# Use of criteria of class validity with the Possibilistic C Means algorithm

A. DEVILLEZ - P. BILLAUDEL - G. VILLERMAIN LECOLIER
Laboratoire d'Automatique et de Microélectronique
Faculté des Sciences
Moulin de la Housse, B. P. 1039 - 51687 REIMS Cedex 2
FRANCE
E-mail : {arnaud.devillez ; patrice.billaudel ; gerard.villermain-lecolier}@univ-reims.fr

*Abstract :* - In this paper we propose to apply on possibilistic partitions, the calculus of the criteria of class validity usually used on fuzzy partitions. After some recalls on the Fuzzy C Means and Possibilistic C Means algorithms, we compute these criteria for four types of data including difficulties which are currently seen in classification. The results obtained by the Fuzzy C Means algorithm are compared with those given by the possibilistic algorithm.

*Key-Words :* - Fuzzy Clustering, Unsupervised Classification, Fuzzy C Means, Possibilistic C Means, Compacity and Separability, Classification Entropy, Fuzzy Hypervolume.

## 1 Introduction

Fuzzy clustering algorithms are used to divide a set of points into a known number of classes [10]. Some criteria allow to validate the obtained partition. They can be used on successive partitions that include an increasing number of classes. In which case they help to find the number of classes giving the best partition as for hard classification methods [8].

The criteria are often used with the Fuzzy C Means algorithm. Experiments have shown that the results are not always in accordance with the visual distribution of points. We propose to extend the application of these criteria to the Possibilistic C Means algorithm. In this paper we will compare the performances of these criteria when applied to partitions coming from Fuzzy and Possibilistic C Means.

## 2 Fuzzy C Means and Possibilistic C Means algorithms

### 2.1 Fuzzy C Means algorithm

Studied essentially by Bezdeck [1] this unsupervised classification method uses as a basic principle the formation from non labeled samples, of a number c of groups. Classes must contain as similar samples as possible, while samples of different groups must be as dissimilar as possible. This is rendered into least square criteria minimization:

$$J(U,V) = \sum_{j=1}^{n}\sum_{i=1}^{c}(u_{ij})^m.(d_{ij})^2$$

with the hypothesis :

$$u_{ij} \in [0,1] \ \forall i,\ \forall j \qquad \sum_{i=1}^{c} u_{ij} = 1 \ \forall j$$

$$0 < \sum_{j=1}^{n} u_{ij} < n \ \forall i$$

Matrix U and V are respectively the membership matrix and the matrix of the centroïds of the classes. The element $u_{ij}$ is the membership degree of the point j to the class i. The term $d_{ij}$ is the distance from point $x_j$ to centroïd $V_i$. The obtained classes have spherical shapes when the euclidean distance is used and an elliptical shape when the Mahalanobis distance is used. The variable m is the fuzzyfication degree which takes its values in the interval $[1,+\infty[$. We have used the value m=2 recommended in [4]. Figure 1 presents the algorithm we used.

For class i, the membership degree of sample $x_k$ and centroïd $V_i$ are given by the following expressions :

$$u_{ik} = \frac{1}{\sum_{j=1}^{c}\left(\frac{\|x_k - V_i\|^2}{\|x_k - V_j\|^2}\right)^{\frac{1}{m-1}}} \qquad V_i = \frac{\sum_{k=1}^{n}(u_{ik})^m x_k}{\sum_{k=1}^{n}(u_{ik})^m}$$

The algorithm stops when the partition becomes stable that is to say when it does no longer evolve between two successive iterations :

$$\left|U^{it} - U^{it+1}\right| < \varepsilon$$

ε is the convergence threshold. It is proved that the algorithm converges in all case but the local minima will have to be avoided by a judicious choice of the value of the convergence threshold.

