Visual clustering based on chemical recognition system of ants

Nicolas Labroche, Nicolas Monmarché, Gilles Venturini

Visual clustering based on chemical recognition system of ants

Nicolas Labroche, Nicolas Monmarché, Gilles Venturini
Laboratoire d'Informatique de l'Université de Tours, EA 6300
nicolas.labroche@univ-tours.fr

Abstract

This paper describes the Visual AntClust clustering algorirthm that relies on a modeling of the chemical recognition system of ants to build a partition of a data set. The algorithm associates each artificial ant with a data object to be classified and represent its chemical signature in a 2D euclidian space. It then applies rules that mimic the behavior of real ants to group into the same nest (or cluster) the artificial ants. Therefore similar ants (or data) also tends to have similar coordinates in the 2D space. Experimental results show that this method can achieve good performances on artificial and real data sets and allows for a good visualization tool.

1 Introduction

Number of computer scientists have proposed novel and successfull approaches for solving problems by reproducing biological behaviors. For instance, genetic algorithms have been used in many research fields, such as clustering problems [2],[12] and optimization [11]. Other examples can be found when considering the modeling of collective behaviors of ants as in the well known algorithmic approach Ant Colony Optimization (ACO)([4]) in which pheromone trails were used. Similarly, ant-based clustering algorithms have been proposed ([10], [7], [11]). In these studies, researchers have modeled real ants abilities to sort brood. Artificial ants may carry one or more objects and may drop them according to given probabilities. These agents do not communicate directly with each others, but they may influence themselves through the configuration of objects on the floor. Thus, after a while, these artificials ants are able to construct groups of similar objects, a problem which is known as data clustering.

We focus in this paper on another important real ant collective behavior, namely the construction of a colonial odor and its use for determining the ant nest membership. As far as we know, this model has not been yet applied to any task in problem solving, and we show here how it can be used in data clustering problem and data visualization.

Our previous works led us to create a new clustering algorithm relying on simple rules inspired by real ants behaviors. In this early study, each object of the dataset was associated with an artificial ant (its genetic odor). At each iteration of the algorithm, two ants were randomly chosen and meet. According to the similarity between the ants genetic odors, a learned acceptance threshold and behavioral rules, an evolutive label, representative of the nest of the ant, was updated until each ant was well integrated in its nest. This method was quiet efficient but had two major weaknesses. First, the ant's modeling was too far from the reality : a single value stood for the label expressed by the ant and the initial data vector represented the genetic odor of the ant. Second, the algorithm does not allow the user to see the dynamic construction of the resulting partition, which can be helpfull when the number and the shape of expected clusters are unknown.

The remaining of this article is organized as follows : section 2 sums up the main principles of real ants recognition system. Section 3 presents the visual clustering algorithm that uses this new model : VISUAL ANTCLUST. Section 4 details experimental tests on benchmarks and their comparisons with ANTCLUST and a standard approach. Finally, section 5 concludes on future extensions of this promising model.

2 Main properties of real ants recognition system

Real ants have to solve every days a crucial recognition problem when they meet : they have to decide whether they belong to the same nest or not, in order to guaranty the survival of the nest. This phenomenon is know as "colonial closing".

It mainly relies on continual exchanges and updates of chemical cues on the ants cuticle determining ,as an identity card, their belonging nest.

Thus, each ant has its own view of its colony odor at a given time, and updates it continuously to preserve its nest being attacked by predators or parasites.

We are going now to introduce the main properties of such a system : on which principles the recognition system relies, how is generated real ants odor ("ontogenesis") and what are the mecanisms implied in its evolution. More details are to be found in [8] with a complete related mathematical model.

2.1 Principles of the recognition system

In the ants society, according to [5], nestmate recognition implies a complex system allowing discrimination between individuals, based on three distinct levels of analysis :

1. The existence of an individual chemical odor (or "label"), partially constructed by the ant, species and environment dependant, stocked over its cuticle.

