SACOC: A spectral-based ACO clustering algorithm

Héctor D. Menéndez †*, Fernando E. B. Otero ‡, and David Camacho †

Abstract The application of ACO-based algorithms in data mining is growing over the last few years and several supervised and unsupervised learning algorithms have been developed using this bio-inspired approach. Most recent works concerning unsupervised learning have been focused on clustering, where ACO-based techniques have showed a great potential. At the same time, new clustering techniques that seek the continuity of data, specially focused on spectral-based approaches in opposition to classical centroid-based approaches, have attracted an increasing research interest—an area still under study by ACO clustering techniques. This work presents a hybrid spectral-based ACO clustering algorithm inspired by the ACO Clustering (ACOC) algorithm. The proposed approach combines ACOC with the spectral Laplacian to generate a new search space for the algorithm in order to obtain more promising solutions. The new algorithm, called SACOC, has been compared against well-known algorithms (K-means and Spectral Clustering) and with ACOC. The experiments measure the accuracy of the algorithm for both synthetic datasets and real-world datasets extracted from the UCI Machine Learning Repository.

Key words: Clustering, Data Mining, ACO, Spectral

† Departamento de Ingeniería Informática. Universidad Autónoma de Madrid
{hector.menendez,david.camacho}@uam.es
‡ School of Computing. University of Kent
F.E.B.Otero@kent.ac.uk

* This work has been partly supported by: Spanish Ministry of Science and Education under project TIN2010-19872 and Savier an Airbus Defense & Space project (FUAM-076914 and FUAM-076915).

1 Introduction

Unsupervised data mining techniques compose a complex field, where several different approaches have been tested in order to obtain similar or even better results to supervised techniques. The main difference between these two techniques is that supervised techniques have the label (target) information, which is used during the model generation, providing a more accurate model—the accuracy of the model is determined by comparing the prediction with the label information. Unsupervised techniques, instead, are totally blind in respect to the label information. An advantage of unsupervised techniques is that they can deal with a huge quantity of (unlabeled) data without a feedback of their performance.

Unsupervised techniques have been studied from different perspectives. Over the last few years, bio-inspired techniques are the most representatives, usually based on evolutionary algorithms or swarm intelligence that mimic a natural behaviour—e.g., the evolutionary process in genetic algorithm, collective behaviour in ant colony optimization. This work has been focused on the latter, which is becoming a promising field for unsupervised techniques. ACO algorithms are based on the foraging behaviour of ant colonies when they try to find the optimal path between their nest and a food source. Based on this idea, researchers have created several optimization algorithms in data mining, which have been focused on the path optimization process followed by the ants to create solutions for hard optimization problems [10, 14, 15].

The work presented in this paper is focused on the application of ACO in the unsupervised learning task of clustering, where the goal is to group (cluster) similar data points in the same group and, at the same time, maximise the difference between different clusters. It has been inspired by the Spectral Clustering (SC) algorithm [12] and the ACO-based Clustering algorithm (ACOC), proposed by Kao and Cheng [6]. ACOC is a centroid-based clustering algorithm, which tries to optimize the centroid (central point) position of each cluster. Following this idea, we focused the proposed algorithm on addressing a spectral-based approach. Inspired by other clustering algorithms [11, 13], we reformulated the original ACOC algorithm to create a spectral-based algorithm. Spectral-based clustering algorithms are usually good to define continuity-based clusters. They usually work with similarity graph amongst the data instances, which can be obtained as a Gram matrix of a kernel or a distance measure, and they study the spectrum of the graph in order to find the best cluster discrimination. In order to check the performance of the proposed algorithm, we have compared it against well-known clustering algorithms SC (Spectral Clustering) and [12] and K-means [9], as well as the original ACOC algorithm, in synthetic and real-world datasets.

The rest of the paper is structured as follows: Section 2 introduces the related work, Section 3 presents the new algorithm, Section 4 presents the computational results on synthetic and real-world datasets, and, finally, the last section discusses the conclusions and future work.

