Evolutionary Strategies and Entropy Approach for the Optimization of a Fuzzy Classification

M. Nasri and M. EL Hitmy
Ecole Supérieure de Technologie, B.P 473, OUJDA, MOROCCO.
Email : nasri@est.univ-oujda.ac.ma

Abstract

The fuzzy C-means algorithm is an unsupervised classification algorithm. This algorithm suffers from some difficulties, the number of classes must be known a priori, the initialization phase and the local optimums. We present in this paper some improvements to this algorithm based on the evolutionary strategies and the entropy approach in order to get around these three difficulties. We have designed a new evolutionist fuzzy C-means algorithm. We have suggested a new mutation operator which allows the algorithm to avoid local solutions and to converge towards the global solution in a small amount of computation time. We have proposed a new criterion for an optimal choice of the number of classes. This criterion uses in its derivation the entropy approach. The proposed method is validated on several simulation examples. The experimental results obtained show the rapid convergence and the good performances of this new approach.

1 Introduction

Classification consists of partitioning a set of objects into groups or classes in such a way that all objects belonging to one same class are all resembling between them and different from objects of other classes. This approach requires both a technique for measuring the resemblance between objects and the choice of an adequate criterion which measures the quality of the obtained grouping of objects. The classification problem becomes then a problem of a criterion optimization [1]. The algorithm of fuzzy C-means (FCM) is an unsupervised classification algorithm based on this approach [2] [3] [4], it is widely used for classification problems. Unlike other classification methods, the FCM algorithm uses the fuzzy logic to determine the best possible partition, the choice of the optimal partition is controlled by a fuzzy function. The FCM algorithm gives furthermore the degree of how much an object is a membership to its allocated class.

However the FCM algorithm requires the *a priori* determination of the number of classes [3] and suffers from the initialization phase and the local optimums [5] [6] [7] [8] :

- This algorithm requires the optimal choice of the classes number. This optimal choice guides the algorithm to provide a partition with the smallest error value possible.
- This algorithm converges in a finite number of iterations but the obtained solution depends on the initial values chosen for the algorithm, if indeed, we reinitialize the algorithm with a set of other values, it will converge to an other local solution which is entirely different from the first one.

We present in this work some improvements to this algorithm based on the evolutionary strategies and the entropy approach in order to get around these three difficulties. We have designed a new evolutionist fuzzy C-means algorithm (EFCM) which has so many advantages over the FCM algorithm. These are viewed in its generality, its parallelism and the genetic operations. The FCM algorithm deals with one unique solution at each iteration, while the proposed EFCM algorithm deals with a population of solutions in the same time. These solutions are subjected, during the iterations steps, to a Gaussian perturbation, which makes it then possible to avoid the local solutions. We have proposed a new mutation operator in order to be able to control the Gaussian disturbance level and to reduce the computation time required to converge towards the global solution. We have also proposed a new criterion for the optimal choice of the classes number based on the entropy approach.

Section 2 introduces the evolutionary strategies. In section 3, we give some definitions, and we recall the fuzzy C-means algorithm. Section 4 describes our evolutionist fuzzy C-means algorithm. In section 5, we present the new proposed criterion for an optimal choice of the classes number. While in section 6, the performances of this new method are evaluated by some experimental results. Finally, we give a conclusion.

2 Evolutionary strategies

Evolutionary strategies (ES) are particular methods for optimizing functions. These techniques are based on the evolution of a population of solutions which under the action of some precise rules optimize a given behavior, which initially has been formulated by a given specified function called *fitness function* [9].

An ES algorithm manipulates a population of constant size, formed by candidate points called *chromosomes*. Each chromosome represents the coding of a potential solution. It is composed by a set of elements called *genes*, these are *reals* [3].

At each iteration (*generation*) a new population is created from its predecessor by applying the genetic operators: se*lection* and *mutation*. The mutation operator perturbs with a Gaussian disturbance the chromosomes to generate a population optimizing further the fitness function. This procedure allows the algorithm to avoid the local optimums. While the selection operator constructs the next generation made by the pertinent individuals [3][9].

Figure 1 illustrates the different operations to be performed in a standard ES algorithm [9] [10] :

> Random generation of the initial population
> Fitness evaluation of each chromosome
> **Repeat**
> Select the parents
> Update the genes by mutation
> Select the next generation
> Fitness evaluation of each chromosome
> **Until** Satisfying the stop criterion

Figure 1 : *Standard SE algorithm.*

3 Fuzzy classification

3.1 Descriptive elements

Consider a set of M objects $\{O_1, O_2, ..., O_i , ..., O_M\}$ characterized by N attributes, grouped in a line vector form $V = (a_1\ a_2\ ...\ a_j\ ...\ a_N)$. Let $R_i = (a_{ij})_{1\leq j\leq N}$ be a line vector of $\boldsymbol{R}^N$ where a_{ij} is the value of the attribute a_j for the object O_i. Let *mat_va* be a matrix of M lines (representing the objects O_i) and N columns (representing the attributes a_j), defined by :

$$mat\_va = \left(a_{ij}\right)_{\substack{1\leq i\leq M\\ 1\leq j\leq N}}$$

V is the *attribute vector*, R_i is the *observation associated* with the object O_i or the *realization* of the attribute vector V for this object, $\boldsymbol{R}^N$ is the *observations space* [1] and *mat_va* is the *observation matrix associated* with V. The i[th] line of *mat_va* is the observation R_i. Each R_i belongs to a class CL_s, s=1, …, C.