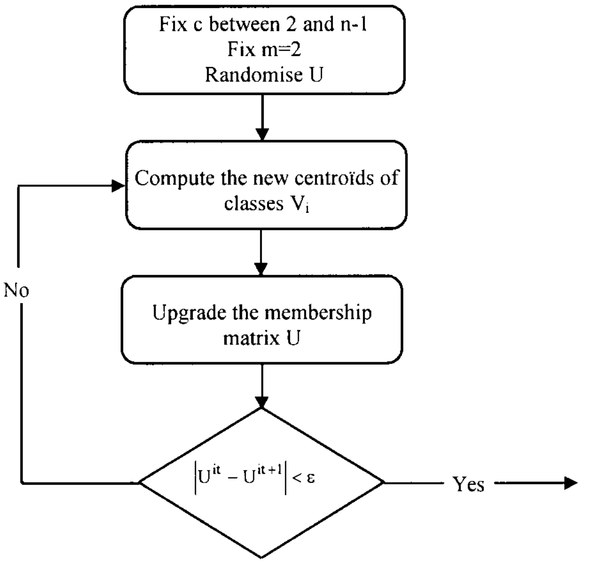



Fig.1 Fuzzy C Means algorithm

The probabilistic hypothesis of the fuzzy C means algorithm imposes that the sum of the membership degrees of each point should be unitary. The maximal membership degree of point can not be lower than 1/c whatever its distance to centroïds of classes. Consequently a point far away from classes has a maximal membership degree too high to be rejected in belonging. So it will be affected to a class.

The defuzzycation is realised by the maximum membership rule [9]. Each point is affected to the class for which it has the maximal membership degree.

### 2.2 Possibilistic C means algorithm

A solution to the problem of points which are far away from classes was proposed by Krishnapuram in 1993 [7]. It is based on a possibilistic approach. The probabilistic hypothesis is changed into an hypothesis warranting that each point belong to one class at least. The degree $u_{ij}$ do not reflect the membership degree anymore but the compatibility degree or membership possibility between a point and a class. The hypothesis become :

$$u_{ij} \in [0,1] \ \forall i,\ \forall j \qquad \max_{i=1,c}(u_{ij}) > 0 \ \forall j$$

$$0 < \sum_{j=1}^{n} u_{ij} \leq n \ \forall i$$

The new criteria to minimise becomes :

$$J(U,V) = \sum_{j=1}^{n}\sum_{i=1}^{c}(u_{ij})^m.(d_{ij})^2 + \sum_{i=1}^{c}\eta_i.\sum_{j=1}^{n}(1-u_{ij})^m$$

In this expression the first term corresponds to the criteria of the Fuzzy C Means. The second imposes the highest possible values to the membership possibilities . Parameter $\eta_i$ determines the distance from which the membership possibility is equal to 0.5. Krishnapuram proposes to use a value proportional to the average intraclass distance :

$$\eta_i = \frac{\sum_{j=1}^{n}(u_{ij})^m.(d_{ij})^2}{\sum_{j=1}^{n}(u_{ij})^m}$$

The algorithm is initialized with the partition matrix obtained by the Fuzzy C Means method. Centroïds and membership possibilities are computed with iterative method by the following expressions :

$$V_i = \frac{\sum_{j=1}^{n}(u_{ij})^m.x_j}{\sum_{j=1}^{n}(u_{ij})^m} \qquad u_{ij} = \left[1 + \left(\frac{(d_{ij})^2}{\eta_i}\right)^{\frac{1}{m-1}}\right]^{-1}$$

The calculus is achieved when the difference between each membership possibility and the possibility of the previous iteration is lower than a defined threshold ε. As for the Fuzzy C Means algorithm, the m value is equal to 2 and the defuzzycation is made according to the maximum membership rule.

## 3 Criteria of class validity

Before the defuzzyfication of U, it is necessary to validate the obtained partition by heuristic criteria. These were studied by Bezdeck and are shown in table 1.