2 Related Work

Ant Colony Optimization (ACO) has become a promising field for data mining problems. In this context, ACO algorithms combine the ants foraging behaviour to generate patterns that describe the data according to a supervised or unsupervised learning criteria—depending on the type of algorithm, classification or clustering, respectively. This paper focuses on clustering problems.

Clustering [7] is based on a blind search within the data. Clustering techniques try to join similar data points into groups (clusters) according to a cost or objective function, which is usually minimized or maximized, making this clusters different from each other at the same time. There is a large number of clustering approaches depending on the goal that the algorithm should achieve. The most classical algorithms are K-means [9] and EM [3]. Both K-means and EM usually try to optimize estimator parameters to define clusters. Over the last decades, new non-parametrical algorithms such as Spectral Clustering [8] are gaining prominence. These algorithms study the graph spectrum generated by a similarity graph, usually extracted from a Kernel function and the Gram matrix associated by the application of the kernel to the data instances. The study of the spectrum maps the original data points to a projective space, where a simple K-means can be applied to group the data into clusters.

There are also bio-inspired algorithms that deal with the clustering problem, several of them focused on genetic algorithms. Hruschka et al. [4] presents a survey of clustering algorithms from different genetic approaches. From other bio-inspired perspectives, ACO algorithms have also produced promising results. Kao and Cheng [6] introduced a centroid-based ACO clustering algorithm; and Ashok and Messinger focused their work on graph-based clustering [1]; several other approaches are discussed in [5].

3 Spectral-based ACO Clustering Algorithm (SACOC)

This section presents the proposed Spectral-based ACO Clustering Algorithm (SACOC). This algorithm is similar to Spectral Clustering. The goal of the algorithm is to choose the data discrimination representing the information as a similarity graph and cutting it in different clusters.

3.1 ACOC algorithm

The ACOC algorithm is the base of SACOC. It has a search space based on instances and centroids, and can be defined as a graph whose associated

matrix is a N×M matrix, where N is the number of instances and M is the number of centroids (clusters).

The algorithm works with several ants looking for the best path in the graph Each ant (k) has the following features: a list of visited objects (tb^k), a set of chosen centroids C^k and a Weighted matrix W^k (related to the assignation of objects to clusters).
An ant k has two possible strategies: exploration and exploitation. It choose the strategy according to the following formula:

$$j = \begin{cases} argmax_{u \in N_i}\{[\tau(i,u)][\eta^k(i,u)]^\beta\} & , \text{ if } q \leq q_0 \\ S & , \text{ otherwise} \end{cases} , \tag{1}$$

where N_i is the set of nodes associated to object i, j is the chosen cluster, $\tau(i,u)$ is the pheromone value between i and u, q_0 is the exploitation probability, q is a random number for strategy selection,β is a parameter, $\eta^k(i,u)$ is the heuristic value between i and u for ant k defined by the formula:

$$\eta^k(i,u) = 1/d(x_i, c_j^k) = ||x_i - c_j^k|| \ , \tag{2}$$

where x_i is a data instance and c_j^k is a centroid from the ant centroid list. and S is the exploration defined by:

$$S = P^k(i,u) = \frac{[\tau(i,u)][\eta^k(i,u)]^\beta}{\sum_{j=1}^{m}[\tau(i,j)][\eta^k(i,j)]^\beta} \ . \tag{3}$$

The algorithm steps can be divided by:

1. Initialize pheromone matrix.
2. Initialize ants: (tb^k, C^k, W^k), for each ant k in the colony. Then, each ant repeats until tb^k is full:
   a. Select (randomly) a data object i satisfying $i \notin tb^k$.
   b. Select a cluster j: first the ant chooses a strategy; then, it calculates the transition probability and, finally, it visits a node.
   c. Update tb^k, C^k and W^k.
3. Choose the best solution. First, calculate the objective function for each ant:

$$J^k = \sum_{i=1}^{n}\sum_{j=1}^{m} w_{ij}^k d(x_i, c_j^k) \ , \tag{4}$$

   where w_{ij}^k is a weight value of the assignation matrix W^k. Next, rank ants solutions. Choose the iteration-best solution, apply local search[2] to improve the solution and, finally, compare it with the best-so-far solution and update this value with the maximum between them.

