The Artificial Bee Colony Algorithm for Unsupervised Classification of Meteorological Satellite Images

Rafik Deriche
Department Computer Science
University of Sciences and the Technology
Mohamed Boudiaf of Oran, Algeria

Hadria Fizazi
Department Computer Science
University of Sciences and the Technology
Mohamed Boudiaf of Oran, Algeria

ABSTRACT
The processing of satellite images plays an important role. It uses different attributes of images, and multiple mathematical and physical tools. Among the techniques of image processing, there exists the unsupervised classification which is based on the combination of image pixels into categories or themes. The classification algorithm designed and implemented in this article says Artificial Bee Colony (ABC) is one of bio inspired algorithms, it is very simple and very flexible compared to existing algorithms bio-inspired, and it has the advantage of rapid convergence and reduced memory size. We will apply the ABC algorithm to achieve an unsupervised classification of the meteorological satellite images of the Météosat-9 satellite.

Keywords
Unsupervised classification, Artificial Bee Colony, bio inspired algorithms, satellite images, Meteosat-9.

1. INTRODUCTION
The problems treated in computer are mostly solved by simple mathematical methods. However, these methods are not always adapted for complex problems, which appear for example in the domain of robotics, remote sensing or operational research. It is then necessary to invent new methods of resolution, and a source of inspiration used in the branch of artificial intelligence that is the modeling of complex natural systems. These are particularly biological systems that are studied: after an observation of a living being or of a group of living beings, a model is extracted and improved to a better resolution of the problem addressed. In this study of the existing, we will be interested at models that copy the social behavior of animals, i.e. the collective intelligence or swarm intelligence which it is possible to observe in the bees, which will be the main species studied in this article for make unsupervised classification of satellite images. Similar work [1] have been proposed in the terrestrial domain by supervised classification, our ABC is applied in a new domain of remote sensing which is the domain of meteorology to achieve an unsupervised classification.

2. UNSUPERVISED CLASSIFICATION
The unsupervised classification is based on the combination of image pixels into categories or themes; it is performed by computer processing based solely on image without the use of statistical training samples. The unsupervised classification produces a natural grouping of pixels of the image, is called spectral clustering or class. Thus, it is assumed that the regions of the image with the same spectral signature have a similar type of soil occupation. The analyst must determine the identity of the spectral clusters.

In problems of unsupervised classification, the number of classes is not known. This study area (Figure 1) was chosen for our thematic mastery of the region and its varied landscape that may be interest for the evaluation of our method.

The RGB color composite image (Figure 1) uses three channels: NIR 1.6, VNIR 0.8 and VIS 0.6. This combination of colors, vegetation appears greenish because of its much higher reflectance in the VNIR 0.8 channel (displayed in green) than in the NIR 1.6 channel (displayed in red) or VIS 0.6 channel (displayed in blue). The clouds composed of water droplets have a high aspect ratio in the three channels reflection, and therefore are whitish, while the clouds of snow and ice are cyan blue color due to the high absorption factor of the ice in NIR 1.6 channel (less red color signal). Bare soil appears brown because of higher reflection factor in the NIR 1.6 channel than VIS 0.6 channel, and the ocean appears black because of the low reflectance in all three channels.

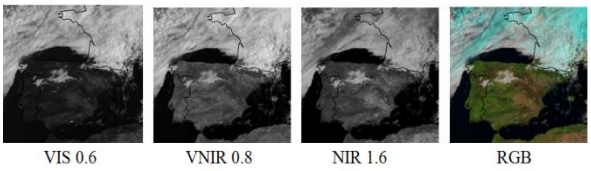


Fig 1: The study area

So the number of class is five: class1 (vegetation), class2 (bare soil), class 3 (Ocean), class 4 (water cloud), class 5 (ice cloud).

3. ABC ALGORITHM
A bee colony can extend over long distances (up to 14 km) and in several directions simultaneously for exploit a large number of food sources. In principle, the surfaces of flowers with abundant amounts of nectar or pollen that can be collected with less effort should be visited by more bees, whereas surfaces with less nectar or pollen should receive fewer bees. [2] [3] [4].

The foraging process begins in a colony of bees by scout bees sent the search for promising flower surfaces. Scout bees move randomly from one surface to another. When they return to the hive, the scout bees who found a surface that is classified above a certain quality threshold (measured as a combination of some components, such as the sugar content) deposit their nectar or pollen and run a dance. [2]

This dance can diffuse into the colony a lot of information or messages, the place of origin of food (quantity, location and floral origin) for example. The definition of the objective by the dance is precise so that the bees which followed the dance can find the objective, even if they make detours to reach. This analysis of bee behavior has led to fundamental discoveries about the human language.