The number of classes could be determined by making successive clusterings with increasing values of c. The best partition is obtained when criteria attain their minima. Consequently they allow to determine the number of classes existing in a set of points. Therefore the use of these criteria can be local only because they have a tendency to attain their global optima when c is near n-1 [5].

| Criteria | Expressions |
|---|---|
| Compacity and separability (CS) | $CS(U,c) = \frac{\pi}{s}$ |
| Classification entropy (H) | $H(U,c) = -\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{c} u_{ik}.\log(u_{ik})$ |
| Fuzzy hypervolume (FHV) | $F_{HV}(U,c) = \sum_{i=1}^{c}[\det(\sum_i)]^{(1/2)}$ |

Table 1 Criteria of class validity

In this table $\Sigma_i$ is the fuzzy covariance matrix defined by :

$$\sum_i = \frac{S_i}{\sum_{k=1}^{n}(u_{ik})^m}$$ where $S_i$ is the fuzzy dispersion matrix defined by :

$$S_i = \sum_{k=1}^{n}(u_{ik})^m.(x_k - V_i).(x_k - V_i)^T \quad i=1,c$$

Compacity $\pi$ and separability s are respectively defined by :

$$\pi = \frac{1}{N}.\sum_{i=1}^{c}\sum_{k=1}^{n}\left[u_{ik}.\|x_k - V_i\|\right]^2$$

$$s = \left[\min_{i=1,c;j=1,c;j\neq i}\|V_i - V_j\|\right]^2$$

# 4 Applications and comparative results

## 4.1 Presentation of the data

### 4.1.1 Plastic material data (PM)

Plastic bottles are generally made with one of the three following polymers : P.E.T., P.V.C., P.E.H.D.. To be recycled these different categories of bottles must be separated. Infrared spectrometry has been used to caracterize the different bottles. The analysis of the spectrum of each material has enabled to select 7 discriminating wavelengths. The study of the correlations has allowed us to reduce this number to 2. We have 90 transmission values for the two wavelengths $\lambda_1$ and $\lambda_2$. These values are distributed into three classes of 30 points as it is shown in figure 2.

The classes are well separated. They have elongated and tilted shapes. Classes include empty areas which can raise some problems for the determination of the number of classes.

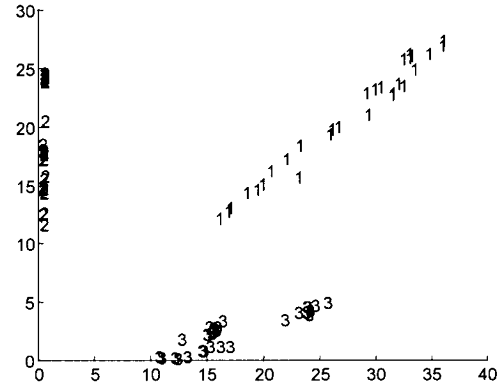
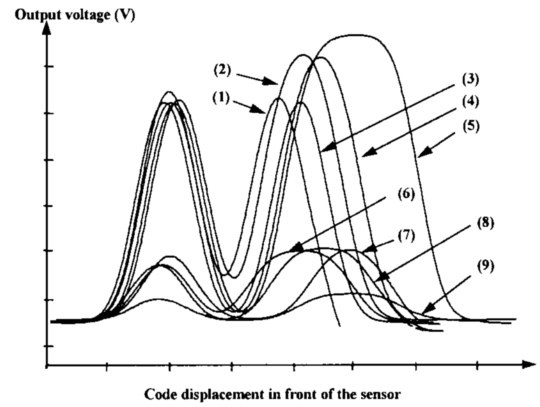
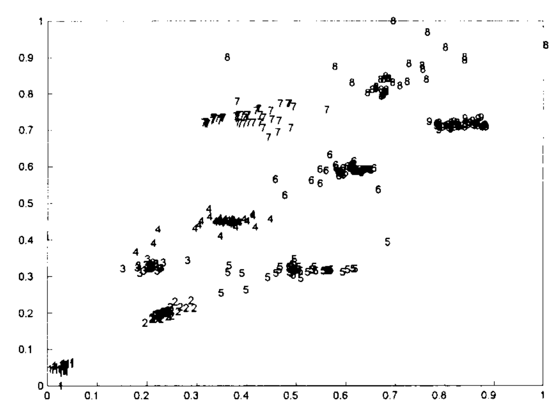