[2] For more details of local search see [16].

4. Update pheromone trails (global updating rule). Only the best r ants are able to add pheromones. Let ρ be the pheromone evaporation rate, $(0 < \rho < 1)$, t the iteration number, r is the number of elitism ants and $\Delta\tau_{ij}^{h} = 1/J^{h}$:

$$\tau_{ij}(t+1) = (1-\rho)\tau_{ij}(t) + \sum_{h=1}^{r} w_{ij}^{h}\Delta\tau_{ij}^{h} \ . \tag{5}$$

5. Check termination condition: if the number of iterations is greater than the maximum limit, finish; otherwise, go to step 2.

3.2 The Spectral hybridisation

The original ACOC algorithm uses the euclidean space as a search space. However, the algorithm can be modified to consider any kernel in a similar way that K-means is modified to generate the Spectral Clustering algorithm. Consider a graph G and its associated weighted matrix W, which is a pairwise similarity graph amongst the data. The similarity is calculated using a similarity function defined by a kernel $k(x_i, x_j)$. The Spectrum of the graph is calculated in a similar fashion used by Ng et al. [12] to create the original Spectral Clustering algorithm. First, we calculate the Laplacian matrix defined by:

$$L = I - D^{-1/2}WD^{-1/2} \ , \tag{6}$$

where I is the identity matrix and D represents the diagonal matrix whose (i,i)-element is the sum of the similarity matrix i-th row. After the creation of the Laplacian matrix, we extract the $v_1, \ldots, v_z$, which corresponds with the z largest eigenvectors of L—chosen to be orthogonal to each other in the case of repeated eigenvalues—and form the matrix $V = [v_1\, v_2\, \ldots\, v_z] \in \mathbb{R}^{n \times z}$ by stacking the eigenvectors in columns. Finally, we form the matrix Y from V by renormalizing each row of V to have unit length (i.e., $Y_{ij} = V_{ij}/(\sum_j V_{ij}^2)^{1/2}$). Then, we can consider Y as a projection of the original space and apply ACOC to the representation of each point.

4 Experiments

This section shows the experimental results. First, the synthetic and real-world datasets are described. Then, the experimental setup is shown. Finally, the computational results for both synthetic and real-world datasets are discussed.

4.1 Datasets Description

For the synthetic experiments we have used the following datasets [11]:

- *Aggregation*: This dataset is composed by 7 clusters, some of them can be separated by parametric clustering;
- *Jain*: This dataset is composed by two surfaces with different density and a clear separation;
- *Spiral*: In this case, there are 3 spirals close to each other.

For the real-world experiments, we have chosen three datasets from UCI Machine Learning Repository [2]:

- *Iris*: Contains 50 instances distributed over 3 classes, with 4 attributes each;
- *Haberman*: Contains 306 instances distributed over 2 classes, with 3 attributes each;
- *Breast Tissue* (Bre. Tis.): Contains 106 instances distributes over 6 classes, with 10 attributes each.

4.2 Experimental Setup

We have chosen K-means [9], Spectral Clustering (SC) [12] and ACOC [6] clustering algorithms to compare the results of SACOC.[3] K-means is an iterative algorithm based on centroids. The goal of the algorithm is to find the best centroids position. It involved two steps: it assigns the data to the closest centroid (cluster) and then, it calculates the new position of the centroid as a centroid of the data which has been assigned to it. SC generates a similarity graph and extracts its spectrum as a projective space in order to applied a simple clustering algorithm (in this case K-means) to the projective data.

