ELSEVIER

Fuzzy Sets and Systems 128 (2002) 323–338

FUZZY sets and systems

www.elsevier.com/locate/fss

A fuzzy hybrid hierarchical clustering method with a new criterion able to find the optimal partition

Arnaud Devillez[a,*], Patrice Billaudel[b], Gérard Villermain Lecolier[c]

[a] *LPMM-UMR CNRS 7554, Institut Supérieur de Génie Mécanique et Productique, Ile du Saulcy, 57045 Metz, Cedex 01, France*

[b] *Laboratoire d'Automatique et de Microélectronique, Institut de Formation Technique Supérieure, 7, boulevard Jean Delautre, 08000 Charleville, Mezieres, France*

[c] *Laboratoire d'Automatique et de Microélectronique, Faculté des Sciences, Moulin de la Housse, B.P. 1039-51687 Reims, Cedex 2, France*

Received 27 June 2000; received in revised form 6 April 2001; accepted 29 August 2001

Abstract

Classical fuzzy clustering methods are not able to compute a partition in a set of points when classes have nonconvex shape. Furthermore we know that in this case, the usual criteria of class validity such as fuzzy hypervolume or compactness–separability, do not allow to find the optimal partition. The purpose of our paper is to provide a clustering method able to divide a set of points into nonconvex classes without knowing a priori their number. We will show that it is possible to reconcile a fuzzy clustering method with a hierarchical ascending one while maintaining a fuzzy partition by a method called unsupervised fuzzy graph clustering. To that effect, we shall use the Fuzzy C-Means algorithm to divide the set of points into an overspecified number of subclasses. A fuzzy relation is then established between them in order to extract the structure of the set of points. It can be represented by a graduated hierarchy. Finally, we present a new criterion to find the cut of the hierarchy giving the optimal regrouping. This one allows to find the real classes existing into the set of points. The given results are compared with those obtained by other classical cluster validity criteria and we propose to study the influence of the number of initial subclasses on the final computed partition. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Cluster validity; Fuzzy clustering; Hierarchical classification; Optimal partition

1. Introduction

Our works deal with unsupervised classification methods used to identify the functioning modes of industrial processes. In this field classes may have nonconvex shape. A lot of works deal with supervised and unsupervised classification methods for classes of elliptic shape. Nagy has provided the problem of classes of

* Corresponding author. Tel.: +33-3-87-54-70-09; fax: +33-3-87-31-53-66.
E-mail address: devillez@lpmm.sciences.univ-metz.fr (A. Devillez).

PII: S0165-0114(01)00187-7

nonconvex shape in 1965 [20]. Since a lot of works dealing with hierarchical methods [22] and structurally constrained clustering [10] has been proposed. But we know that fuzzy clustering methods give good results when the shape of the classes is elliptic and when these classes are separable by an hyperplane. In any other cases it is absolutely necessary to use a classic hierarchical method. This one has several deficiencies. Indeed what we obtain as a result is a hard partition of the set of points, where the notion of graduated membership does not exist. Moreover, the hierarchical classification methods require a very important memory space when the number of samples increases, which makes its applications difficult in real time. Also the addition of a point changes possibly in a major way, the structure of the graduated hierarchy what may entail some important calculations. To make up for these inconveniences, we have shown that it is possible to reconcile the fuzzy techniques of clustering with those of graduated hierarchies [4,13]. The purpose of our paper is to present the method we have developed from the creation of the subclasses up to their fusion which implies the use of a proximity graph built according to a graduated hierarchy. The major difficulty of such a method comes from the determination of the optimal cut of the hierarchy. In order to compute it we introduce a new criterion. We will see that this one allows to find the number of classes existing in the set of points. To finish, we will apply our method called unsupervised fuzzy graph clustering (UFGC) on several examples in order to show its ability to compute a fuzzy partition into a set of points without knowing the number of classes a priori.

2. Division of the set of points into subclasses

The first stage of our algorithm consists in using the Fuzzy C-Means algorithm [4] to divide the set into c' subclasses. This principle has been used by Frigui and Krishnapuram in the URCP algorithm [11]. The result of this algorithm is a fuzzy membership matrix U'. The element u'_{kj} represents the membership degree of sample x_j to subclass SC_k. Each subclass must belong to one real class only.

Therefore, the number of subclasses must be much greater than the number of real classes. In our example presented in Fig. 1, c' must be higher than 3. It is possible to use some heuristic criteria to choose c'. We used the fuzzy hypervolume, compactness and separability criteria to that effect [6]. The partition we obtained into 12 subclasses for our example, is presented in Fig. 1. We can note that each subclass belongs to one real class only.

3. Similarity degrees between subclasses

In [8], we studied the notion of hard neighbourhood between subclasses. We noted that our algorithm includes a major restriction: the initial set of points must not present ambiguous points otherwise, we will observe fusions that should not exist between subclasses. The clustering algorithm based on a hard graph endures a lack of hardiness that can greatly decrease its performances. That is why we have introduced the concept of fuzzy neighbourhood. The neighbourhood or proximity between two subclasses can be quantified by a coefficient. The one we are using is the similarity degree defined by Frigui and Krishnapuram [11].

$$n_{kl} = 1 - \frac{\sum_{x_i \in \mathrm{SC}_k \text{ or } x_i \in \mathrm{SC}_l} |u'_{ik} - u'_{il}|}{\sum_{x_i \in \mathrm{SC}_k} u'_{ik} + \sum_{x_i \in \mathrm{SC}_l} u'_{il}}.$$

We obtain a neighbourhood matrix N, the size of which is $c' \times c'$. The element n_{kl} corresponds to the proximity value between subclass SC_k and subclass SC_l. Nearer the value of n_{lk} is to 1, the closer the subclasses are to one another. The terms on the diagonal are all equal to 1 because each subclass is a neighbour of itself. Furthermore, the matrix N is symmetrical.