Fig.2 Representation of the 3 classes of plastic materials :1 PVC, 2 PET, 3 PEHD

### 4.1.2 Metallic code data (MC)




a



b

Fig.3 Response of the dispositive to the 9 codes (a) and representation of the 9 classes (b)

The Laboratory of Automatic and Microelectronic has developed a device using a sensor with eddy currents to read the codes made by a succession of metallic fine bands, separated by an insulating zone [6]. Figure 5a shows the response of the device to the 9 codes defined by the manufacturer. Two characters were selected among the 52 computed, for their power of discrimination. Figure 4b shows the 9 classes of 45 points that we obtained. The classification difficulty of these data comes from the different densities of the different classes.

#### 4.1.3 Iris data (ID)

Those data are widely used for the comparison of classifiers. They are composed of three classes of flowers : IRIS Setosa, Versicolor and Virginia. Each class is constituted of 50 samples characterized by 4 attributes : length and width of the bot sepals, length and width of the petals. Two classes among the three are not well separated.

#### 4.1.4 Washing Machine data(WM)

These data come from a study achieved in the Laboratory of Automatic and Microelectronic. The aim is to detect the unbalance failures in a washing machine by measuring the movements coming from the vibrations of the machine during its washing process [2]. Measures are realised by eddy current sensors. We obtain a set of 204 points including 4 classes. Each class corresponds to an unbalance and to its position in the tub. The set of points is shown on figure 4. Classes are elongated and tilted. Two of them are not well separated.

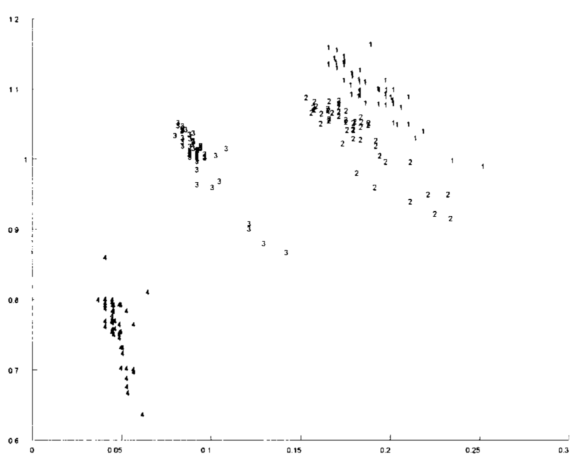


Fig.4 WM data

### 4.2 Results of the Fuzzy C Means algorithm

For the PM and WM data we have used the Mahalanobis distance because of the elongated and tilted shape of the classes. For the other serials of data we have retained the euclidean distance. Tables 2-5 present the values of the criteria. The value of the convergence threshold is 0.01 for all data.

| c | CS(U,c) | H(U,c) | $F_{HV}(U,c)$ |
|---|---|---|---|
| 2 | 0.0032 | 0.1366 | 94.1043 |
| 3 | **0.0014** | **0.1151** | 27.8385 |
| 4 | 0.0055 | 0.1653 | 34.8269 |
| 5 | 0.0025 | 0.2003 | 27.5009 |
| 6 | 0.0081 | 0.2427 | **14.5998** |
| 7 | 0.0063 | 0.2724 | 25.0723 |

Table 2 Criteria of class validity for MP data

| c | CS(U,c) | H(U,c) | $F_{HV}(U,c)$ |
|---|---|---|---|
| 2 | **0.0542** | **0.0851** | 0.0216 |
| 3 | 0.1371 | 0.1719 | **0.0205** |
| 4 | 0.1958 | 0.2440 | 0.0209 |
| 5 | 0.3998 | 0.3073 | 0.0210 |
| 6 | 0.3281 | 0.3552 | 0.0217 |
| 7 | 0.3733 | 0.3943 | 0.0190 |