From a geometric point of view, if we represent each observation by a point in the observations space $\boldsymbol{R}^N$, the set of observations will then provide a cloud of points in this space.

3.2 FCM algorithm

Consider M observations $(R_i)_{1\leq i\leq M}$ to be associated with C classes $(CL_s)_{1\leq s\leq C}$ of centers $(g_s)_{1\leq s\leq C}$. The centers are line vectors of dimension N. Unlike other classification methods, the FCM gives the degree of membership of R_i to a given class CL_s of center g_s, denoted μ_{is}, $\mu_{is} \in [0,1]$. The FCM is based on the minimization of the following optimization criterion [2][3][4] :

$$J=\sum_{i=1}^{M}\sum_{s=1}^{C}(\mu_{is})^{df}\left\|R_i-g_s\right\|^2$$

under the constraints :

$$\sum_{s=1}^{C}\mu_{is}=1 \text{ for } i=1 \text{ to } M \text{ and } 0<\sum_{i=1}^{M}\mu_{is}<M \text{ for } s=1 \text{ to } C$$

where $\|\cdot\|$ is an Euclidean distance. df is the *"fuzzy degree"* and may vary from 1 to infinity. Often, df is equal to 2 [3].

The FCM algorithm supposes that the number of classes C is known, *a priori*. After an initialization with a random value, the centers and the degrees of membership are updated iteratively. Let μ_{is}^* and g_s^* be the new values, we have [3][4] :

$$\mu_{is}^*=\frac{\left(\left\|R_i-g_s\right\|^2\right)^{\frac{-1}{df-1}}}{\sum_{k=1}^{C}\left(\left\|R_i-g_k\right\|^2\right)^{\frac{-1}{df-1}}} \quad \text{and} \quad g_s^*=\frac{\sum_{i=1}^{M}\left(\mu_{is}^*\right)^{df}R_i}{\sum_{i=1}^{M}\left(\mu_{is}^*\right)^{df}}$$

Figure 2 gives the FCM algorithm flowchart [3] [4] :

> Fix the number of classes C
> Initialize randomly the centers $(g_s)_{1\leq s\leq C}$
> Stop = false
> **While** Stop = false **Do**
> Adjust the degrees of membership
> Adjust the centers
> **If** $\max_s \left\|g_s-g_s^*\right\|<\varepsilon$ **Then** Stop = true
> **End While**
> Attribute each observation R_i to the class CL_s with the condition $R_i \in CL_s$ if $\mu_{is}=\max\mu_{ir}$, $1\leq r\leq C$.

Figure 2 : *Flowchart of the FCM algorithm.*

4 Evolutionary fuzzy classification

4.1 Proposed coding

The FCM algorithm consists of selecting among all of the possible partitions the optimal partition by minimizing a criterion. This yields the centers $(g_s)_{1\leq s\leq C}$. Thus we suggest the real coding as :

$$chr=(g_{sj})_{1\leq s\leq C, 1\leq j\leq N}$$
$$=\left(g_{11}\cdot\cdot g_{1N}\, g_{21}\cdot\cdot g_{2N}\cdot\cdot g_{s1}\cdot\cdot g_{sN}\cdot\cdot g_{C1}\cdot\cdot g_{CN}\right)$$

The *chr* chromosome is a real line vector of dimension $C\times N$. The genes $(g_{sj})_{1\leq j\leq N}$ are the component of the g_s center :

$$g_s=(g_{sj})_{1\leq j\leq N}$$
$$=(g_{s1}\, g_{s2}\cdot\cdot g_{sj}\cdot\cdot g_{sN})$$

To avoid that the initial solutions to be far away from the optimal solution, each of the *chr* chromosome of the initial population should satisfy the condition :

$$g_{sj}\in\left[\min a_{ij\,1\leq i\leq M}, \max a_{ij\,1\leq i\leq M}\right]$$

In the evolutionist fuzzy C-means algorithm, we must discard any chromosome of the initial population having a gene which does not satisfy this constraint. This gene, if any, is replaced by an other one which complies with the constraint.