The parameters of ACOC and SACOC are: the ants number is 10, the elitism is 1, the exploitation probability is 0.0001, the initial pheromone values have been set to $1/m$—where m is the number of clusters, $\beta = 2.0$, $\rho = 0.1$, the local search probability is 0.001 and the maximum number of iterations is 1000. These values have been chosen according to the original ACOC paper [6].

All algorithms need the number of cluster as an initial parameter. The experiments have been carried out 50 times using the Euclidean distance as the metric, except for Spectral Clustering and SACOC which use the Radial Basis Function. The evaluation of the experiments has been focused on two different ideas: the synthetic datasets have been evaluated according to the cluster discrimination and the performance of the algorithm in discriminating

[3] We used the K-means and SC implementation available in R; the author's implementation of ACOC and SACOC is available upon request.

Table 1 Minimum, Maximum, Median, Mean and Standard Deviation accuracy results of the application of the algorithms to the synthetic datasets. The p-values for the Wilcoxon test applied to SACOC and SC results are: Aggregation ($p = 1.062 \times 10^{-8}$), Jain ($p = 0$) and Spiral ($p = 0.02225$)—statistical significant improvements are indicated by a ▲ symbol.

SACOC	Min	Max	Median	Mean		SD
Aggregation	**98.60**%	**99.62**%	**99.24**%	**99.28**%	▲	± 0.0022
Jain	**100.0**%	**100.0**%	**100.0**%	**100.0**%		± 0.0000
Spirals	**100.0**%	**100.0**%	**100.0**%	**100.0**%	▲	± 0.0000
SC	Min	Max	Median	Mean		SD
Aggregation	63.96%	99.37%	88.39%	90.30%		± 0.0716
Jain	**100.0**%	**100.0**%	**100.0**%	**100.0**%		± 0.0000
Spirals	35.26%	**100.0**%	**100.0**%	93.20%		± 0.1724
K-means	Min	Max	Median	Mean		SD
Aggregation	66.88%	88.07%	78.55%	77.93%		± 0.0495
Jain	78.28%	78.28%	78.28%	78.28%		± 0.0000
Spirals	33.97%	34.94%	34.29%	34.41%		± 0.0020
ACOC	Min	Max	Median	Mean		SD
Aggregation	62.18%	86.17%	77.73%	77.07%		± 0.0516
Jain	73.19%	76.68%	74.80%	74.97%		± 0.0067
Spirals	33.65%	36.54%	35.26%	35.14%		± 0.0063

the original clusters; the real-world datasets have been evaluated using the accuracy rate, in order to check how close the algorithm is to real criteria.

4.3 Synthetic Experiments

Figure 1 presents the visual (best) results of the SC and SACOC algorithms when applied to the synthetic datasets. Table 1 shows the accuracy results on the same datasets.

Aggregation results show that SACOC achieved the best results and outperforms all the other algorithms—SACOC results are statistically significantly better than SC ($p = 1.062 \times 10^{-8}$). K-means and ACOC usually have the worse results in this dataset. These algorithms are not able to define clusters on the left (see Fig. 1), where the cluster boundaries are not clear.

Jain results show that both SC and SACOC are able to discriminate the clusters in all cases (see Table 1), both algorithms achieving the same results without statistically significant differences between them. K-means achieves stable results (the standard deviation is 0), while ACOC obtains the worst results.

Spirals shows that both SC and SACOC are able to define the clusters continuity—SACOC achieved the best and most stable results (0 standard