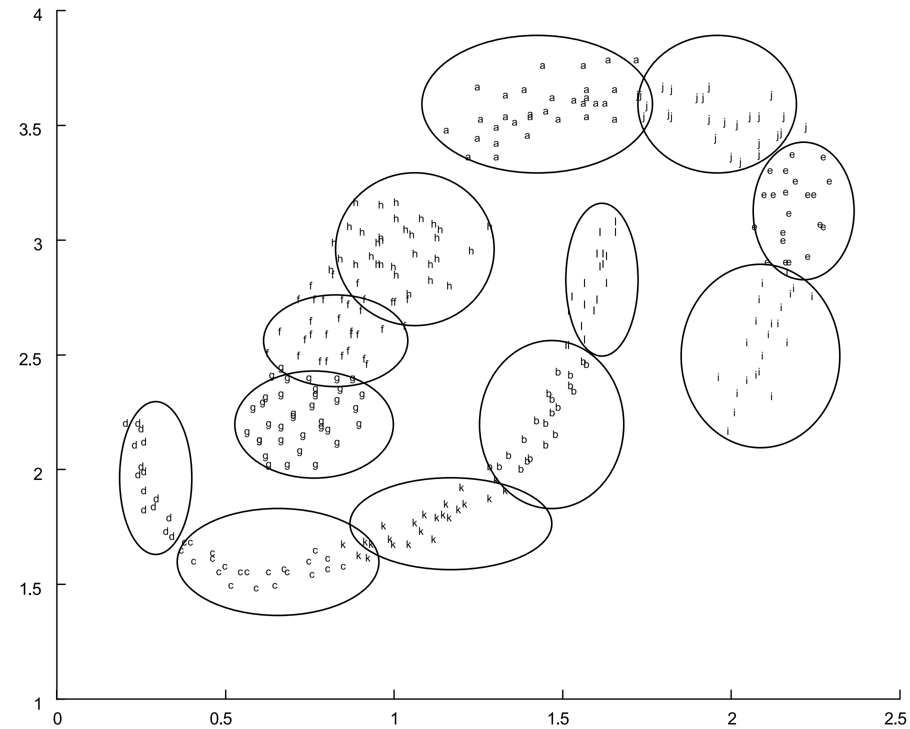

Fig. 1. Division of the set into 12 subclasses. All points marked by a similar letter belong to the same subclass.

The obtained matrix N, which is calculated by using the fuzzy similarity degree for our example, is the following:

$$N = 10^{-2}.\begin{pmatrix} 100 & 1.44 & 0.53 & 0.68 & 3.42 & 2.00 & 1.06 & 5.12 & 1.96 & 12.5 & 0.73 & 4.31 \\ 1.44 & 100 & 2.05 & 1.60 & 1.71 & 3.74 & 4.04 & 2.82 & 5.79 & 1.22 & 9.40 & 8.90 \\ 0.53 & 2.05 & 100 & 11.5 & 0.43 & 1.71 & 4.90 & 0.95 & 0.85 & 0.40 & 10.5 & 0.86 \\ 0.68 & 1.60 & 11.5 & 100 & 0.45 & 2.79 & 8.35 & 1.36 & 0.78 & 0.46 & 3.09 & 0.91 \\ 3.42 & 1.71 & 0.43 & 0.45 & 100 & 0.99 & 0.71 & 1.51 & 10.8 & 15.1 & 0.70 & 5.00 \\ 2.00 & 3.74 & 1.71 & 2.79 & 0.99 & 100 & 14.5 & 15.6 & 1.52 & 1.05 & 2.40 & 3.42 \\ 1.06 & 4.04 & 4.90 & 8.35 & 0.71 & 14.5 & 100 & 3.29 & 1.25 & 0.69 & 5.70 & 1.96 \\ 5.12 & 2.82 & 0.95 & 1.36 & 1.51 & 15.6 & 3.29 & 100 & 1.80 & 1.86 & 1.37 & 5.87 \\ 1.96 & 5.79 & 0.85 & 0.78 & 10.8 & 1.52 & 1.25 & 1.80 & 100 & 2.97 & 1.72 & 7.26 \\ 12.5 & 1.22 & 0.40 & 0.46 & 15.1 & 1.05 & 0.69 & 1.86 & 2.97 & 100 & 0.58 & 3.68 \\ 0.73 & 9.4 & 10.5 & 3.09 & 0.70 & 2.40 & 5.70 & 1.37 & 1.72 & 0.58 & 100 & 1.68 \\ 4.31 & 8.93 & 0.86 & 0.91 & 5.00 & 3.42 & 1.96 & 5.87 & 7.26 & 3.68 & 1.68 & 100 \end{pmatrix}.$$

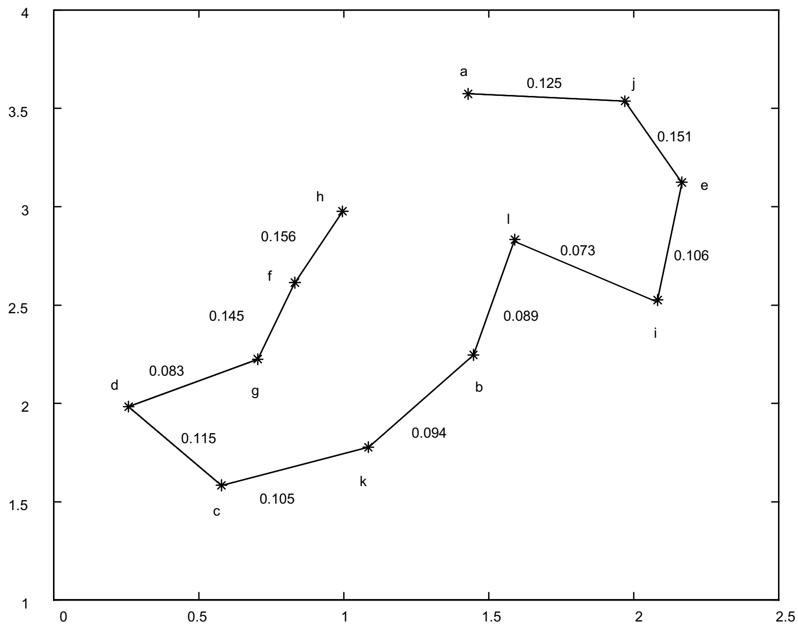

Fig. 2. Reduced fuzzy proximity graph for our example.