4.2 The proposed fitness function

Let *chr* be a chromosome of the population formed by the centers $(g_s)_{1\leq s\leq C}$, for computing the fitness function value associated with *chr*, we define the *fitness function F* which expresses the behavior to be optimized (*J* criterion) :

$$F(chr)=\sum_{i=1}^{M}\sum_{s=1}^{C}(\mu_{is})^{df}\left\|R_i-g_s\right\|^2$$

with

$$\mu_{is}=\frac{\left(\left\|R_i-g_s\right\|^2\right)^{\frac{-1}{df-1}}}{\sum_{k=1}^{C}\left(\left\|R_i-g_k\right\|^2\right)^{\frac{-1}{df-1}}}$$

The chromosome *chr* is *optimal* if *F* is *minimal.*

4.3 The proposed mutation operator

The performances of an algorithm based on evolutionary strategies are evaluated according to the mutation operator used [11]. One of the mutation operator form proposed in the literature [7] [12] [13] is given by :

$$chr^{*}=chr+\sigma\times N(0,1)$$

where *chr** is the new chromosome obtained by a Gaussian perturbation of the old chromosome *chr*. $N(0,1)$ is a Gaussian disturbance of mean value 0 and standard deviation value 1, σ is the strategic parameter. σ is high when the fitness value of *chr* is high. When the fitness value of *chr* is low, σ must take very low values in order to be not far away from the global optimum.

We have been inspired from this approach to propose a new form of the mutation operator. The fact that we have proposed a new mutation operator is motivated by our interest to reach the global solution in a small computational time.

Let *chr* be a chromosome of the population formed by the centers $(g_s)_{1\leq s\leq C}$. Let $CL_s=\{R_i \,/ \mu_{is}=\max \mu_{ir}\ ,1\leq r\leq C\}$, i.e. the class formed by the observations R_i which have a membership degree to this class of center g_s higher than that if they were to belong to other classes CL_r having centers g_r. Let $g°_s$ be the gravity center of CL_s (figure 3).

$$g_S^{\circ}=\frac{\sum_{R_i\in CL_s}R_i}{l_s}\ \text{where}\ l_s=card(CL_s)$$

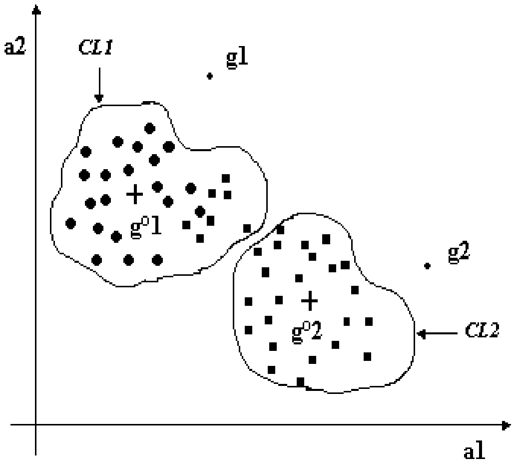

Figure 3 : *Illustration example in a two dimensional space.*

The mutation operator which we propose in this work consists in generating, from the *chr*, the new chromosome *chr** formed by the centers $(g^{*}_s)_{1\leq s\leq C}$, as :

$$g_s^{*}=g_s+f_m\times(g_s^{\circ}-g_s)\times N(0,1)$$

where f_m is a constant in [0.5,1]. The new strategic parameter proposed $\sigma' = f_m \times (g°_s - g_s)$ is low when g_s gets closer to $g°_s$ and is high when g_s is far from $g°_s$. The proposed σ' parameter has two advantages :

- When *chr* is far from the global solution, *chr* is subjected to a strong Gaussian perturbation allowing *chr* to move more quickly in the research space and in the same time to avoid local solutions.
- σ' controls the Gaussian perturbation level. Indeed, as the chromosome *chr* gets closer to the global solution, the Gaussian perturbation level is reduced until becoming null at convergence.

In order to choose the population parents chromosomes which will be mutated in order to generate children chromosomes we have adopted the technique of *choice by ordering.* We have also used the e*litist* technique [14].

4.4 The proposed EFCM algorithm

Figure 4 shows the different steps of the proposed EFCM algorithm.

Stage 1 :

1.1. Fix :
- The size of the population *maxpop.*
- The maximum number of generations *maxgen.*
- The fuzzy degree *df* (often $df = 2$).
- The constant f_m ($f_m \in [0.5 , 1]$).
- The number of classes *C.*

1.2. Generate randomly the population *P* :

$$P = \{chr_1, .., chr_k, ..., chr_{maxpop}\}$$

1.3. Verify for each *chr* of *P* the constraint :

$$g_{sj}\in[\min a_{ij}, \max a_{ij}],\ 1\leq i\leq M$$

1.4. Attribute for each *chr* of *P*, the observations R_i to the corresponding classes :

$$CL_s = \{R_i / \mu_{is} = max \mu_{ir}, 1 \le r \le C\}$$

1.5. Update the population P, for each chr of P do :

$$g'_s = \frac{g_s + \sum_{R_i \in CL_s} R_i}{1 + l_s} \text{ where } l_s = card(CL_s)$$

1.6. Compute for each chr of P its fitness value $F(chr)$.

Stage 2 :