2. A mecanism of chemical reception allowing the reading of the encountered odors and an associated model of reference (or "template"), which is either learned during the very first hour of existence or imposed at birth, and used during every ants meetings.

3. A set of decision rules leading discrimination between ants and behaviors to have according to their similarities.

According to Sherman and Holmes ([6]), the recognition between two ants relies on the detection of phenotypic differences ("phenotype matching"). Thus, each ant compare the other's label to its reference model to resolve the recognition problem.

Labels are made of chemical substances : mainly hydrocarbons that ants are able to synthetise, but also extracts from environment and food. Otherwise, hydrocarbons may vary qualitatively and quantitatively from one specy to another.

Lots of external factors can modify the odor and hence influence inter-individuals recognition process. We cite among these :

1. the colony's Queen : its role may vary from one specy to another.

   In some of the species, it has been demonstrated that Queen do not participate to the construction of the colonial odor but it contributes to its diffusion among all the ants of the colony (see works of [9]).

   In other species, as noticed by Carlin et Hölldobler in ([6],[1]), Queen's odor is the major recognition discriminator among all,when it is mature enougth. In this case, it adds a chemical odor over "its" ants designing them whith no doubt as belonging to its nest.

2. the food : there are evidences that two distinct ant species under similar diet tend to be less agressive with each other than if they had kept their old ones which were significantly different.

3. the environment : if discrimination process between two colonies relies heavily on the chemical substances from environment, then two ants society living in the same neighbourhood wood accept each other massively, what does not reflect biological reality (in [6]). In fact, ants wood rather consider their own nest odor than the odor of the environment near the nest to set their recognition cues.

4. the genetic information : gentic factors have a role in chemical odor construction but this influence might tend to be less than the influence of other factors or else halflings wood mutually consider themselves as intruders (in [6]).

To have a complete view of the main principles of "colonial closing", we must also consider ontogenesis of the ants odor and its evolution via internal ans external circuits.

2.2 Ontogenesis

At the early stage of their life, young ants does not have any cuticular label nor recognition template. On the other hand, they possess a "brood masking odor" over their cuticle, which attracts the other members of the colony who will then feed them in return. By the mean of this social act, young ants will impregnate physically odors from other colony members and learn their odors as a first template. This period lets time for young ants, to develop their own label simultaneously with the growth of their post-pharyngeal gland (PPG).

2.3 Internal circuits

They gather all the interactions between the different organs of the ants, at an internal level. Each ant possess the ability to generate hydrocarbons in their biosynthesis organ, according to their genetic information. The goal is to reinforce the ant's odor with its own synthetised substances during a period called "individual licking". The latter corresponds to an exchange of chemical substances between the ant's PPG and its cuticle. PPG being the place were all odors gather, it is more likely to be used in the definition of the ant's template at colony level ("Gestalt odor", which represents a sort of colonial recognition model reference).

2.4 External circuits

They gather all the chemical exchange mecanisms between two ants. Thus, according to the level of similarity existing between the two ants, they will either do a trophallaxis during which one of the two will decant its PPG contents in the other's PPG, or a "social liking" during which each ant spread a portion of its PPG over the other's cuticle, or else a simple contact during which only cuticular substances are exchanged.

3 Clustering and visualization algorithm Visual AntClust

3.1 The clustering problem

We focus in this paper, the unsupervised clustering problem in which we consider a set of objects $O = \{o\}$ not knowing in how many clusters they are. Thus, the goal is to find groups of similar objects closest to natural partition of the starting dataset. No assumption are made about the representation of the objects. They may have numerical or symbolic values. All we need here, is the definition of a similarity measure which takes as input a couple of objects and outputs a value between 0 and 1. Value 0 means that the two objects are totally different, 1 means that they are identical.

The following equations present the computing method used for similarity in our case.

We consider that each object is represented by a set of attributes, each of them having a data type among Θ (which is the set of existing data types).