4. Construction of the fuzzy proximity graph and of the dendogram

4.1. Fuzzy proximity graph and reduced graph

The neighbourhood matrix N defines a fuzzy proximity graph in which the vertices represent the subclasses and the arcs represent the links. The arcs are graduated by the proximity degrees. The graph corresponding to the neighbourhood matrix of our example includes 12 vertices and 66 arcs, knowing that the arcs linking a subclass to itself are not represented and that matrix N is symmetrical. The representation of a graph of this size is not comfortable. That is why it is possible to associate a reduced graph representing the highest $c'-1$ proximity degrees linking the c' vertices. In our case, it is a chain linking the 12 vertices as it is shown in Fig. 2. We note that this graph reflects structures existing in the set of points. Calculations are based on matrix N, consequently, they take account of the complete graph. Furthermore it is possible to define the remoteness matrix F. It is the complementary matrix to matrix N where $f_{kl} = 1 - n_{kl}$ whatever k and l. The higher the value of f_{kl} is, the more distant subclasses SC_k and SC_l are from each other. Graph and graduated hierarchy associated to matrix F are, respectively, shown in Figs. 2 and 3.

4.2. Dendogram

We call dendogram or graduated hierarchy a tree the leaves of which are samples subject to classification. The hierarchy is graduated if any part h is associated to a numerical value $v(h) \geqslant 0$ compatible with the following relation: If $h \subset h'$ then $v(h) < v(h')$.

A hierarchy is associated to the graph shown in Fig. 2. It is graduated according to the remoteness values of matrix F. Two subclasses or two sets of subclasses are associated at each level. This operation is repeated until all subclasses are merged into one class only. For our example, we have obtained the dendogram shown in Fig. 3.

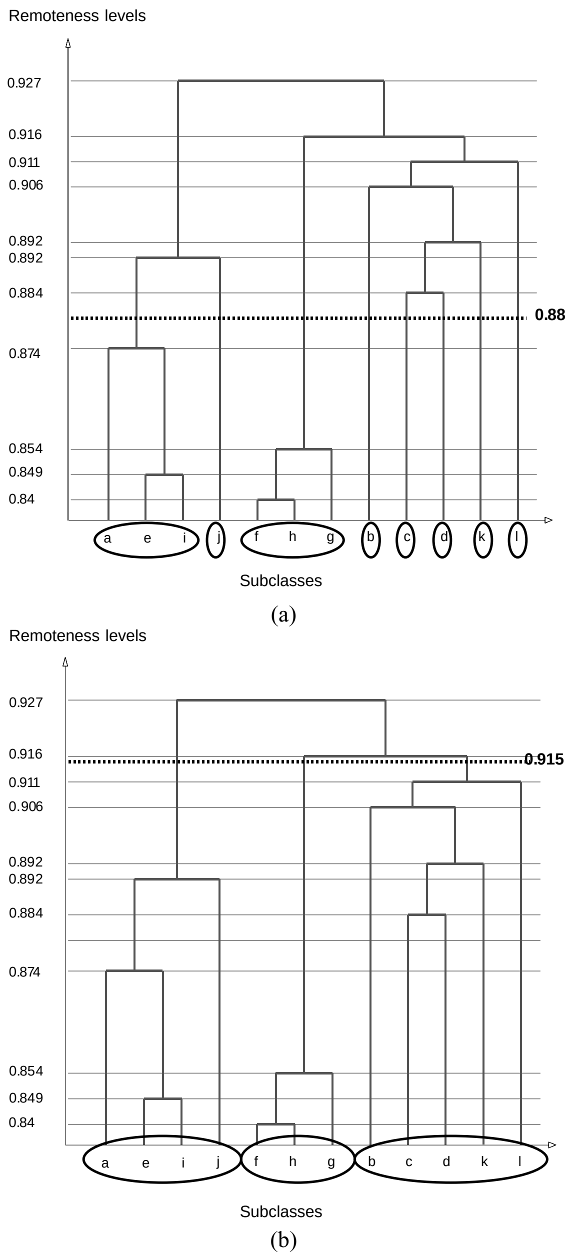

Fig. 3. Dendograms for our example for the two levels of cutting: (a) $\beta = 0.88$ and (b) $\beta = 0.915$.

5. Defuzzyfication of the proximity graph

5.1. Level of the cut of the graduated hierarchy

To recover real classes, it is absolutely necessary to cut the dendogram or the graduated hierarchy at a level β. For example the two levels $\beta = 0.88$ and 0.915 give the sets of subclasses shown in Fig. 3. We note

$n^{\alpha}_{kl} = 1$ if $n_{kl} \geq \alpha$ and $n^{\alpha}_{kl} < \alpha$.

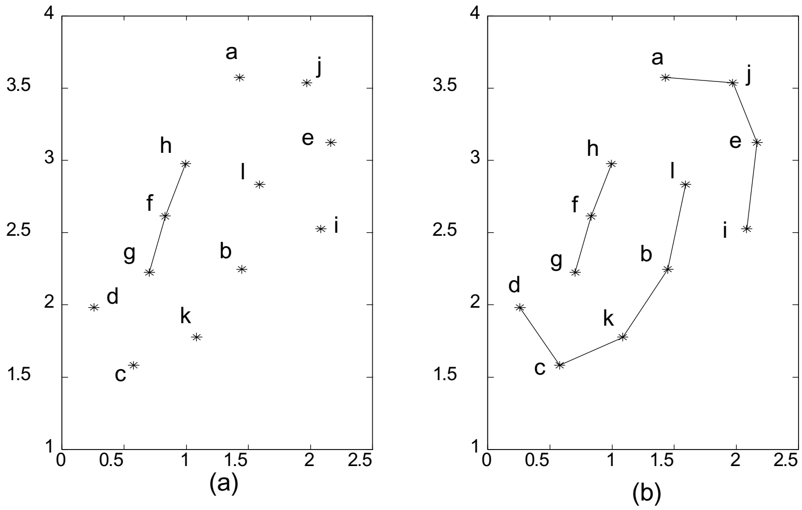

Fig. 4. Graphs obtained by the relations of different levels: α_1 (a); and α_2 (b) for our example.

that these sets are completely different from one cut to another. Consequently the final partition depends on the level of the cut.