Repeat

2.1. Order the chromosomes chr in P from the best to the poor (*in an increasing order of F*).
2.2. Choose the best chromosomes chr.
2.3. Attribute for each chr of P, the observations R_i to the corresponding classes :

$$CL_s = \{R_i / \mu_{is} = max \mu_{ir}, 1 \le r \le C\}$$

2.4. Mutation of all the chromosomes chr of P except the first one (*elitist technique*) :

$$g_s^* = g_s + f_m \times (g_s^\circ - g_s) \times N(0,1)$$

2.5. Attribute for each chr of P except the first one, the observations R_i to the corresponding classes:

$$CL_s = \{R_i / \mu_{is} = max \mu_{ir}, 1 \le r \le C\}$$

2.6. Update the population P, for each chr of P except for the first one, do :

$$g'_s = \frac{g_s + \sum_{R_i \in CL_s} R_i}{1 + l_s} \text{ where } l_s = card(CL_s)$$

(*The population P obtained after the updating is the population of the next generation*)
2.7. Compute for each chr of P its fitness value $F(chr)$.

Until *Nb_gen* (*generation number*) > *maxgen*

Figure 4 : *The proposed EFCM algorithm.*

5 Determination of the optimal number of classes

Choosing the right number of classes C, In many partition problems, is a difficult task. Several criteria for choosing the optimal number of classes, based on different approaches, have been proposed in the literature [15] [16] [17] [18].

For a given value of C, the EFCM algorithm obtains at convergence the global optimal partition for which the value of F is minimal. Let $f(C) = \min_{chr} F(chr)$ this value (chr is a real line vector of dimension $C \times N$ formed by the centers $(g_s)_{1 \le s \le C}$). The aim is to find a number $C_{opt} << M$ which gives rise to an optimal fuzzy classification. This ensures a partition with the lowest error value possible.

In [19], Ruspini interpret μ_{is} (degree of membership of the observation R_i to the class CL_s) as the probability $p(R_i \in CL_s)$ that R_i belongs to CL_s. Similarly, in this paper, we interpret μ_{is} as the probability $p(R_i \in CL_s)$ that R_i belongs to CL_s. Indeed, we have :

$$\mu_{is} \in [0,1] \text{ and } \sum_{s=1}^{C} \mu_{is} = 1, \forall \text{ i}$$

The *a priori* probability of each observation is $p(R_i)$:

$$p(R_i) = \frac{1}{M}$$

Thus the *a priori* probability of the class CL_s is [20]:

$$p_s = \sum_{i=1}^{M} p(R_i) p(R_i \in CL_s) = \frac{1}{M} \sum_{i=1}^{M} \mu_{is}$$

A class is represented in the observations space as a cloud of points. The position (repartition) of these clouds in the observations space may be characterized by the intra-cloud distances and the entropy. Palubinskas et al. [21] have carried out several works on this subject but in the non fuzzy case. The first term of their criterion is the sum of the intra-cloud distances. The second term is the entropy associated with all the clouds present.

We have been inspired by this approach to define the entropy of the clouds of points in the fuzzy case. We define the entropy in our case as :

$$E(C) = -2M \sum_{s=1}^{C} p_s \log(p_s)$$

Thus, to determine the optimal number of classes, we propose the criterion :

$$J_{HE}(C) = f(C) + E(C)$$

The first term $f(C)$ of the criterion characterizes the homogeneity of the clouds of points. This term becomes smaller if a particular cloud contains points which are more similar between them.

The second term $E(C)$ is the entropy of the clouds of points. This term becomes smaller if all the points which are similar to each other belong to the same cloud of points and all the other points are outside**.**

The EFCM algorithm runs for several values of C, $C \in [C_{min}, C_{max}]$ ($2 \le C_{min}$ and $C_{max} << M$). For each value of C, this algorithm obtains at convergence the values of $f(C)$ and μ_{is}. Once the values of μ_{is} are obtained, we compute $E(C)$ and we determine $J_{HE}(C)$. The optimal number of classes C_{opt} is chosen such as :

$$J_{HE}(C_{opt}) = \min_C J_{HE}(C) \Leftrightarrow C_{opt} = \arg\min_C J_{HE}(C)$$

6 Experimental results and evaluations

6.1 Introduction

We have considered three simulation tests in the observations space of dimension 2. These tests are different from each other by the repartition type of the classes in the observations space. In each test, the classes are generated randomly by Gaussian distributions and each class contains 100 observations.

6.2 Test 1

In this test, the number of classes chosen is C=3 and the overlapping degree between the classes is null. The classes are well separated between them. Table 1 gives the real centers of the classes and figure 5 shows the repartition of the observations in the observations space.

