, as an approximate topographic mapping (as opposed to a pure clustering). However, it is our contention that the results presented in this paper, although understandably leading to initial excitement and activity, donot suffice in demonstrating a true topology-preserving capability. In particular, the analytical measures used (mean local fit and grid entropy) do not capture the preservation of inter-cluster relationships at all and reflect only a limited or loose intra-cluster sorting.

3 ATTA: An improved ant-based clustering and sorting algorithm

Questions of interest

Previous research on ant-based clustering and sorting has left several questions related to the algorithm’s performance broadly unanswered, many of which fall into one of the following three categories.

3.1Clustering performance:

Seen purely as a clustering algorithm, how does the ant algorithm perform? How do its results compare to those obtained using classical clustering methods from the data-mining literature? In particular, is the algorithm competitive in terms of its runtime and of the quality of its solutions?

3.2Sorting performance:

To what degree is the spatial embedding generated by the ant algorithm topology-preserving? Are neighborhood relations preserved on a local or a global scale, or both? How do its results compare to those obtained using classical methods for topographic mapping from the data-mining literature? In particular, is the algorithm competitive in terms of its runtime and of the quality of its solutions?

3.3 Sensitivity to data properties:

How robust is the ant algorithms clustering and sorting performance with respect to different data properties? In particular, how strongly is it affected by the use of high dimensional and/or large data sets, by increasing overlap between clusters, or distinct deviations in the sizes of individual clusters?

Basic ant algorithm

```
Procedure for basic algorithm/*
INITIALIZATION PHASE*/
Randomly scatter data items on the toroidal grid
for j = 1 to #agents do
    i := randomselect(freedataitems)
    pickup(agent(j), i)
    g := randomselect(emptygridlocations)
    placeagent(agent(j), g)
end for
/* MAIN LOOP */
for itctr = 1 to #iterations do
    j := randomselect(allagents)
    ste p(agent(j), stepsize)
    i := carrieditem (agent(j))
    drop := dropitem?(f(i)) // see equations 2 and 3
    if drop = TRUE then
        drop(agent(j), i)
        pick := FALSE
        while pick = FALSE do
            i := randomselect(freedataitems)
            pick := pickitem?(f(i)) // see equations 1 and 3
        end while
        picku p(agent(j), i)
    end if
end for
end procedure
```

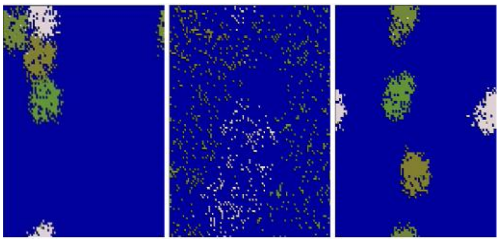

Figure 1: Spatial distribution on the grid at different stages of the run of our algorithm (with all

modifications). (a) After the initial clustering phase: the clusters touch. (b) After the interlude with the modified neighborhood function: the clusters have dissolved with data items spread out but remaining sorted.

(c) Final result: the clusters are clearly separated.

3.4 ATTA-TM: topographic mapping

To differentiate between the different algorithms for clustering and topographic mapping, we use the denomination ATTA-TM for the topographic mapping version of ATTA, which differs only in that it does not include the cluster-retrieval step described above.

4.Analyzing ATTA-C’s clustering performance:

Having explicitly defined an improved algorithm for ant-based clustering and sorting, ATTA, in the previous section, we are now in a position to proceed with our performance analysis. In this section, the clustering performance of ATTA-C is evaluated and compared to three alternative clustering methods and to one statistical method for the automatic determination of the number of clusters in the data. A range of synthetic and real data sets, and three distinct analytical evaluation functions, are used for this purpose. Both the experimental setup and the clustering results are presented in this section.