Table 4 Criteria of class validity for ID data

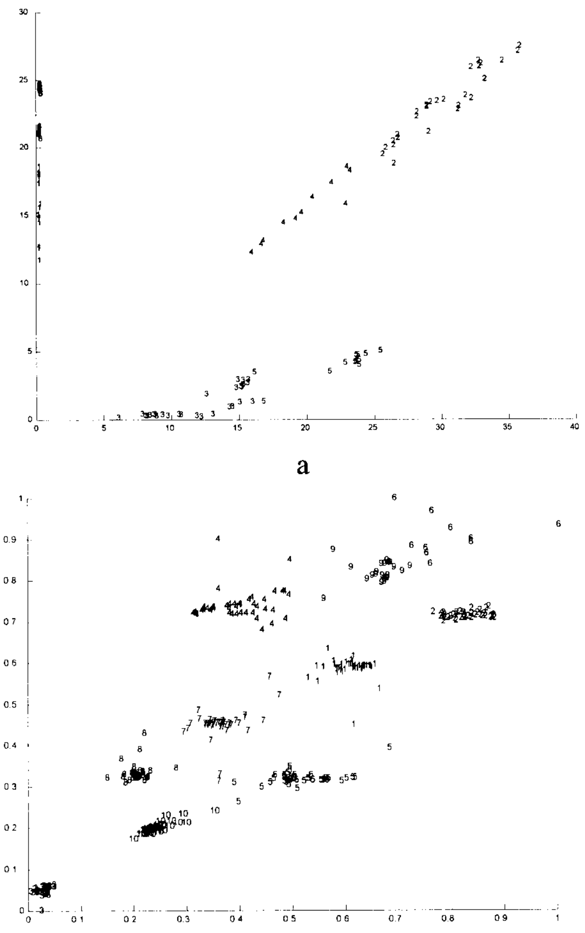
| c | CS(U,c) | H(U,c) | $F_{HV}(U,c)$ |
|---|---|---|---|
| 7 | 0.1234 | 0.2650 | 0.0153 |
| 8 | 0.1575 | 0.2029 | 0.0118 |
| 9 | 0.0872 | **0.1750** | **0.0081** |
| 10 | **0.0698** | 0.1821 | 0.0088 |
| 11 | 0.7868 | 0.2119 | 0.0090 |
| 12 | 0.4070 | 0.2082 | 0.0087 |

Table 3 Criteria of class validity for MC data

| c | CS(U,c) | H(U,c) | $F_{HV}(U,c)$ |
|---|---|---|---|
| 2 | **20.3545** | 0.1128 | 0.0035 |
| 3 | 72.58417 | **0.0645** | 0.0013 |
| 4 | 149.2862 | 0.0954 | **0.0011** |
| 5 | 203.0406 | 0.1631 | 0.0019 |
| 6 | 27.5606 | 0.1728 | 0.0019 |
| 7 | 360.1878 | 0.2335 | 0.0021 |