For instance, we can have $\Theta = \{\theta\} = \{\mathbb{R}^+, \mathbb{R}^-, Symbolics\}$. Global similarity between two objects o_i and o_j can then be defined :

$$Sim(o_i, o_j) = \frac{1}{|\Theta|} \times \sum_{\forall \theta \in \Theta} Sim_\theta(o_i, o_j) \tag{1}$$

$$Sim_\theta(o_i, o_j) = 1 - \left(\frac{1}{|\theta|} \times \sum_{k=1}^{|\theta|} \Delta_\theta(o_i, o_j) \right) \tag{2}$$

where $|\Theta|$ represents the cardinality of the set of data types, $|\theta|$ the number of times that data type θ is used to describe an object o and finally Δ_θ a function which computes the similarity between two attributes of the compared objects o_i and o_j having data type θ. Description of the Δ_θ functions won't be detailled further, because they are not under the scope of this paper.

3.2 Main principles of Visual AntClust algorithm

The principle of this visual clustering algorithm is as follows : one object is assigned to each artificial ant and represents its genetic information and therefore a "part" of its recognition template. Moreover, each ant possess its own chemical odor (*"Odor"*) defined as a 2D vector to allow a simple visual representation in a continuous 2D plan. The goal of our algorithm is then to generate meetings against all ants so as to bring their *Odors* closer if they accept each others and further if they do not. After a while, groups of ants, representing the same type of objects, are gathered in the same regions of the *Odor's plan* and thus define a partition of the starting set of objects.

Thus, for one ant i, we define and explain the following parameters :

1. A genetic odor G_i entirely determined by an object o; G_i is time invariant.

2. A cuticular odor (or *Label*, *Odor*) C_i which will evolve according to the issue of the meetings. At the beginning, this odor is set randomly in the 2D plan, because no assumption concerning the objects partition is made.

3. A template T_i represented by an acceptance threshold. It is learned during an initialization phase, similar to real ants ontogenesis period, during which each artificial ant will meet others and each time will evaluate their similarity. The resulting template is a function of the similarities observed during this period.

4. A local estimator of the success of the ant meetings, B_i in its portion of the 2D odors plan. If an ant fails all its meetings the estimator value will be very low (near 0), otherwise it will be near 1, indicating that all the ants in the neighborhood represents similar objects. At the beginning, B_i is set to 0.

5. A threshold V_i indicating which odors the ant can perceive. Thus an ant can not meet an other one which has a cuticular odor completely different. This threshold is linked whith B_i, because we consider here that in nature, ants can reinforce their odor only with nestmates having similar odors through "trophalaxies" or "social lickings".

The detail of Visual AntClust main algorithm and meeting function are given hereafter.

VISUAL ANTCLUST()

(1) Initialize all the Nb_{Ants} ants using parameters described herebefore
(2) $t \leftarrow 1$, $Nb_{Finished} \leftarrow 0$
(3) **while** $(t \leq Nb_{ITER}) \wedge (Nb_{Finished} \leq 0.95 \times Nb_{Ants})$ **do**
(4)   Draw all ants in the 2D Odor plan
(5)   **for** $i \leftarrow 1$ to Nb_{Ants} **do**
(6)   Choose randomly an ant $j \neq i$
(7) $\sigma \leftarrow Sim(G_i, G_j)$
(8)   Meeting(i,j,σ)
(9)   **if** (Ant i has not yet finished evolving in the odor plan) $\wedge$ $(B_i > 0.999)$ **then**
(10)   Ant i will not change its odor anymore
(11) $Nb_{Finished} \leftarrow Nb_{Finished} + 1$
(12)   **endif**
(13)   Repeat steps (9) to (12) for ant j
(14)   **endfor**
(15) $t \leftarrow t + 1$
(16) **endwhile**
(17) Compute D_{MAX}, the maximal distance in the odor plan under which an ant can find a nestmate
(18) Gather in the same nest all the ants within a local perimeter of value D_{MAX}
(19) Reaffect each ant having no nest or a too little one, to the nest of the most similar ant found having a valid one

We have now to detail the fundamental underlying process of VISUAL ANTCLUST : the meeting of two ants according to their current odor and their similarity.