5.2. Existence of a fuzzy order relation

The neighbourhood matrix N defines a fuzzy relation [9] which expresses the idea of neighbourhood between subclasses. It can be decomposed into a series of relations of level α. If N is a fuzzy order relation where $n_{kl} \in [0, 1]$, the relation of level α which is associated to N is defined by N^{α} such as

$$n^{\alpha}_{kl} = 1 \quad \text{if } n_{kl} \geqslant \alpha \text{ and } n^{\alpha}_{kl} < \alpha.$$

This decomposition entails the defuzzyfication of the proximity graph into a hard graph which determines the structure of the set of points. The remaining arcs are no longer graduated. Two subclasses are then considered as neighbouring and consequently merged if an arc exists between the two vertices which correspond to them. For the two levels of cut $\alpha_1 = 0.12$ and $\alpha_2 = 0.085$, we have obtained the two hard neighbourhood graphs shown in Fig. 4. We can note that the connected components of these graphs correspond to the sets of subclasses obtained by the cuts of the dendogram. Furthermore, values α_1 and α_2 are, respectively, complementary to values β_1 and β_2. Consequently, the relation of level α gives the same result than the cut of level β into the dendogram. In our exercises, we have used the fuzzy relation defined by matrix N to recover the structure of the set of points.

6. Fusion of subclasses

The last stage of the algorithm consists in searching for the connected components [14] of the hard proximity graph or sets of subclasses in the dendogram. For each point we have at our disposal a membership degree for each subclass. It is necessary to achieve a fusion in order to obtain a membership degree for each real class. This fusion is made by a limited sum [18]. When a number s_i of subclasses SC_k exists within a class C_i, the membership function of a point x to this class is defined by

$$u_i(x) = \min\left[1, \sum_{k=1}^{s_i} u'_{\mathrm{SC}_k}(x)\right].$$

The new membership matrix is calculated from the matrix U' resulting from the Fuzzy C-Means algorithm. A sum is achieved between lines corresponding to the subclasses which must be merged. We obtain a membership matrix U the size of which is $c \times n$ with c being the number of connected components and n the number of points. The element u_{ij} represents the membership degree of the point x_j to the new class C_i. The obtained membership matrix depends on the chosen level of the cut.

7. Research on the optimal cut

Few criteria can help us to determine the cut to be made into the graduated hierarchies. Those proposed in [4,19] give good results with classes of spherical or elliptical shapes. With classes of complex shapes, there exists no adequate criteria to find the optimal partition. The criterion we propose, uses the principle of the compactness criterion defined in [25]. The global compactness of a partition is defined by

$$\mathrm{co}_{\mathrm{gl}} = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij}.d_{ji})^2,$$

where d_{ji} represents the distance from the point x_j to the nearest centroid of class C_i because of the multi-prototype approach used in our algorithm. $\mathrm{co}_{\mathrm{gl}}$ allows one to measure the global compactness of the set of points.

According to the same model, we define a compactness measure for each class C_i:

$$\mathrm{co}_{C_i} = \frac{\sum_{j=1}^{n}(u_{ij}.d_{ji})^2}{|C_i|} \quad \text{where } |C_i| \text{ is the fuzzy cardinal defined by } |C_i| = \sum_{j=1}^{n} u_{ij}.$$

The average compactness of the created classes into the set of points is defined by the following expression:

$$\mathrm{co}_{\mathrm{av}} = \frac{1}{c}\sum_{i=1}^{c}\mathrm{co}_{C_i} = \frac{1}{c}\sum_{i=1}^{c}\frac{\sum_{j=1}^{n}(u_{ij}.d_{ji})^2}{\sum_{j=1}^{n}u_{ij}}.$$

If the number of classes c is equal to 1, the global compactness and the average compactness are equivalent. Indeed $u_{ij} = 1\ \forall j$ which induces

$$\sum_{j=1}^{n} u_{ij} = n.$$

Consequently we have

$$\mathrm{co}_{\mathrm{av}} = \frac{1}{n}\sum_{j=1}^{n}(u_{ij}.d_{ji})^2 \quad \text{that is to say: } \mathrm{co}_{\mathrm{av}} = \mathrm{co}_{\mathrm{gl}}.$$

Our study concerns the general case where the number of researched classes is higher than 1. The case where c is equal to 1 remains marginal and most of the criteria do not function for this value. The global compactness defined by Frigui and Krishnapuram represents the average distribution of the points around the centroids of all the classes according to the membership levels. The considered cardinal using for calculating the average value is the total number of points n. The compactness we have defined for one class is based on the same model. It represents the average distribution of the points around the centroids of one class. The considered cardinal is now the fuzzy cardinal $|C_i|$ of class C_i. Then the average compactness is computed from the compactness of all the classes in order to obtain a value which translates the distribution of the points around the centroids of the classes. These definitions are equivalent on condition that the partitioning is optimal that is to say when the computed clusters correspond to the real classes. The definition of our criteria is based on this hypothesis of equivalence between average and global compactness when the optimal partition into c^* classes is attained. In this case, its ratio tends towards 1 and we propose the minimisation of the following criterion for determining the number of classes existing in the set of points:

$$K_c = \left|1 - \frac{\mathrm{co}_{\mathrm{av}}}{\mathrm{co}_{\mathrm{gl}}}\right| = \left|1 - \frac{1/n \sum_{i=1}^{c} \sum_{j=1}^{n} (u_{ij}.d_{ji})^2}{1/c \sum_{i=1}^{c} [\sum_{j=1}^{n} (u_{ij}.d_{ji})^2 / \sum_{j=1}^{n} u_{ij}]}\right|.$$

We will show that this criterion functions for the case where all the classes include the same number of points. It is supposed that the optimal partition is attained for a number c^* of classes. The membership degrees tending towards 1 give the following property:

$$\sum_{j=1}^{n} u_{ij} \to n_i \quad \text{with } n_i \text{ the number of points into each class } C_i.$$

This tendency induces that

$$\mathrm{co}_{\mathrm{av}} \to \frac{1}{c^*} \sum_{i=1}^{c^*} \frac{\sum_{j=1}^{n} (u_{ij}.d_{ji})^2}{n_i}. \tag{1}$$

As all the n_i are equal, the average compactness becomes

$$\mathrm{co}_{\mathrm{av}} \to \frac{1}{c^*.n_i} \sum_{i=1}^{c^*} \sum_{j=1}^{n} (u_{ij}.d_{ji})^2 \quad \text{that is to say: } \mathrm{co}_{\mathrm{av}} \to \mathrm{co}_{\mathrm{gl}}. \tag{2}$$

If the partition includes too many classes, relation (1) is not verified and the criterion cannot reach the optimal value. Indeed subclasses are very near one from each other and because of the orthogonality condition, the membership degrees u_{ij} are very lower than the value 1. If the partition does not include enough classes, relation (1) is verified. The additional fusions which appeared, unbalance the sizes of classes. Relation (2) is no more verified, which does not allow one to find the optimal number of classes.