Table 5 Criteria of class validity for WM data

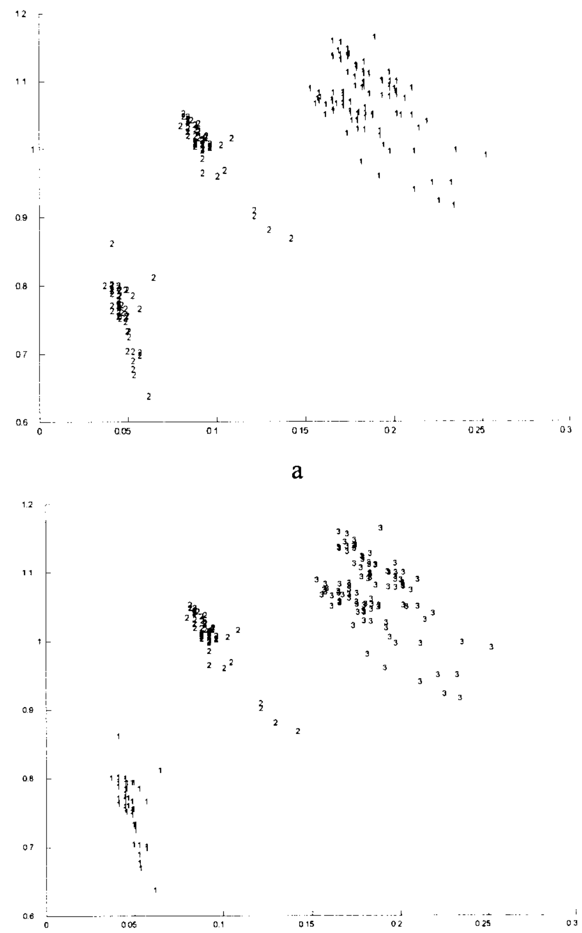
For PM data the first two criteria give 3 classes as expected. On the other hand the fuzzy hypervolume criterion reaches its optimum for 6 classes. We can explain this result ; indeed we can obtain 6 classes by dividing each class into two subclasses. This clustering is presented on figure 5a.



a

b

Fig.5 Clustering of PM data into 6 classes (a) and of MC data into 10 classes (b)



a

b

Fig.6 Clustering of the WM data into 2 (a) and 3 classes (b)

Concerning MC data two criteria reach their optima for 9 classes which is true to reality. The compacity separability gives 10 classes. This result comes from the low compacity of class 6. The clustering into 10 classes is presented on figure 5b.

For ID data, we notice that the optima are reached for 2 or 3 classes. These results are true to reality, indeed two of the three classes overlap themselves. The fuzzy hypervolume gives the best result. For WM data criteria give 2, 3 and 4 classes. The clusterings are presented on figure 6. For the clustering into 4 classes the algorithm does not make any classification mistake. Clustering into 3 classes is obtained by merging classes 1 and 2, which are not well separated. The fusion of classes 3 and 4 gives the clustering into 2 classes.

During our tests we noticed that the results of validity criteria were not always reliable when they were used with the Fuzzy C Means algorithm. Indeed all of them do not converge on the same number of classes. The best criterion seems to be the fuzzy hypervolume.

### 4.3 Results of the Possibilistic C Means algorithm

The possibilistic algorithm is initialised in all cases with the membership matrix coming from the Fuzzy C Means algorithm. The convergence threshold is equal to 0.7. We computed the criteria of validity of classes for this algorithm. The results are shown in tables 6-9.

We notice that for PM and CM data, the criteria give the real number of classes existing in the set of points. For ID data, the proximity of two classes among the three explains that two of the three criteria attain their optima for 2 classes. However the classification entropy reaches its minimum for 3 classes. For WM data, the criteria reach their optimal values for 3 classes. As for ID data, the possibilistic algorithm has a tendency to merge two neighbour classes. Only the fuzzy hypervolume attains a second optimal value in 4 classes.

| c | CS(U,c) | H(U,c) | $F_{HV}(U,c)$ |
|---|---|---|---|
| 2 | 0.1447 | 0.2112 | 47.9712 |
| 3 | **0.0106** | **0.1543** | **5.8804** |
| 4 | 0.0803 | 0.2417 | 28.9280 |
| 5 | 0.0564 | 0.2520 | 12.0897 |
| 6 | 0.1542 | 0.2883 | 9.0676 |
| 7 | 0.6524 | 0.3002 | 15.3853 |

Table 6 Criteria of class validity for MP data

| c | CS(U,c) | H(U,c) | $F_{HV}(U,c)$ |
|---|---|---|---|
| 2 | **0.0146** | **0.1761** | **0.0123** |
| 3 | 0.0657 | 0.2664 | 0.0157 |
| 4 | 0.1341 | 0.3505 | 0.0195 |
| 5 | 0.1878 | 0.4190 | 0.0251 |
| 6 | 0.3037 | 0.4655 | 0..0243 |
| 7 | 0.4654 | 0.5722 | 0.0325 |