3.3 Ants meetings resolution

The crucial point of our method concerns resolution of meetings. It allows ants to share a cuticular odor with individuals with which they are closest genetically. It has been made possible by the establishment of rules which will be detailled hereafter in the MEETING algorithm.

We consider thereafter two ants i and j. We define that there is acceptance (or recognition) between i and j, σ being the similarity value computed between the two objects relatives to ant i and j, in the following case :

$$Acceptance(i, j) \Leftrightarrow (\sigma > T_i) \wedge (\sigma > T_j) \tag{3}$$

```
MEETING(Ant i, Ant j, σ_{i,j})
Compute View, the Euclidian distance between Odor_i and Odor_j
if View ≤ (1 − Max(B_i, B_j)) then
    if Acceptance(i, j) then
        Increment B_i and B_j
        if B_i > 2 × B_j then
            Reinforce Odor_j with Odor_i
        else
            if B_j > 2 × B_i then
                Reinforce Odor_i with Odor_j
            else
                Reinforce mutually Odor_i and Odor_j
            endif
        endif
    endif
endif
```

This algorithm needs little explanations. The bigger an ant's i indicator B_i is, the less chance this ant have to make encounters. In fact, we consider here that at the beginning an ant is not at the right place in the 2D odor plan. But as time goes on, and meetings are successfull, B_i increases signifying that there is no more need for ant i to change its cuticular odor. This period similar to ontogenesis allows each ant to find the better place in the Odor plan and hence ensure the convergence of the algorithm by preventing well placed ant to evolve.

4 Experiments and results

In this section, we will compare VISUAL ANTCLUST with a well-known method : the K-MEAN algorithm. The latter is initialized with 10 clusters generated randomly, so we will refer to it as 10-Means hereafter. Before, detailling experiments settings, benchmarks used for evaluation must be introduced.

Benchmarks and Experimental settings In order to test and compare the clustering abilities of the two methods, we use randomly generated and real data sets. For more details, see ([11]). Namely, there are : $\text{ART}_{i,i\in[1,8]}$ as artificial data sets and for real ones : IRIS, GLASS, PIMA, SOYBEAN and THYROID. However, the main characteristics of datas are summarized in the table (4).

All evaluation have been conducted over 50 tests for each data set and each method. Concerning VISUAL ANTCLUST, each test corresponds to an optimized number of iterations (max. 5000) during which all artificial ants realize random meeting with each other. Results are shown in table (4). The following fields are introduced in the table for each data file : the number of objects (*"#O"*) and their associated number of attributes(*"#A"*), the number of clusters expected to be found in the data (*"#C"*), the number of clusters effectivelly found by the two methods (*"#CF"*) with the standard deviation ("σ_{cf}") and finally the error generated by both algorithms (*"%E"*) associated with its standard deviation too("σ_e").