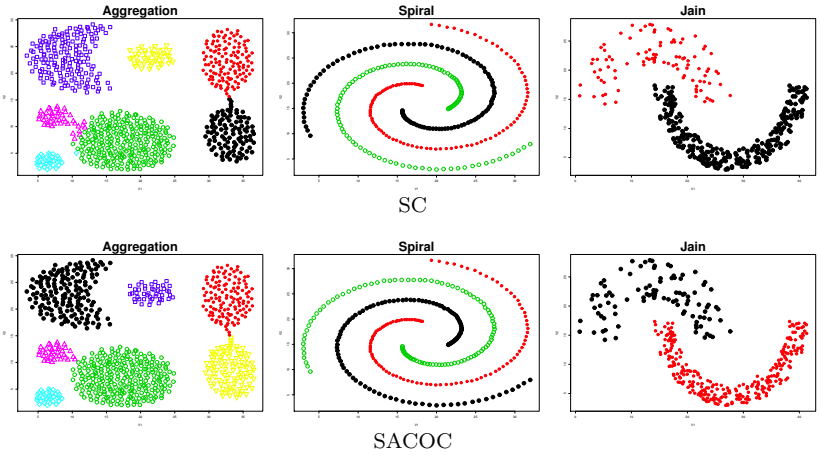

Fig. 1 Graphical representation of the best results on the synthetic datasets.

deviation), which are statistically significantly better than SC ($p = 0.02225$). K-means and ACOC are not able to define the continuity of the data due to the use of the Euclidean space.

Overall, the results for the synthetic datasets show that SACOC achieved best results, with statistically significant differences when compared to SC. In the next section we will compare the algorithms in real-world datasets.

4.4 Real-world Experiments

Table 2 shows the results of the algorithms applied to real-world datasets from UCI Machine Learning repository [2].

Breast Tissue dataset is more a spectral-like dataset. The data is continuous and the clusters do not intersects in several parts. In this case, both SACOC and SC achieved good results—SACOC results are statistically significantly better than SC ($p = 7.689 \times 10^{-9}$). K-means and ACOC have problems in discriminating the clusters information.

In Haberman case, SACOC achieved the best results, however, SC achieves the highest maximum value. This datasets shows more stable results for SACOC than SC (the standard deviation of SACOC is 0). There is also a high statistical significance between them ($p = 3.919 \times 10^{-10}$). K-means and ACOC achieved the worse results again in this case.

Iris dataset shows intersecting results. The best results are achieved by ACOC and K-means, while SACOC and SC achieved the worse results. This problem is likely due to the data projection, since it affects both SC and SACOC. Usually, when there are places with cluster intersections, the data projection is generally—it worse produces a big cluster and a cluster with a

Table 2 Minimum, Maximum, Median, Mean and Standard Deviation accuracy results of the application of the algorithms to the synthetic datasets. The p-values for the Wilcoxon test applied to SACOC and SC results are: Breast Tissue ($p = 7.689 \times 10^{-9}$), Haberman ($p = 3.919 \times 10^{-10}$) and Iris ($p = 5.371 \times 10^{-8}$)—statistical significant improvements are indicated by a ▲ symbol.

SACOC	Min	Max	Median	Mean	SD
Breast Tissue	**39.62**%	**60.38**%	**48.11**%	**48.43**%	▲ ± 0.0432
Haberman	**73.53**%	73.53%	**73.53**%	**73.53**%	▲ ± 0.0000
Iris	66.67%	68.67%	68.67%	68.53%	▲ ± 0.0038
SC	Min	Max	Median	Mean	SD
Breast Tissue	36.79%	48.11%	41.51%	41.30%	± 0.0321
Haberman	51.31%	**75.82**%	52.12%	52.37%	± 0.0341
Iris	68.00%	68.00%	68.00%	68.00%	± 0.0000
K-means	Min	Max	Median	Mean	SD
Breast Tissue	33.02%	34.91%	33.02%	33.02%	± 0.0032
Haberman	50.00%	52.29%	51.96%	51.52%	± 0.0020
Iris	58.00%	89.33%	89.33%	84.95%	± 0.1098
ACOC	Min	Max	Median	Mean	SD
Breast Tissue	30.19%	40.57%	33.02%	33.42%	± 0.0184
Haberman	50.65%	52.94%	51.96%	51.90%	± 0.0048
Iris	**89.33**%	**92.67**%	**90.00**%	**90.23**%	± 0.0079

couple of outliers. Even in this case, SACOC discriminates the clusters better than SC with statistically significant differences ($p = 5.371 \times 10^{-8}$).