Class	CL_1	CL_2	CL_3
Center vector	4 4	6 2	8 4

Table 1 : *Real centers of the classes.*

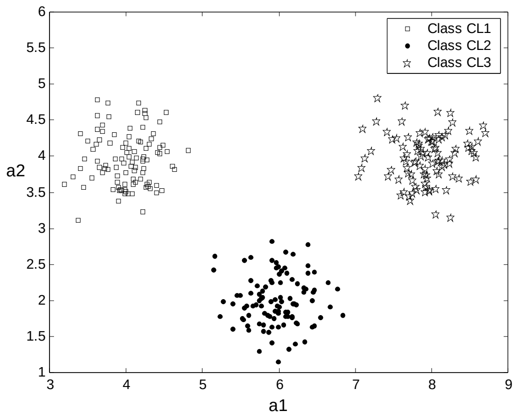

Figure 5 : *Repartition of the observations in the space.*

The proposed evolutionist algorithm runs quickly. Figure 6 shows the evolution of the fitness value of the best chromosome of the current population as long as the generations progress. The optimal chromosome chr_{opt} obtained is :

chr_{opt}= (4.0179 3.9522 5.9718 1.9959 7.9442 3.9672)

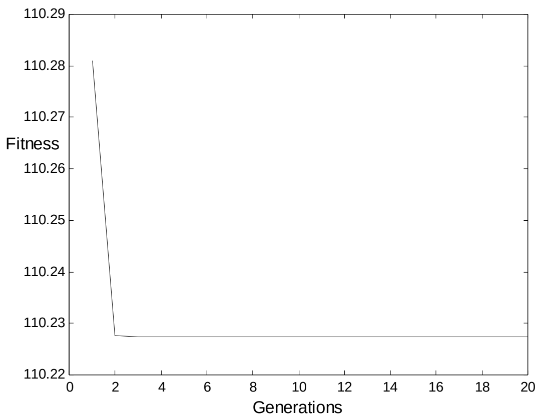

Figure 6 : *Fitness evolution.*

We noticed that in very few generations, the EFCM algorithm converges to the global optimum and determines the class centers. This is due to the parallel nature of the evolutionist algorithm and also to the nature of the proposed mutation operator which has rapidly guided the algorithm, by means of an adapted Gaussian perturbation, to the global solution. The local solutions have well been avoided. The centers obtained are slightly shifted from the real centers.

The classification results obtained by the proposed EFCM algorithm are summarized in figure 7 and table 2.

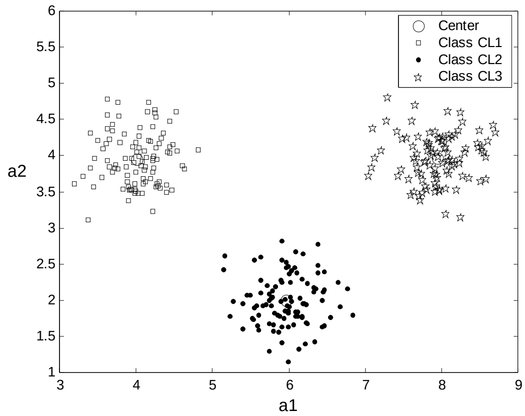

Figure 7 : *Optimal classes and centers obtained using EFCM.*

	Estimated CL_1	Estimated CL_2	Estimated CL_3
CL_1	100	0	0
CL_2	0	100	0
CL_3	0	0	100

Table 2 : *Confusion matrix.*

These results show that all the observations are correctly attributed to their corresponding classes, the error rate obtained is null. The initialization problem is removed, the result obtained is the same for different initializations. The proposed mutation operator has permitted to the algorithm to avoid local optimums and to converge rapidly to the global solution.

6.3 Test 2

In this test, we have considered three other classes, but the overlapping degree in this case is high. The classes are very close to each other and have the same centers as the classes of test 1. Figure 8 shows the repartition of the observations in the observations space. We notice that it is difficult to find the optimal partition in this case.

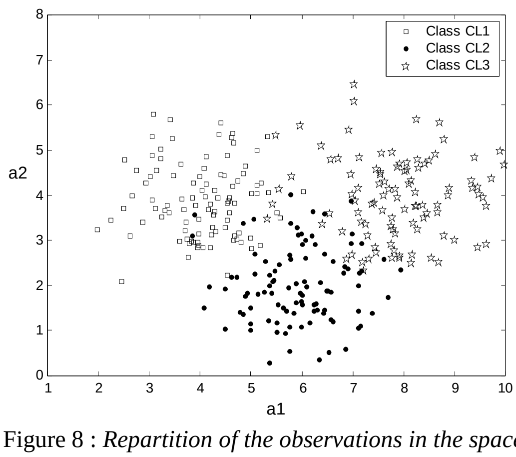

Figure 8 : *Repartition of the observations in the space.*

Figure 9 shows the evolution of the fitness value of the best chromosome of the current population with respect to the progressing generations. It shows that the proposed algorithm converges rapidly to the global solution. The rapidity of the algorithm is not sensitive to the overlapping degree. The optimal chromosome chr_{opt} obtained is :

chr_{opt}= (4.1012 3.9225 5.9447 1.8709 7.9207 3.9308)