Datas	#O	#A	#C	# CF		%E	
				10_M $[\sigma_{cf}]$	V_{AC} $[\sigma_{cf}]$	10_M $[\sigma_e]$	V_{AC} $[\sigma_e]$
Art1	400	2	4	8.58 [0.98]	5.42 [0.99]	0.18 [0.01]	0.15 [0.04]
Art2	1000	2	2	8.52 [0.96]	5.08 [1.69]	0.38 [0.01]	0.21 [0.07]
Art3	1100	2	4	8.28 [0.96]	7.08 [1.51]	0.31 [0.01]	0.27 [0.04]
Art4	200	2	2	6.38 [0.75]	2.14 [0.40]	0.32 [0.02]	0.03 [0.05]
Art5	900	2	9	8.82 [0.91]	6.58 [1.39]	0.08 [0.01]	0.16 [0.04]
Art6	400	8	4	8.46 [1.08]	3.92 [0.27]	0.10 [0.02]	0.01 [0.04]
Art7	100	2	1	7.76 [1.03]	4.84 [1.28]	0.87 [0.02]	0.74 [0.09]
Art8	1000	2	1	8.78 [0.83]	5.18 [1.81]	0.88 [0.01]	0.69 [0.15]
Iris	150	4	3	7.12 [1.11]	2.28 [0.54]	0.18 [0.03]	0.19 [0.05]
Glass	214	9	7	9.44 [0.70]	6.12 [0.87]	0.29 [0.02]	0.32 [0.02]
Pima	798	8	2	9.90 [0.36]	12.04 [4.75]	0.50 [0.01]	0.49 [0.01]
Soybean	47	35	4	8.82 [0.97]	4.00 [0.00]	0.13 [0.02]	0.00 [0.00]
Thyroid	215	5	3	9.56 [0.57]	9.52 [2.12]	0.42 [0.02]	0.36 [0.08]

Table 1: Results obtained after 50 iterations of each method applied over each data.

Results Our algorithm Visual AntClust tend to perform better than the 10-Means method. It seems to be mainly because Visual AntClust manage to have, in general, a better appreciation of the number of clusters in the data. 10-Means founds too much clusters because it starts from 10 and it does not manage to reduce this number because of the too little difference existing between the objects. In fact, 10-Means does really perform better than Visual AntClust only one time : for Art5 because the number of clusters expected is quiet near 10 and that our algorithm does not manage to reach this number. Moreover, the main advantage of our method, unless the fact that no information concerning the number of expected clusters is needed, is that a user can see the dynamic evolution of the partitionning of the objects he(she) is studying (see fig. 1). That can provides informations concerning the size of the clusters or their shapes and help detecting objects not attached to any groups. To end with, these results show that Visual AntClust can treat from little to big sets of datas with a great success (see Soybean, Art1, Art4, Art6 and Art6) but demonstrate too, that it does not manage to find the right number of clusters when there are overlapping groups of objects or when data sets are made of white noise. For instance, our method fails with Art8 in which approximately 5 clusters are found when only one was expected.

5 Conclusion

We describe in this paper a new model for the ant recognition system and its application to the unsupervised clustering and data visualization problems. Results are good when compared with those of 10-Means algorithm for the clustering part. Moreover, our approach does not make any assumption about the nature of the datas to be clustered. That allows us to test our method in numerous application fields. The first one will be the web mining problem