These results show that SACOC achieved better and more stable results than SC in the datasets where the cluster assignation has clear boundaries and low cluster intersection. However, when there are intersection, it is harder for the algorithm to discriminate the data—in the same way that it is harder for SC.

5 Conclusions and Future Work

This paper presented a transformation of the ACOC algorithm into a Spectral algorithm. The new algorithm, called SACOC, uses spectral transformations of the original search space in order to apply the clustering in the projective space. The transformation consists on converting the original data in a graph-based representation (through a similarity graph) and calculate its Laplacian matrix. Once the Laplacian has been obtained, the eigenvectors are extracted and normalized to generate the projective space.

The proposed SACOC algorithm showed good results for synthetic datasets. It is able to discriminate continuity-based clusters with more stable results,

when compared to Spectral Clustering (SC). Also, the SACOC shows good results for real datasets, except in those cases where there are cluster intersections. In this situation, it has the same problems to discriminate the data than SC.

The future work will be focused on some improvements of the algorithm, such as, an initial centroid selection and to improve the spectral projections of the algorithm, in order to avoid cluster intersection problems. Also we will study a comparison on the algorithms performance, in order to improve memory consumption and running time.

References

1. Ashok, L., Messinger, D.W.: A spectral image clustering algorithm based on ant colony optimization. pp. 83,901P–83,901P–10 (2012)
2. Bache, K., Lichman, M.: UCI machine learning repository (2013). URL `http://archive.ics.uci.edu/ml`
3. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society. Series B (Methodological) **39**(1), 1–38 (1977)
4. Hruschka, E., Campello, R., Freitas, A., de Carvalho, A.: A survey of evolutionary algorithms for clustering. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on **39**(2), 133–155 (2009)
5. Jafar, O.M., Sivakumar, R.: Ant-based clustering algorithms: A brief survey. International Journal of Computer Theory and Engineering **2**, 787–796 (2010)
6. Kao, Y., Cheng, K.: An aco-based clustering algorithm. In: Ant Colony Optimization and Swarm Intelligence, *Lecture Notes in Computer Science*, vol. 4150, pp. 340–347. Springer Berlin Heidelberg (2006)
7. Larose, D.T.: Discovering Knowledge in Data. John Wiley & Sons (2005)
8. Luxburg, U.: A tutorial on spectral clustering. Statistics and Computing **17**(4), 395–416 (2007)
9. Macqueen, J.B.: Some methods of classification and analysis of multivariate observations. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, pp. 281–297 (1967)
10. Martens, D., Baesens, B., Fawcett, T.: Editorial survey: swarm intelligence for data mining. Machine Learning **82**(1), 1–42 (2011)
11. Menéndez, H.D., Barrero, D.F., Camacho, D.: A genetic graph-based approach for partitional clustering. Int. J. Neural Syst. **24**(3) (2014)
12. Ng, A., Jordan, M., Weiss, Y.: On Spectral Clustering: Analysis and an algorithm. In: T. Dietterich, S. Becker, Z. Ghahramani (eds.) Advances in Neural Information Processing Systems, pp. 849–856. MIT Press (2001)
13. Orgaz, G.B., Menéndez, H.D., Camacho, D.: Adaptive k-means algorithm for overlapped graph clustering. Int. J. Neural Syst. **22**(5) (2012)
14. Otero, F., Freitas, A., Johnson, C.: Inducing decision trees with an ant colony optimization algorithm. Applied Soft Computing **12**(11), 3615–3626 (2012)
15. Otero, F., Freitas, A., Johnson, C.: A New Sequential Covering Strategy for Inducing Classification Rules With Ant Colony Algorithms. IEEE Transactions on Evolutionary Computation **17**(1), 64–76 (2013)
16. Shelokar, P., Jayaraman, V.K., Kulkarni, B.D.: An ant colony approach for clustering. Analytica Chimica Acta **509**(2), 187–195 (2004)