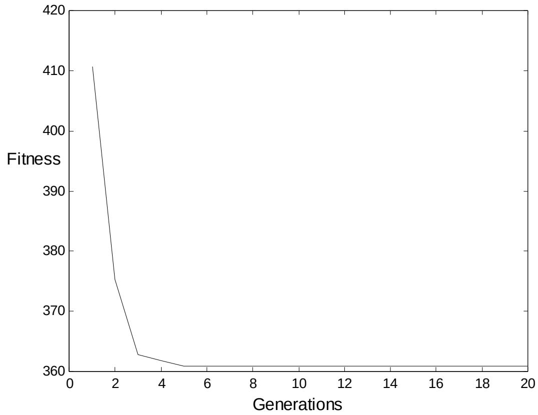

Figure 9 : *Fitness evolution.*

Figure 10 and table 3 summarize the classification results obtained by the proposed algorithm.

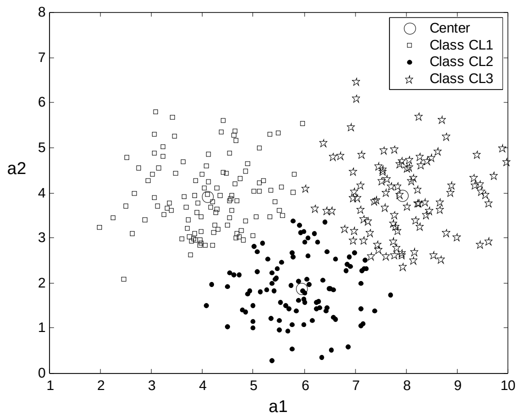

Figure 10 : *Optimal classes and centers obtained using EFCM.*

	Estimated CL_1	Estimated CL_2	Estimated CL_3
CL_1	96	3	1
CL_2	5	87	8
CL_3	6	5	89

Table 3 : *Confusion matrix.*

The number of misclassified observations in this case is 28. The corresponding error rate is :

$$\tau = \frac{28}{300} = 9.33\%$$

The error rate has increased with the overlapping degree. By analyzing the repartition of the classes, we noticed that the misclassified observations are situated :

- Either far away from the space of their corresponding classes, for instance the class CL_1 contains 4 observations of class CL_2 (figure 8).
- Either in the boundaries of separation between the classes, for instance the boundary which separates the two classes CL_2 and CL_3 (figure 8).

It is then normal that these observations are misclassified, this explains the high error rate value obtained.

6.4 Test 3

In this test, the aim is to evaluate the performances of the EFCM algorithm for a high number of classes, we have chosen C = 6. The overlapping degree between the classes is high. The real centers of the 6 generated classes are shown in table 4.

Class	CL_1	CL_2	CL_3	CL_4	CL_5	CL_6
Center vector	4 5	4 7	6 3	6 6	8 5	8 7

Table 4 : *Real centers of the classes.*

Figure 11 shows the repartition of the classes in the observations space, it shows that it is difficult to find the best partition for such a case. The observations of each class are indeed not concentrated around their class center. It is then possible to find observations of a class CL_s which are more close to the center of an other class $CL_{s'}$ than they are to their own center (figure 11). These observations are generally misclassified.

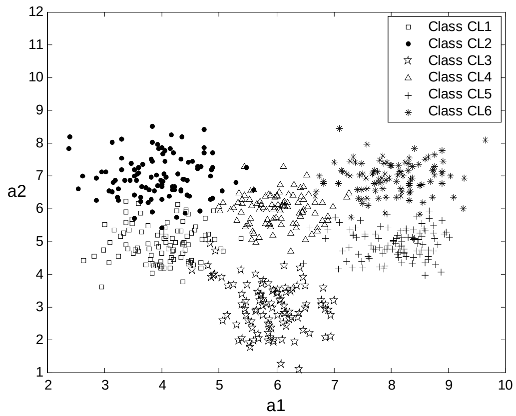

Figure 11 : *Repartition of the observations in the space.*

The proposed EFCM algorithm converges in a small number of generations (not more than 6) towards the global optimum (figure 12). The optimal chromosome chr_{opt} obtained in this case is :

chr_{opt} = (4.1071 4.8030 3.9031 6.9612 6.0048 2.8858 6.0564 5.9842 8.1286 4.9510 8.0197 6.9866)

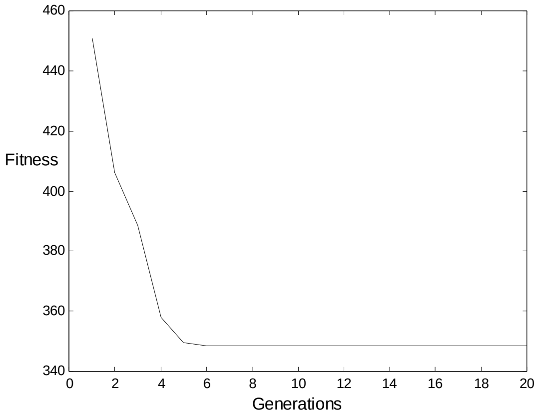

Figure 12 : *Fitness evolution.*

The classification results obtained by the EFCM algorithm are summarized in figure 13 and table 5.