These results imply that when the optimal partition is attained, the criterion K_c tends toward 0. The minimisation of the criterion allows one to find the optimal partition of the set of points and consequently to determine the number of classes existing in it on condition that all of them would have the same number of points. We have verified experimentally that this criterion functions even if classes do not have the same number of points.

Table 1 illustrates the results we obtained for our example. We can note that the minimum is attained for three classes. The corresponding partition is presented in Fig. 5. The obtained classes are similar to the real classes.

Table 1
Values of the criterion K_c as a function of the number of classes

$\alpha \times 10^{-2}$	8	8.5	9	10	10.5	11	12	14	15	15.5	16
$K_c \times 10^{-2}$	0.5	0.25	6.48	9.46	12.2	8.6	11	6.47	3.22	1.37	1
Number of clusters	2	3	4	5	6	7	8	9	10	11	12

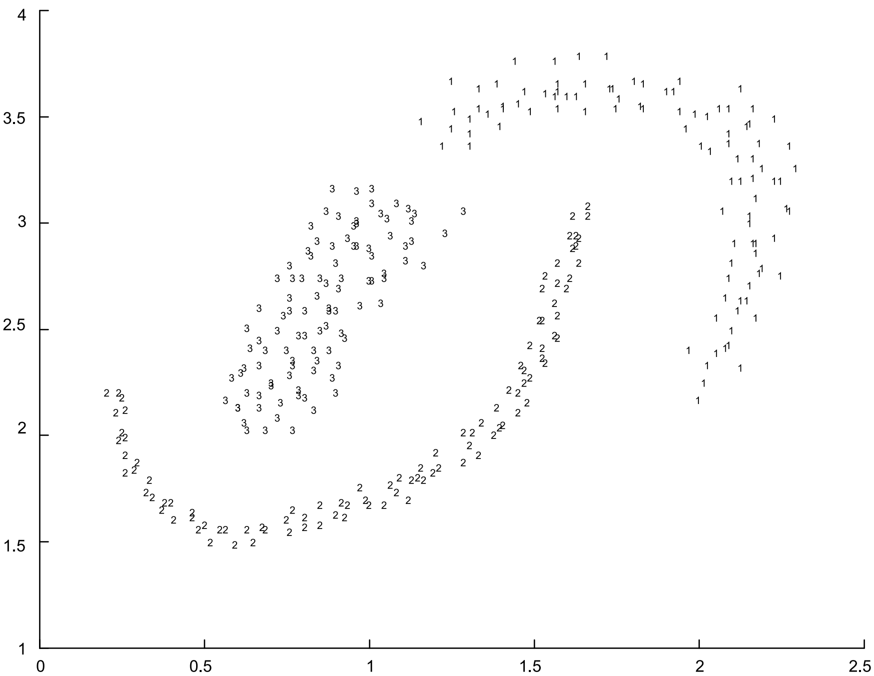

Fig. 5. Graphic results of classification. Points marked by a similar number belong to the same class.

8. Comparison with other cluster validity criteria

A lot of cluster validity were proposed during the last 10 years. They come from different studies dealing with the number of clusters existing in a set of points [5,15,17,23]. These studies started with hard partitions. In 1996 Hardy realised a comparative appraisal of the hard approaches allowing to determine the number of clusters [16]. These ones were translated to the fuzzy partitions [1,2,4,24]. Among the criteria which we can use to determine the number of clusters, we can notably cite the ones which are used with the classical Fuzzy C-Means algorithm [4]. We have chosen to compare our criterion to the most used in the field of fuzzy clustering: partition coefficient [2], partition entropy [3], proportion exponent [24], Fukuyama and Sugeno index [12], compactness separability [25] and an index called “Compose Within and Between Scattering” [21]. All these indexes are described in Table 2. We have adapted those using the distance between point x_k and centroid v_i of class C_i to the multiprototype approach. Indeed d_{ki} represents the distance between point x_k and the nearest centroid of class C_i in our case.

Table 2
Description of the cluster validity criteria[a]

Validity criteria	Functional description	Optimal cluster number
Partition coefficient	$F=\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}^2$	$\max(F,U,c)$
Partition entropy	$H=-\frac{1}{n}\sum_{k=1}^{n}\sum_{i=1}^{c}u_{ik}\log(u_{ik})$	$\min(H,U,c)$
Fuzzy hypervolume	$F_{hv}=\sum_{i=1}^{c}[\det(\Sigma_i)]^{1/2}$	$\min(F_{hv},U,c)$
Compactness separation	$\mathrm{CS}(U)=\frac{\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^2.d_{ki}^2}{n(\min_{v_i\in C_i,v_j\in C_j,i\neq j}\{v_i-v_j\})}$	$\min(\mathrm{CS},U,c)$
Fukuyama and Sugeno	$\mathrm{FS}(U)=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}^2(d_{ki}^2-\|v_i-\bar{v}\|^2)$	$\min(\mathrm{FS},U,c)$
Compose within and between scattering	$\mathrm{CWB}(U)=\alpha\,\mathrm{Scat}(c)+\mathrm{Dis}(c)$	$\min(\mathrm{CWB},U,c)$

[a] $\bar{v}$ is the grand mean of all data and v_i is the nearest centroid of the class C_i to the point x_k.