4.1 Experimental data:

Two different types of benchmark data sets are used. First, seven real data sets from the Machine Learning Repository [1], which we summarize in Table 1. Second, a range of two-dimensional synthetic data sets that permit the modulation of specific data properties. In the latter, each cluster is described by a two-dimensional normal distribution N (μ, σ). The number of clusters, the sizes of the individual clusters, and the mean vector μ~ and vector of the standard deviation σ~ for each normal distribution are manually fixed. In each run of the experiments, a new set of data is sampled from these distributions. These synthetic benchmarks are variations of the Square data set, a data set that has been frequently employed in the literature on ant-based clustering and sorting. It is two-dimensional and consists of four clusters of equal size (250 data items each), which are generated by normal distributions with a standard deviation of 2 in both dimensions and are arranged in a square. The data sets Square1 to Square 7 only differ by the distance between the individual clusters (i.e., the length of the edges of the square), which are 10, 9, 8, 7, 6, 5 and 4 respectively. They were generated in order to study the relative sensitivity of the algorithms to increasing overlap Between clusters

4.2 Data pre-processing:

All types of data are subject to a number of preprocessing steps: missing values are replaced by zeros; the data vectors are normalized in each dimension; dissimilarities between data vectors are computed using Euclidean distance (for the synthetic data) and Cosine similarity respectively (for the real data), and are normalized to lie within the interval [0, 1].

4.3 Results:

We begin by presenting results for the partition data sets, focusing on the sensitivity of the algorithms to two important data properties, degree of overlap of clusters, and differences in cluster sizes. Results on the real data sets based on the following partitions .

DATASET-1

1	Nagendra	475	950	870	280

DATASET-2

1	Sachin	400	946	500	261
2	Dhoni	320	1050	470	279
3	Suresh	310	950	410	270
4	Yuvraj	315	878	510	265

5	Sehwag	320	900	500	250
6	Zaheer	300	900	515	250

Without partition dataset

Name	Hits	Strikeouts	Homeruns	Walks	Year
Sachin	200	20	20000	20	2012

From the datasets and partition datasets,,ANT Algorithm find Euclidean distance for each data set. The dissimilarity with Euclidean distance is represented by

0.0	529.194	342.90	376.96	53.62
529.194	0.0	54.543	180.84	570.55
342.90	227.488	0.0	54.5436	376.167
376.96	180.848	376.167	0.0	414.548
53.62	570.555	283.213	414.58	0.0

Based on the above datasets and other items, if we enter nodelist with clustering values of alpha and beta with 0.2 and 0.8 ,Then the Graphical format is formed in Nodelist.

The Graphical Format for Node list is represented by

20 6 8 5

6 16 5

5. Conclusion

In the proposed algorithm, the objects are preprocessed by principal component analysis and their two principal components are processed as the projecting coordinates. Moreover, each object is looked as an ant. After the objects are projected to a plane, an artificial force field is added to the plane. The ant is attracted by its similar companies and repelled by dissimilar ones. The ant moves to a certain place according to the composite of all the forces. After many cycles, the similar ants will group and create a cluster.There are a number of directions in which research on ant-based clustering and sorting can be continued. Indeed, we are convinced that there is still room for improvement of the algorithm itself, though it will become increasingly difficult to obtain more than marginal performance gains. In our opinion, the hybridization of the algorithm with alternative clustering methods might therefore be a more rewarding and promising line of research. This work shows a previously unknown relation of two topographic mapping techniques. The assessment of algorithms conforms with experimental results from other researchers. Yet, it is based on the assumption that stochastic agents, e.g. ants, are nothing more than an arbitrary sampling technique that is to be omitted for further analysis of formulae.

6.References:

1) Halkidi,M.,Vazirgiannis,M.,&Batistakis,I.(2000).Qualityschemeassessmentinthe clustering process. In Proceedings of the Fourth European Conference on Principles of Data Mining and Knowledge Discovery, Vol. 1910 of Lecture Notes in Computer Science(pp. 265–267). Heidelberg, Germany: Springer-Verlag.

2) Handl,J.(2003). Ant-based methods for tasks of clustering and topographic mapping: extensions, analysis and comparison with alternative methods. Masters thesis. University of Erlangen-Nuremberg, Germany. http://www.handl.julia.de

3) Handl,J.,Knowles,J.,&Dorigo,M.(2003). On the performance of antbased clustering. DesignandApplicationofHybridIntelligentSystems,Vol.104ofFrontiersinArtificialIntelligence and Applications(pp.204–213). Amsterdam, The Netherlands: IOS Press.