The ABC algorithm is an optimization algorithm inspired by the natural behavior of bees for search nectar, to find the optimal solution, proposed by Karaboga in 2005. [5]

The ABC algorithm is a population-based algorithm, the position of a food source represents a possible solution to the optimization problem and the nectar amount of a food source corresponds to the solution quality (fitness) associated.

In its basic version, the ABC algorithm performs a kind of neighborhood search combined with random search. The colony consists of three groups of bees:

- Scout bees.
- Recruited bees.
- Selected bees.

It is assumed that there is only one artificial selected bee for each food source. In other words, the number of selected bees in the colony is equal to the number of food sources around the hive.

The selected bees go to their food sources and return to the hive and dance on this sector. The selected bee whose food source has been abandoned becomes a scout bee and begins searching for a new source of food. Recruited bees watch the dances of selected bees and choose food sources according dances.

The pseudo-code for the ABC algorithm in its simplest form: [6]

- `Initialize the population with random solutions.`
- `Evaluate fitness of the population.`
- `While (stopping criterion is not met) // forming new population.`
- `Select sites for neighborhood search.`
- `Recruit bees for selected sites (more bees for best sites) and evaluate fitness.`
- `Select the fittest bee from each surface.`
- `Assign remaining bees to search randomly and evaluate their fitness.`
- `End while.`

The algorithm requires a number of parameters to be set:

- The number of scout bees (n).
- The number of sites selected out of n visited sites (m).
- The number of best sites out of m selected sites (e).
- The number of bees recruited for best e sites (nep).
- The number of bees recruited for the other (me) selected sites (nsp).
- The initial size of surface (ngh) which includes site and its neighborhood.
- Stopping criterion.

In the first step, the ABC algorithm starts with the scout bees (n) being placed randomly in the search space.

In step 2, the fitness of the sites visited by scout bees are evaluated.

In step 4, bees that have the highest fitness are chosen as selected bees and sites visited by them are chosen for neighborhood search.

Then, in step 5, the algorithm conducts searches in the neighborhood of the selected sites, assigning more recruited bees to search near the best (e) sites. This differential recruitment is a key operation of the ABC Algorithm.

However, in step 6, for each surface only bee with the highest fitness will be selected to form the next bee population.

In step 7, the remaining bees in the population are assigned randomly around the search space scouting for new potential solutions.

These steps are repeated until a stopping criterion is met. At the end of each iteration, the colony will have two parts to its new population: those that were the fittest representatives from a surface and those that have been sent out randomly. [6]

4. ABC APPLICATION

The ABC algorithm is a algorithm based on the population, the position of a food source represents a possible solution to the optimization problem, and the nectar amount of a food source corresponds to the associated solution quality (fitness). In our case, the possible solution represents the coordinates of the pixel and the quality (fitness) represents the class of the pixel.

In our ABC model, the colony of artificial bees consists of three groups of bees: scout, selected and recruited bees.

- Scout bees (S) search around the hive to find new food sources (a new pixel).
- Selected bees (L) outweigh with them information about their food sources: distance and direction of the hive (the coordinates (i, j) of the pixel), and the quality of the nectar source (the class of the pixel).
- Recruited bees (R) watch the dances of selected bees and choose food sources depending on dances (feedback phase).

The two key behaviors in ABC are: the recruitment of a food source, and the abandonment of a source. If the studied pixel is a cloudy pixel this causes recruitment, otherwise if the pixel is a clear pixel this causes abandonment of that pixel.

Recruitment (feedback phase) of food source is done in two ways:

- Positive feedback (R+) if the pixel c (i, j) is a water cloud (Figure 2).
- Negative feedback (R-) if the pixel c (i, j) is an ice cloud (Figure 2).

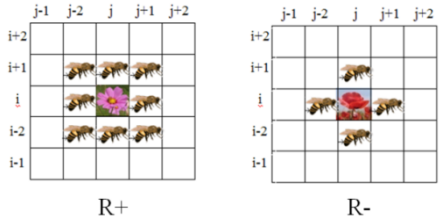


Fig 2: Structure of the neighboring pixel (i, j)

The search process to find the best solutions by ABC can be summarized as follows:

In the first step, initialize the bee colony: The algorithm starts with scout bees (n) being placed in our satellite image.

In step 2, calculate the fitness of each scout bee using the following formula:

$$fit(i) = \min d(x_{ij}, C_k)$$

Where:

- $fit(i)$: The fitness function of the bee i.
- x_{ij} : Pixel c (i, j).
- C_k : Class k.
- d : Any normalized distance.

In step 3. Verify the stopping criterion:

- If the condition is satisfied stop, (Condition: the total area of the image is explored and the optimum of the objective function is reached).
- Otherwise continue.