Table 8 Criteria of class validity for ID data

| c | CS(U,c) | H(U,c) | $F_{HV}(U,c)$ |
|---|---|---|---|
| 7 | 0.0745 | 0.4104 | 0.0173 |
| 8 | 0.0631 | 0.3023 | 0.0084 |
| 9 | **0.0238** | **0.2481** | **0.0040** |
| 10 | 1.1389 | 0.3106 | 0.0060 |
| 11 | 2.9399 | 0.3379 | 0.0069 |
| 12 | 3.3606 | 0.3884 | 0.0068 |

Table 7 Criteria of class validity for MC data

| c | CS(U,c) | H(U,c) | $F_{HV}(U,c)$ |
|---|---|---|---|
| 2 | 0.0597 | 0.1799 | 0.0014 |
| 3 | **0.0154** | **0.1447** | **0.0004** |
| 4 | 0.0567 | 0.1834 | **0.0004** |
| 5 | 0.3533 | 0.2862 | 0.0015 |
| 6 | 0.8938 | 0.3500 | 0.0014 |
| 7 | 0.5309 | 0.3127 | 0.0007 |

Table 9 Criteria of class validity for WM data

# 5 Conclusion

The tests we have made show that the criteria of classes validity have more performance when they are used with the Possibilistic C Means algorithm. This result can be explained by the principle of the algorithm which consists in decreasing the similarity degrees of points which are far away from the class. Consequently the influence of these points in the calculus of the criteria is decreased. So the used criteria with the possibilistic approach allow to determine the number of existing classes in the set of points without having to divide classes of high density, such as the class 8 of MC data.

However the Possibilistic C Means algorithm does not allow to distinguish two neighbour classes. That is the case of ID and WM data. For this type of set of points the use of the fuzzy hypervolume criterion with the Fuzzy C Means algorithm would be better. Finally these methods allow to determine the exact number of clusters only if the data include elliptical or spherical classes. The use of the criteria of validity of classes with the Possibilistic C Means algorithm can constitute a good solution to the problem of the determination of the number of existing classes in a set of points [3].

*References :*


[1] J.C. BEZDEK, *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New-York, 1981.

[2] P. BILLAUDEL, F. KLEIN, E. PERRIN, G. VILLERMAIN LECOLIER, Influence des défauts sur le mouvement d'une machine tournante : étude de faisabilité, *JESA*, Vol. 31, n°2, pp 373-391, 1997.

[3] P. BILLAUDEL, A. DEVILLEZ, G. VILLERMAIN LECOLIER, An unsupervised fuzzy classification algorithm for the non elliptic classes, *EURISCON'98*, June 22 - 25, 1998, Athens, Greece.

[4] H. CHOE et J.B. JORDAN, On the optimal choice of parameters in the fuzzy C-means algorithm, *Proc. of fuzzy IEEE 92*, pp 349 354, 1992.

[5] C. FRELICOT, *Un système adaptatif de diagnostic prédictif par reconnaissance de formes floue*, Thèse en contrôle des systèmes, Université Technologique de Compiègne, 1992.

[6] M. GOUTIER, *Optimisation de la tête de détection d'un capteur à balance d'induction. Identification de la réponse d'un code métallique*, Thèse, Université de Reims, 1995.

[7] R. KRISHNAPURAM, J. M. KELLER, A possibilistic approach to clustering, *IEEE Transactions on Fuzzy Systems*, Vol. 15, n°4, pp 98-110, 1993.

[8] G. W. MILLIGAN, M. C. COOPER, An examination of procedures for determining the number of clusters in data set, *Psychometrica*, Vol. 50, pp 159-179, 1985.

[9] PAL S., Fuzzy sets and decision making approaches in vowel speaker recognition, *IEEE Transaction on Systems, Man and Cybernetics*, pp. 625-629, 1977.

[10] E. H. RUSPINI, A new approach to clustering, *Information and Control*, Vol. 15, n°1, pp 22-32, 1969.