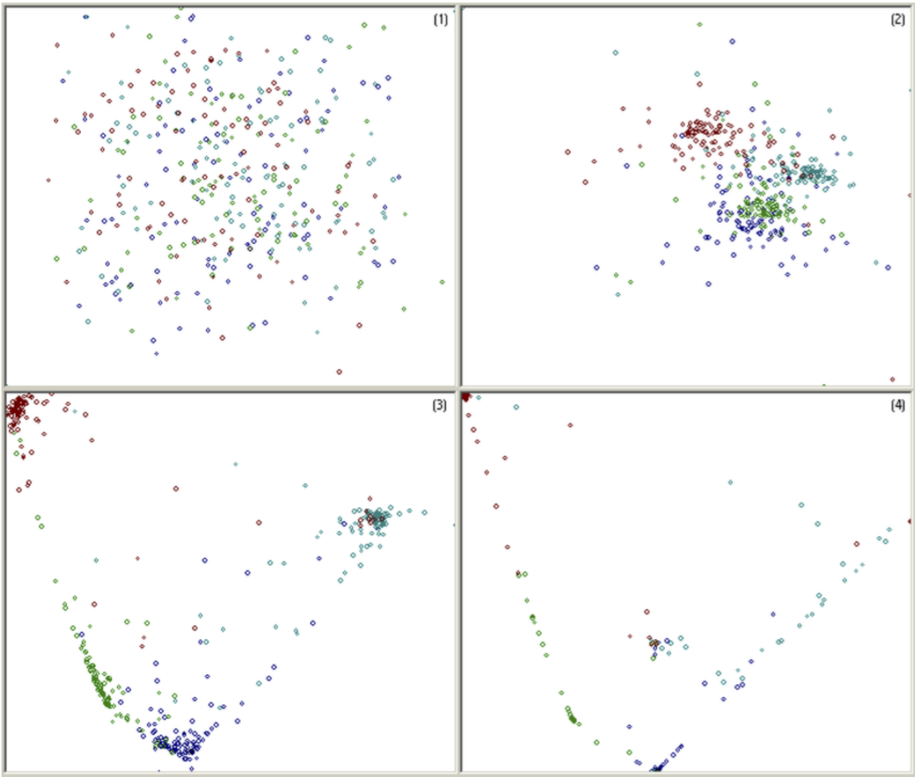

Figure 1: Examples of dynamic visualization of clustering for file Art1

and more precisely the study of Internet user behavior, because of the growing necessity of such tools for webmasters and because it provides a huge source of datas.

The visualization tool that our method provides can be extend to 3D, and we are currently working on adaptation of this method in this context. We hope then improving ergonomy and functionnalities to explore data sets.

Clustering method used at the end of the evolution of odors phase can be improved to better evaluate the distance used in the construction of colonies which leads to partitionning the starting set of objects. Finally, biological model can be also more precisely adapted to obtain better results in the future.

References

[1] N.F. Carlin and B. Hölldobler. The kin recognition system of carpenter ants(camponotus spp.). i. hierarchical cues in small colonies. *Behav Ecol Sociobiol*, 19:123–134, 1986.

[2] Y. Chiou and L. W. Lan. Genetic clustering algorithms. *European journal of Operational Research*, (135):413–427, 2001.

[3] D. Cliff, P. Husbands, J.A. Meyer, and Stewart W., editors. *Third International Conference on Simulation of Adaptive Behavior: From Animals to Animats 3*. MIT Press, Cambridge, Massachusetts, 1994.

[4] A. Colorni, M. Dorigo, and V. Maniezzo. Distributed optimization by ant colonies. In F. Varela and P. Bourgine, editors, *Proceedings of the First European Conference on Artificial Life*, pages 134–142. MIT Press, Cambridge, Massachusetts, 1991.

[5] D. Fresneau and C. Errard. L'identitloniale et sa reprntation chez les fourmis. *Intellectica*, 2:91–115, 1994.

[6] B. Hölldobler and E.O. Wilson. *The Ants*, chapter Colony odor and kin recognition, pages 197–208. Springer Verlag, Berlin, Germany, 1990.

[7] P. Kuntz and D. Snyers. Emergent colonization and graph partitioning. In Cliff et al. [3], pages 494–500.

[8] N. Labroche, N. Monmarché, and G. Venturini. Modsation du syst de reconnaissance chimique des fourmis. Rapport interne, Laboratoire d'Informatique de l'Université de Tours, E3i Tours, january 2002. 45 pages.

[9] S. Lahav, V. Soroker, R.K. Vander Meer, and A. Hefetz. Nestmate recognition in the ant *cataglyphis niger*: Do queens matter? *Behav Ecol Sociobiol*, 43:203–212, 1998.

[10] E.D. Lumer and B. Faieta. Diversity and adaptation in populations of clustering ants. In Cliff et al. [3], pages 501–508.

[11] N. Monmarché. *Algorithmes de fourmis artificielles : applications à la classification et à l'optimisation.* Thèse de doctorat, Laboratoire d'Informatique, Universit Tours, décembre 2000.

[12] L. Y. Tseng and S. B. Yang. Genetic clustering algorithms. *European journal of Operational Research*, (135):413–427, 2001.