For the fuzzy hypervolume Σ_i is the fuzzy covariance matrix of class C_i defined by

$$\Sigma_i=\frac{S_i}{\sum_{i=1}^{c}u_{ik}^2},$$

where S_i is the fuzzy dispersion matrix defined by

$$S_i=\sum_{i=1}^{c}u_{ik}^2(x_k-v_i)\cdot(x_k-v_i)^{\mathrm{T}}.$$

For the index CWB we have used

$$\mathrm{Scat}(c)=\frac{1/c\sum_{i=1}^{c}[(\sigma(v_i))^{\mathrm{T}}\sigma(v_i)]^{1/2}}{[(\sigma(X))^{\mathrm{T}}\sigma(X)]^{1/2}}$$

and

$$\mathrm{Dis}(c)=\frac{D_{\max}}{D_{\min}}\sum_{i=1}^{c}\frac{1}{\sum_{z=1}^{c}\min_{v_l\in C_i,\,v_m\in C_z,\,l=1-c',\,m=1-c',\,l\neq m}\|v_l-v_m\|}$$

where

$$\sigma(X)=\begin{pmatrix}\frac{1}{n}\sum_{k=1}^{n}(x_k^1-\overline{x^1})^2\\ \vdots\\ \frac{1}{n}\sum_{k=1}^{n}(x_k^p-\overline{x^p})^2\end{pmatrix},\qquad \sigma(v_i)=\begin{pmatrix}\frac{1}{n}\sum_{k=1}^{n}u_{ik}(x_k^1-v_i^1)^2\\ \vdots\\ \frac{1}{n}\sum_{k=1}^{n}u_{ik}(x_k^p-v_i^p)^2\end{pmatrix},$$

Table 3
Values of different criteria for our example

Number of clusters	2	3	4	5	6	7	8	9	10	11	12
K_c	0.05	0.0025	0.065	0.095	0.122	0.086	0.110	0.065	0.032	0.013	0.010
F	0.887	0.782	0.765	0.746	0.725	0.702	0.688	0.659	0.623	0.601	0.572
H	0.207	0.417	0.485	0.551	0.614	0.680	0.726	0.799	0.894	0.954	1.034
F_{hv}	0.033	0.047	0.061	0.072	0.082	0.094	0.103	0.116	0.128	0.139	0.150
CS	0.087	0.092	0.088	0.084	0.080	0.077	0.083	0.078	0.110	0.105	0.107
FS	−78.2	−164.2	−162.6	−159.5	−152.8	−146.5	−140.3	−131.1	−128	−119.9	−118.7
CWB	4.387	4.256	4.098	5.407	6.071	5.855	6.187	6.519	7.365	7.213	7.481

$$D_{\max} = \max_{i=1-c, z=1-c, i\neq z} \left[\min_{v_l\in C_i, v_m\in C_z, l=1-c', m=1-c', l\neq m} \|v_l - v_m\| \right],$$

$$D_{\min} = \min_{i=1-c, z=1-c, i\neq z} \left[\min_{v_l\in C_i, v_m\in C_z, l=1-c', m=1-c', l\neq m} \|v_l - v_m\| \right],$$

$$\alpha = \mathrm{Dis}(c_{\max}).$$

Table 3 presents the results we have obtained for our example including two classes of parabolic shape and one class of elliptic shape. The set of points was divided into 12 subclasses before application of UFGC algorithm. The usual criteria attain their optimal value for a number of 2, 7 or 12 clusters. Criterion CWB attains its optimal value for four clusters whereas criteria K_c and FS attain their optimal values for three clusters. On this example criterion K_c is classified among the most performing criteria for validating the computed fuzzy partition.

9. Influence of the initial number of subclasses

In order to validate our criterion, we have studied its sensibility to the number of subclasses computed by Fuzzy C-Means in our algorithm. We can suppose that this number has an influence on the final partition and on the optimal value of criterion K_c. We know that this number must be enough high in order that all the subclasses belong to one real class only. As a consequence, we have chosen to provide this study from 12 to 30 subclasses. For 10 or 11 subclasses, our algorithm does not function because one subclass may belong to 2 real classes. Table 4 presents the evolution of criterion K_c for different number of subclasses. We can note that this criterion gives always three clusters. To finish, we conclude that our methodology with criterion K_c is not sensible to the initial number of subclasses if this last one is high enough. Nevertheless, the determination of this parameter believes a weak point of the method; even some heuristics could be used to determine it.

10. Applications

We have applied our criterion to three examples in order to show its ability to determine the number of classes existing into a set of points. These examples include difficulties currently met in the classification field.

Table 4
Values of criterion K_c for different numbers of subclasses

		Number of clusters										
		2	3	4	5	6	7	8	9	10	11	12
Number	14	1.51	0.37	7.56	12.89	15.84	11.71	13.58	17.04	11.05	9.05	2.29
of	16	1.75	0.79	1.77	10.30	15.58	23.95	16.96	20.88	25.35	18.07	15.21
initial	18	2.21	0.44	3.74	12.80	17.66	25.37	18.44	21.30	25.38	20.45	17.69
subclasses	20	2.13	0.1	6.88	15.16	20.37	23.80	24.80	17.72	18.51	20.33	21.38
	22	1.27	0.14	5.59	8.37	20.76	23.94	29.20	34.00	25.91	21.22	23.99
	30	0.33	0.02	16.63	16.86	24.38	38.61	37.43	36.92	41.56	43.44	44.27

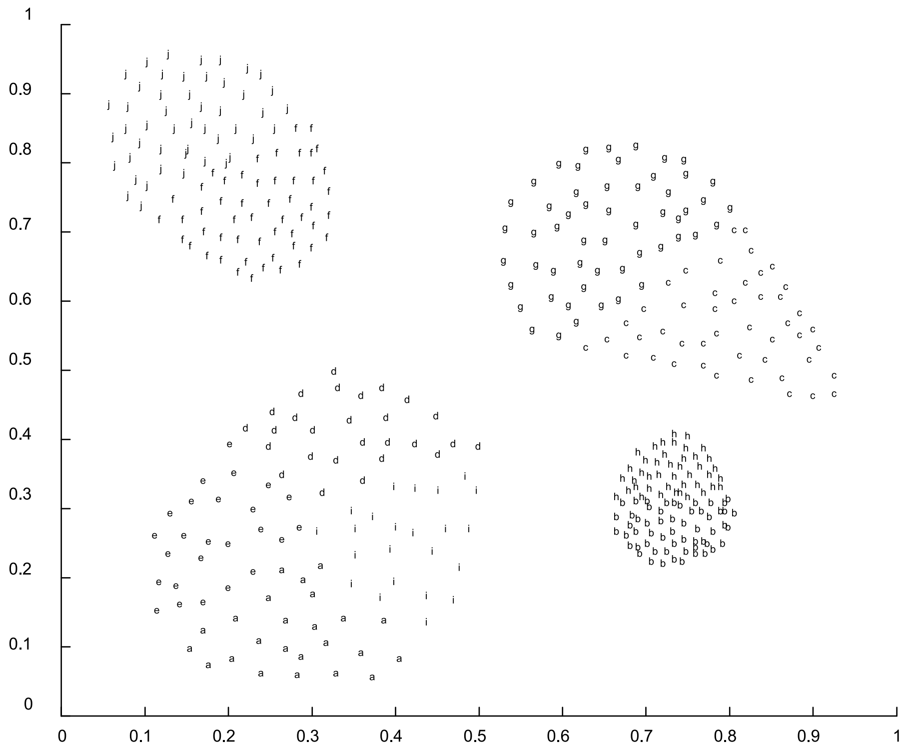

Fig. 6. Set of points of example 1. All the points marked by the same letter belong to the same subclass.