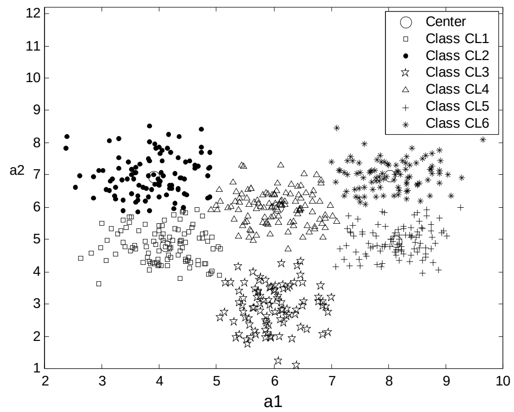

Figure 13 : *Optimal classes and centers obtained using EFCM.*

	Estimated CL_1	Estimated CL_2	Estimated CL_3	Estimated CL_4	Estimated CL_5	Estimated CL_6
CL_1	91	7	0	2	0	0
CL_2	5	91	0	4	0	0
CL_3	8	0	92	0	0	0
CL_4	0	0	0	98	0	2
CL_5	0	0	1	5	94	0
CL_6	0	0	0	5	4	91

Table 5 : *Confusion matrix.*

The number of misclassified observations is 43, the corresponding error rate is :

$$\tau = \frac{43}{600} = 7.17\%$$

Whilst the number of classes increases with a high overlapping degree between the classes, the error rate value obtained remains low. This confirms the good performances of the EFCM algorithm presented even when the number of classes is high.

6.5 Estimation of the optimal number of classes

We here evaluate the performances of the proposed J_{HE} criterion. For this, we have retained the three experimental tests previously presented. In each test, the EFCM algorithm was run for several values of C, $C\in[2,6]$ for tests 1 and 2 and $C\in[2,10]$ for test 3. Figures 14 to 16 show the evolution of the J_{HE} function with respect to the number of classes C for each test.

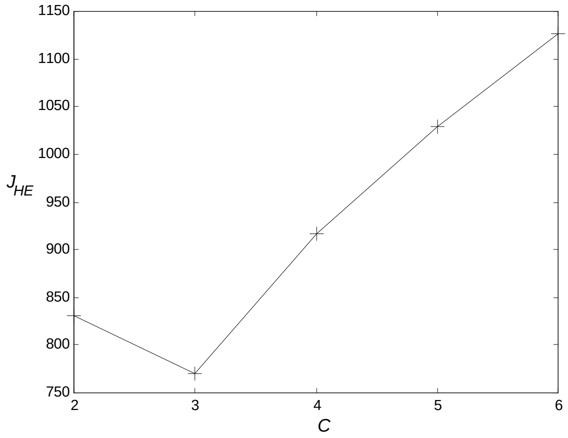

Figure 14 : *Evolution of J_{HE} with respect to C for test 1, C_{opt} =3.*

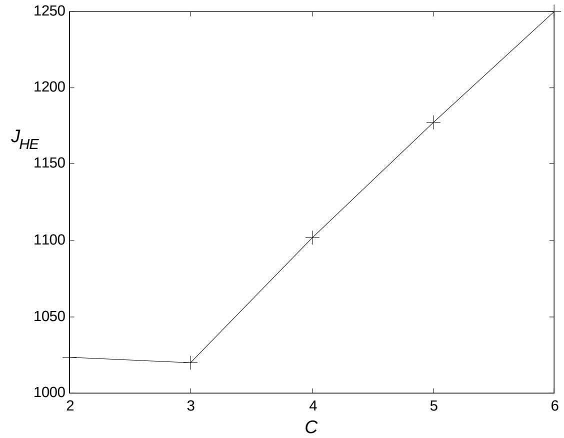

Figure 15 : *Evolution of J_{HE} with respect to C for test 2, C_{opt}* =3.

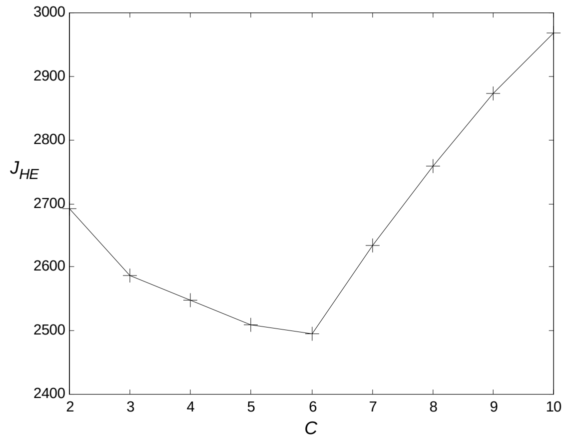

Figure 16 : *Evolution of J_{HE} with respect to C for test 3, C_{opt}* =6.