4) andlHandl,J.,Knowles,J.,&Dorigo,M.(2004). Strategies for the increased robustness of ant-basedclustering.InEngineeringSelfOrganisingSystems,Vol.2977ofLectureNotesinComputerScience(pp.90–104). Heidelberg, Germany: Springer-Verlag.

5) Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). Swarm Intelligence From Natural to Artificial Systems .NewYork, Oxford University Press.

6) Carreira-Perpin´an, M. ´A. (2001). Continuous latent variable models for dimensionality reduction and sequential data reconstruction. PhD thesis, Department of Computer Science, University of Sheffield, UK.

7) Deneubourg, J.-L., Goss, S., Franks, N., Sendova-Franks, A., Detrain, C., & Chr´etien, L. (1991). The dynamics of collective sorting: Robot-like ants and ant-like robots. In Proceedings of the First International Conference on Simulation of Adaptive Behaviour: From Animals to Animats 1 (pp. 356– 365). Cambridge, MA: MIT Press.

8) Dorigo, M., Bonabeau, E., & Theraulaz, G. (2000). Ant algorithms and stigmergy. Future Generation Computer Systems, 16(8), 851–871.

9) Dorigo, M. & Di Caro, G. (1999). The ant colony optimization meta-heuristic. In Corne, D., Dorigo, M., and Glover, F., editors, New Ideas in Optimization.Handl, J. (2003). Ant-based methods for tasks of clustering and topographic mapping: extensions, analysis and comparison with alternative methods. Masters thesis. University of Erlangen-Nuremberg, Germany. http://www.handl.julia.de.

10) Handl, J. & Meyer, B. (2002). Improved ant-based clustering and sorting in a document retrieval interface. In Proceedings of the Seventh International Conference on Parallel Problem Solving from Nature, Vol. 2439 of Lecture Notes in Computer Science (pp. 913–923). Berlin, Germany: Springer-Verlag.

11) Hoe, K., Lai, W., & Tai, T. (2002). Homogeneous ants for web document similarity modeling and categorization. In Proceedings of the Third International Workshop on Ant Algorithms, Vol. 2463 of Lecture Notes in Computer Science (pp. 256–261). Heidelberg, Germany: Springer-Verlag.

12) Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. ACM Computing Surveys, 31, 264–323.

13) Kuntz,P.&Snyers&Memo,D.(1994)Emergento f colonization andgraphpartitioning. In Proceedings of the Third International Conference on Simulation of Adaptive Behaviour: From Animals to Animats Cambridge, MA: MIT Press.

14) Kuntz,P.&Snyers,D.(1999). Newresultsonanant-basedheuristicfor highlighting the organization of large graphs. In Proceedings of the 1999 Congress on Evolutionary Computation (pp. 1451–1458). Piscataway, NJ: IEEE Press.

15) Kuntz, P., Snyers, D., & Layzell, P. (1998). A stochastic heuristic for visualising graph clusters in a bi-dimensional space prior to partitioning. Journal of Heuristics, 5(3), 327–

351.

16) Lumer, E. & Faieta, B. (1994). Diversity and adaptation in populations of clustering ants. In Proceedings of the Third International Conference on Simulation of Adaptive Behaviour: From Animals to Animats 3 (pp. 501– 508). Cambridge, MA: MIT Press.

17) Lumer, E. & Faieta, B. (1995). Exploratory database analysis via self-organization. Unpublished manuscript. Results summarizedin.

18) MacQueen,L.(1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1(pp. 281–297). Berkeley, CA: University of California Press.

19) Martin,M.,Chopard,B.,&Albuquerque,P.(2002). Formationof anant cemetery: Swarm intelligence or statistical accident? Future Generation Computer Systems, 18(7), 951–959.

20) Kuntz, P., Snyers, D., & Layzell, P. (1998). A stochastic heuristic for visualising graph clusters in a bi-dimensional space prior to partitioning. Journal of Heuristics, 5(3), 327–351.

21) Lumer, E. & Faieta, B. (1994). Diversity and adaptation in populations of clustering ants. In Proceedings of the Third International Conference on Simulation of Adaptive Behaviour: From Animals to Animats 3 (pp. 501– 508). Cambridge, MA: MIT Press.

22) Lumer, E. & Faieta, B. (1995). Exploratory database analysis via self-organization. Unpublished manuscript.