10.1. Example 1

The set of points is presented in Fig. 6. It includes 400 points divided into four classes of varying shape and density. It is divided into 10 subclasses as it is shown in Fig. 6. Table 5 shows the values of criterion K_c as a function of the number of classes. We can note that the minimum is attained for a value of four classes, which corresponds to the real number. The obtained classes are similar to the real ones. So our criterion succeeds for this set of points.

Table 5
Values of the criterion K_c as a function of the number of classes for example 1

$\alpha \times 10^{-2}$	7	8	10	14	16	18	19	20	21	23
$K_c \times 10^{-2}$	1.69	1.69	1.29	1.29	1.29	2.78	2.76	6.44	1.98	2.53
Number of clusters	3	3	4	4	4	5	6	7	8	9

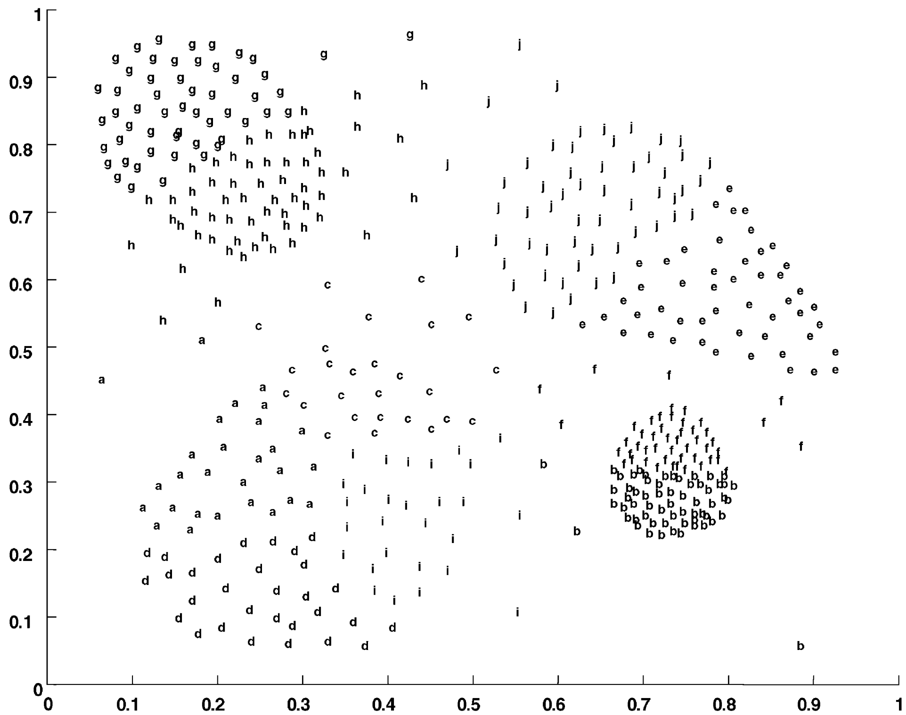

Fig. 7. Set of points of example 2. All the points marked by the same letter belong to the same subclass.

Table 6
Values of the criterion K_c as a function of the number of classes for example 2

$\alpha \times 10^{-2}$	8.5	9	12	14	16	17	20	22	24	26
$K_c \times 10^{-2}$	0.75	1.66	0.27	0.27	0.27	2.57	5.67	2.79	2.01	2.94
Number of clusters	2	3	4	4	4	6	7	8	9	10

10.2. Example 2

This example uses the same set of points as previously to which we have added some ambiguous points between classes. It is divided into 10 subclasses as it is shown in Fig. 7. We note in Table 6 that criterion K_c is minimised for a value of four subclasses which corresponds to the real number.

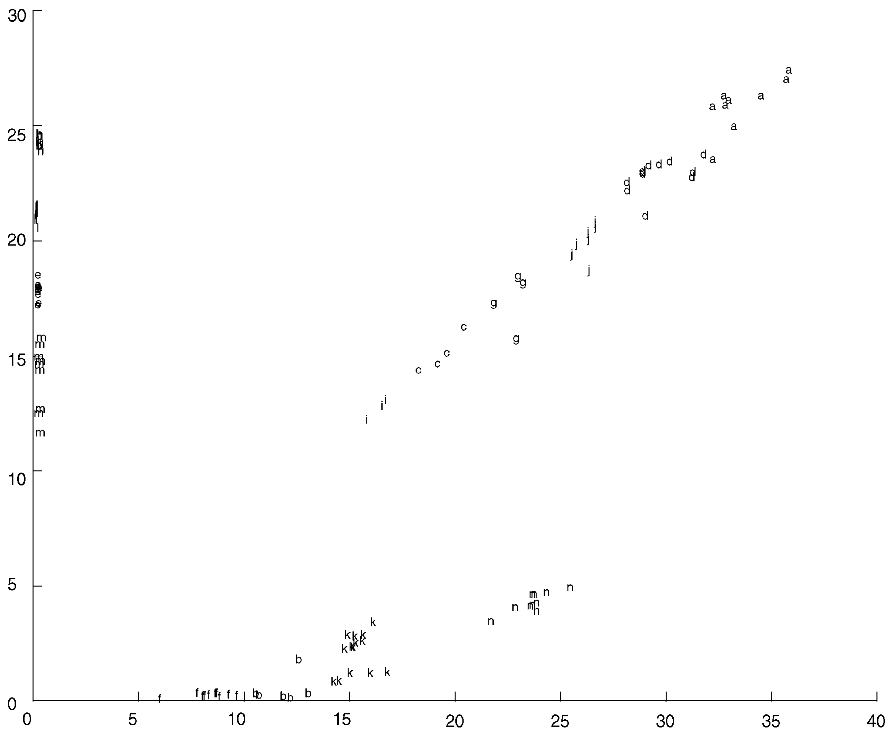

Fig. 8. Set of points of example 3. All the points marked by the same letter belong to the same subclass.