The results obtained, for each case, show that the estimated optimal number of classes C_{opt} coincide with the real number C_{real} (i.e. $C_{opt} = C_{real} = 3$ for tests 1 and 2, $C_{opt} = C_{real} = 6$ for test 3). This confirms the good performances of the proposed J_{HE} criterion.

7 Conclusion

The unsupervised classification by the FCM algorithm requires the *a priori* determination of the number of classes and suffers from the initialization phase and the local optimums.

We have proposed in this paper a new approach to get around these three difficulties. The new approach is based on the evolutionary strategies and the entropy approach. We have proposed a new evolutionist fuzzy C-means algorithm. We have presented a real coding and have defined an adequate fitness function suitable for the behavior to be optimized. We have proposed a new mutation operator which have permitted to the algorithm to avoid local solutions and to converge rapidly to the global solution. Then, we have defined a new criterion for an optimal choice of the classes number. This criterion is based on the entropy approach.

The proposed approach was tested on several simulation examples. The experimental results obtained show the rapidity of convergence and the good performances of the presented classification method. The optimal number of the classes estimated by the proposed criterion coincide with the real number. The two problems of initialization and local optimums are discarded in the EFCM algorithm.

8 References

[1] Hamad, D., El Saad, S. and Postaire, J., G. "Algorithmes d'apprentissage compétitif pour la classification automatique". TISVA'98, Oujda, Maroc, 27-28 April 1998, pp. 159-164.

[2] Bezdek, J., C. "Cluster validity with fuzzy sets". J. Cybernetics, Vol.3, N°3, pp. 58-73, 1974.

[3] Glorennec, P., Y. "Algorithmes d'apprentissage pour systèmes d'inférence floue". Colloque Neuro-mimétisme, Lyon, juin 1994.

[4] Karayiannis, N., B., Bezdek, J., C. and Hathaway, R., J. "Repairs to GLVQ : a new family of competitive learning schemes". IEEE Trans. on Neural Networks, Vol.7, N°5, pp. 1062-1071, 1996.

[5] Bezdek, J., C. and Hathaway, R., J. "Optimization of fuzzy clustering criteria using genetic algorithms". In Proc. First IEEE Conf. Evolutionary Computation, Piscataway, NJ : IEEE Press, Vol.2, pp. 589-594, 1994.

[6] Kuncheva, L., I. and Bezdek, J., C. " selection of cluster prototypes from data by a genetic algorithm". EUFIT, pp.1683-1688, Aachen, Germany, september 1997.

[7] Sarkar, M. et *al.* " A clustering algorithm using an evolutionary-based approach". Pattern Recognition Letters, 1997.

[8] Hall, L., Q., Özyurt, I., B. and Bezdek, J., C. " Clustering with a genetically optimized approach ". IEEE Trans. on Evolutionary Computation, Vol.3, N°2, pp. 103-110, july 1999.

[9] Presberger, T. and Koch, M. " Comparison of evolutionary strategies and genetic algorithms for optimization of a fuzzy controller ". Proc. of EUFIT'95, Aachen, Germany, august 1995.

[10] Glorennec, P., Y. "Constrained optimization of FIS using an evolutionary method ". In Genetic Alogorithms and Soft Computing, F. Herrera and J. L. Verdegay Eds., Physica-Verlag, ISBN 3-7908-0956-X, 1996.

[11] Michalewicz, Z. "Genetic algorithms + data structures = evolution programs ".2nd Eds., Springer-Verlag, 1994.

[12] Schwefel, H., P. " Numerical optimization of computer models". Wiley, 1981.

[13] Fogel, D., B. " An introduction to simulated evolutionary optimization". IEEE Trans. on Neural Networks, Vol. 5, N°1, 1994.

[14] Renders, J., M., *Algorithmes génétiques et Réseaux de Neurones*. Editions HERMES, 1995.

[15] Hoffman, R. and Jain A., K. " Segmentation and classification of range images". IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 9, N°5, pp. 608-620, 1987.

[16] Nguyen, H., H. and Cohen, P. " Gibbs random fields, fuzzy clustering, and the unsupervised segmentation of textured images". CVGIP, Vol. 55, N°1, pp. 1-19, janvier 1993.

[17] Rissanen, J. " The minimal description length principle". Research Report RJ-8236, IBM US research centers, july 1991.

[18] Olivier, C., Jouzel, F., Matouat, A., El. and Courtellemont, P. " Un nouveau critère pour la sélection de l'ordre d'un modèle ". In GRETSI, pp. 451-454, Grenoble, 1997.

[19] Ruspini, E., H., "A new approach to clustering ". Information and Control, Vol. 15, N°1, pp. 22-32, juillet 1969.

[20] Mood, A., Graybill, F. and Boes, D. "Introduction to the theory of statistics ". 1974.

[21] Palubinskas, G., Descombes, X. and Kruggel, F. " An unsupervised clustering method using the entropy minimization". In ICPR, pp. 1816-1818, Australia, august 1998.