In Step 4: According to the fitness function $fit(i)$ and the probability $P(i)$, the bees visited the cloudy pixels are chosen as selected bees and the visited pixels are chosen for neighborhood search to generate new solutions. Using the following formula:

$$P(i) = \frac{1/\ fit(i)}{\sum_{n=1}^{SN}(1/\ fit(n))}$$

Where:

- i : Cloudy pixel.
- $P(i)$: The probability associated with the i_{th} food source.
- $fit(i)$: Represents the fitness function.
- SN : Represents the number of food source that is equal to the number of selected bees.

Then in step 5, the algorithm conducts searches in the neighborhood of the (e) classified pixels (feedback phase), assigning more recruited bees to search in neighborhood of the best (e) pixels (R +).

Each selected bee produces new potential solutions in the neighborhood of the current solution:

- Positive feedback (R+) = {xi / i = 1... nep}.
- Negative feedback (R-) = {xi / i=1,...,nsp}.

However in step 6, only recruited bees with the highest fitness will be selected to form the next bee population.

In step 7 the scout bees or recruited, which have not found solutions, return again to the hive to form the next bee population, and will be affected for new potential pixels search.

At the end of each iteration, the colony has two parts to its new population: those that were the fittest representatives from a surface and those that have been sent out randomly.

5. RESULTS AND DISCUSSION

All experiments were processed on a PC with Intel Pentium Dual CPU E2200 processor (2.20 GHz) with 2 GB of memory, and executed with C++ Builder 6, Microsoft Windows XP SP2.

First, the groups of pixels are obtained by the ABC algorithm. Each pixel of the satellite image is considered a possible source of food for the ABC algorithm. So the total number of the best food sources (e+me) in the ABC algorithm is equal to the total number of cloudy pixels in our meteorological satellite image.

The obtained groups by the ABC algorithm are further classified into five, during the classification; there are three classes that represent the clear sky and two classes cloudy sky that represent our two types of clouds (water cloud and ice cloud).

The accuracy assessment is an important step in the process of classification. In unsupervised classification, several known validity indices are available in the literature. In our application we will use a new objective function (Xie Beni index). The Xie Beni index (XB) measures the separability of groups. [7]

A great value of this index indicates a strong separability of the groups.

The performance of the algorithm ABC is influenced by a certain number of parameters (Table 1), such as: the number of scouts, the neighborhood size, the number of iteration…etc. In what follows we will study the influence of these parameters on the performance of the algorithm in order to evaluate their contribution to unsupervised classification, to the objective function and to the execution time.

Table 1. ABC parameters

Similarity function	Euclidean
The objective function	XB
Neighborhood size	3x3
nep	8
nsp	4
The number of scout	?
The number of iterations	?

To adjust the parameters of ABC, different values were tested, and the best ones are taken into account.

In order to study the influence of the number of Scout on the quality of classification, the ABC algorithm was executed with a different number of scouts. The Table 2 illustrates the effect of the number of scout on the quality of classification for maximization of XB.

Table 2. Influence of the scout on the quality of classification

Scout	XB
10	4,9347
25	4,9534
40	4,9577
55	4,9592
70	4,9577
85	4,9577
100	3,8248

According to Table 2 the optimum number of scout is 55 which maximizes XB for a value of 4,9592.

The algorithm of classification proposed was executed with number of iterations (It) varying between 10 and 100. We can see the results in the figures below:

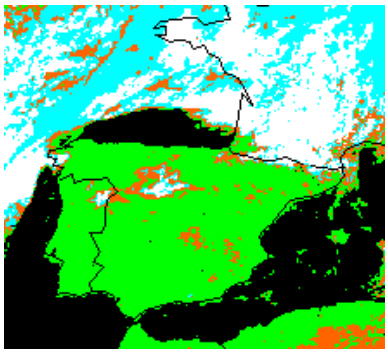


Fig 3: It = 10

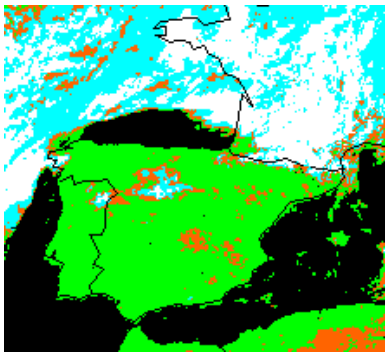


Fig 4: It = 15

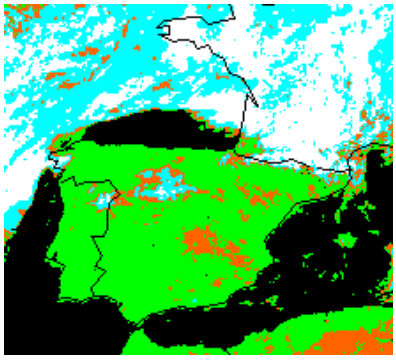


Fig 5: It = 20

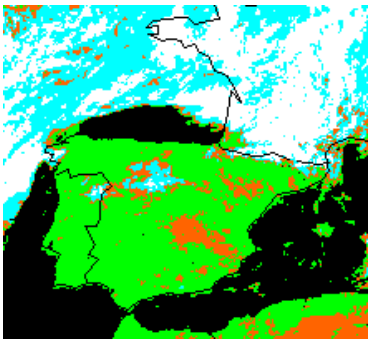


Fig 6: It = 25

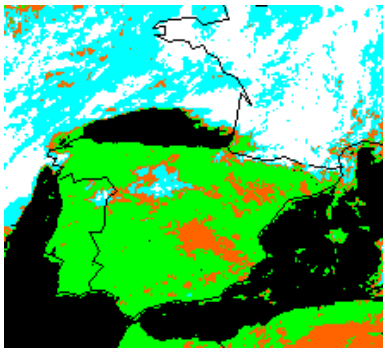
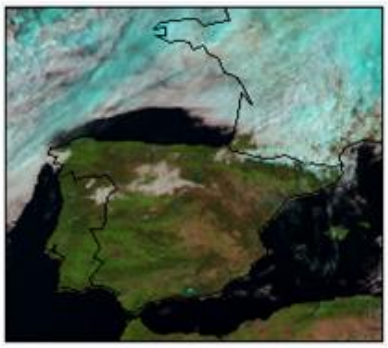


Fig 7: It = 28

According to these five figures (Figure 3-7), we observe that the convergence of our algorithm to the optimal solution is fast (convergence after 11seconde), as well as a stability of the results (after 28 iterations), which makes the algorithm used less material resources.

The best solution is reached after 28 iterations with execution time of 11,68 s (Figure 8).



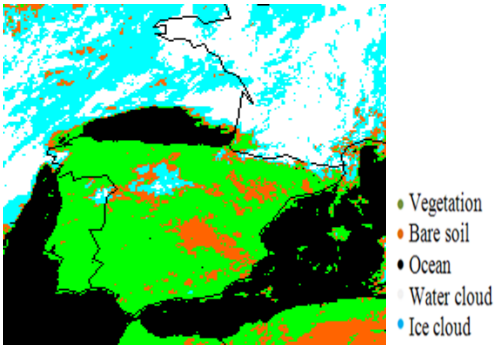


Fig 8: Result of classification by ABC

6. CONCLUSIONS

In the present work, we focused on developing a method for unsupervised classification of meteorological satellite images; the entire work is based on ABC Algorithm.

Our application is achieved on satellite data, original Meteosat-9; five classes were distinguished in the region. Many types of clouds are identified more clearly when the proposed algorithm is used.

The (XB) index was used as the objective function and it has also provided a mechanism to compare the results with other paradigms of unsupervised classification.

The classification speed is fast thanks to ABC algorithm, which has an exceptional performance of convergence. However, as a new meta-heuristic model in swarm intelligence, the ABC algorithm is still weak in mathematics as for the images classification discussed in this study, and certain parameters must be adjusted. These parameters have a direct influence on the speed and the stability of convergence.

The ABC algorithm is implemented using a similarity measure simple that is the Euclidean function; the performance of the algorithm can be increased by using other mathematical functions. The future scopes of the research propose to include some modification of the algorithm so that the (XB) index can be further ameliorated.

7. REFERENCES

[1] S. Banerjee, A. Bharadwaj, D. Gupta, and V. K. Panchal, Remote Sensing Image Classification Using Artificial Bee Colony Algorithm, international journal of computer science and informatics, volume 2, issue 3, 2012.

[2] K. V. Frisch, Bees: Their Vision, Chemical Senses And Language, Cornell University Press, Ithaca, NY, pp. 157, 1971.

[3] T. D. Seeley, The Wisdom of the Hive: The Social Physiology of Honey Bee Colonies, Harvard University Press, Cambridge, 1996.

[4] E. Bonabeau, M. Dorigo, and G. Theraulaz, Swarm Intelligence from Nature to Artificial System, First Edition, Oxford University Press, USA, pp. 1-24, 1999.

[5] D. Karaboga, An Idea Based On Honey Bee Swarm for Numerical Optimization, Technical Report-TR06, Erciyes University, Engineering Faculty, Computer Engineering Department 2005.

[6] D. T. Pham, A. Ghanbarzadeh, E. Koc, S. Otri, S. Rahim, and M. Zaidi, The bees algorithm-a novel tool for complex optimization problems, Proceedings of the 2nd Virtual International Conference on Intelligent production machines and systems (IPROMS 2006), Elsevier, Oxford, pp. 454-459, 2006.

[7] X. L. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Trans, Pattern Anal, Mach, Intell, vol. 13, no. 8, pp. 841-847, 1991.