Table 7
Values of the criterion K_c as a function of the number of classes for example 3

$\alpha \times 10^{-2}$	0.5	0.7	1.3	1.4	4.3	5.1	6.8	7.4	8.5	9	10
$K_c \times 10^{-2}$	14.4	0.07	24.6	24.6	36.8	20.8	23.4	21.3	14.5	10.6	3.91
Number of clusters	2	3	4	5	6	7	8	9	10	12	15

10.3. Example 3

To finish, we have applied the method to the set of points coming from a study concerning the classification of plastic bottles [7] and presented in Fig. 8. It comprises of three classes of elongated and titled shape which constitutes generally a difficulty in classification. It was divided into 14 subclasses as it is shown in Fig. 8. The number of classes found by criterion K_c is 3 as we can see in Table 7.

11. Conclusion

This study presents an unsupervised classification method together with its application to a set of points including classes of complex shapes. The examples we have chosen show the ability of our criterion to determine the number of classes in a set of points whatever the shape of the classes. The classification method is based on the division of the set into subclasses. The number of these initial subclasses has no influence on the final partition on condition that it would be high enough. It can be determined by heuristics methods [7], but we can consider that this determination is a weak point of the method. For all these reasons, we can conclude that the UFGC algorithm is able to compute a fuzzy partition in a set of points without knowing a priori the number of classes. Indeed, we have proposed a criterion more performing than the usual criteria of class validity. However, the method presents several limits when distributions of classes are uneven and when classes do not appear clearly in the set of points. In these cases the proximity measure allowing one to establish a graduated hierarchy or a fuzzy order relation between subclasses is not efficient. A consequence is that the partition computed by the UFGC algorithm does not respect the shape of the real classes. Moreover criterion K_c does not succeed in the computing of the number of clusters. The problem of such classes is the basis of our future works dealing with unsupervised fuzzy clustering.

Finally, we can note that our methodology combines the clustering method of the Fuzzy C-Means and the hierarchical ascending method. It shows several positive aspects: the number of classes may be unknown and classes can have complex shapes, something which was impossible with the Fuzzy C-Means algorithm.

References

[1] E. Backer, A.K. Jain, A clustering performance measure based on fuzzy set decomposition, IEEE Trans. Pattern Anal. Mach. Intell. 3 (1) (1981) 66–75.
[2] J.C. Bezdek, Cluster validity with fuzzy sets, J. Cybernet. 3 (3) (1974) 58–73.
[3] J.C. Bezdek, Mathematical models for systematics and taxonomy, Proc. 8th Internat. Conf. Numerical Taxonomy, San Francisco, 1975, pp. 143–166.
[4] J.C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, New York, 1981.
[5] J.C. Bezdek, N.R. Pal, Some new indexes of cluster validity, IEEE Trans. Systems Man Cybernet. 28 (1998) 301–315.
[6] P. Billaudel, A. Devillez, G. Villermain Lecolier, An unsupervised fuzzy classification algorithm for the non elliptic classes, Proc. EURISCON'98, Athens, Greece, 1998.
[7] P. Billaudel, A. Devillez, G. Villermain Lecolier, Automation of plastic bottles sort using classification methods, Proc. 5th Internat. Conf. of the Decision Science Institute DSI'99, Athens, Greece, 1999, pp. 2016–2018.
[8] P. Billaudel, A. Devillez, G. Villermain Lecolier, Unsupervised fuzzy classification method based on a proximity graph and a graduated hierarchy, Proc. FUZZ-IEEE'99, Seoul, Korea, 1999, pp. 1054–1058.
[9] B. Bouchon-Meunier, La logique floue et ses applications, Addison-Wesley, Paris, 1995.
[10] V. Chepoi, D. Dumistrescu, Fuzzy clustering with structural constraints, Fuzzy Sets and Systems 105 (1999) 91–97.
[11] H. Frigui, R. Krishnapuram, A robust algorithm for automatic extraction of an unknown number of clusters from noisy data, Pattern Recognition Lett. 17 (1996) 1223–1232.
[12] Y. Fukuyama, M. Sugeno, A new method of choosing the number of clusters for the Fuzzy C Means method, Proc. 5th Fuzzy Systems Symposium, 1989, pp. 247–250.
[13] A.B. Geva, Hierarchical–fuzzy clustering of temporal patterns and its application for time-series prediction, Pattern Recognition Lett. 20 (1999) 1519–1532.
[14] M. Gondran, M. Minoux, Graphes et algorithmes, Collection des Etudes et Recherches d'Electricité de France, Paris, Eyrolles, 1985.
[15] R. Gunderson, Application of fuzzy ISODATA algorithms to star-tracker pointing systems, Proc. 7th Triannual World IFAC Congress, Helsinki, 1978, pp. 1319–1323.
[16] A. Hardy, On the number of classes, Comput. Statist. Data Anal. 23 (1996) 83–96.
[17] R. Kothari, D. Pitts, On finding the number of classes, Pattern Recognition Lett. 20 (1999) 405–416.
[18] M.H. Masson, B. Dubuisson, C. Frelicot, Conception d'un module de reconnaissance de formes floues pour le diagnostic, RAIRO-APII-JESA 30 (1996) 319–341.
[19] G.W. Milligan, M.C. Cooper, An examination of procedures for determining the number of clusters in data set, Psychometrika 50 (1985) 159–179.
[20] G. Nagy, State of the art in pattern recognition, Proc. IEEE 56 (1968) 836–882.
[21] M. Ramze Rezaee, B.P.F. Lelieveldt, J.H.C. Reiber, A new cluster validity index for fuzzy C mean, Pattern Recognition Lett. 19 (1998) 237–246.
[22] A. Ribert, Structuration évolutive de données, application à la construction de classifieurs distribués, thèse de doctorat en génie automatique de l'université de Rouen, 1998.

[23] E. Trauwert, L. Kaufman, P. Rousseeuw, Fuzzy clustering algorithms based on the maximum likehood principle, Fuzzy Sets and Systems 42 (1991) 213–227.
[24] M.P. Windham, Cluster validity for fuzzy clustering algorithms, Fuzzy Sets and Systems 5 (1981) 177–185.
[25] L.X. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Trans. Pattern Anal. Mach. Intell. 13 (1991) 841–847.