ELSEVIER Computational Statistics & Data Analysis 23 (1996) 83–96

COMPUTATIONAL STATISTICS & DATA ANALYSIS

On the number of clusters

André Hardy

Unité de Statistique, Département de Mathématique, Facultés Universitaires N.-D. de la Paix, 8 Rempart de la Vierge, B-5000 Namur, Belgium & Facultés Universitaires Saint-Louis, 43 Boulevard du Jardin Botanique, B-1000 Bruxelles, Belgium

Abstract

A large number of classification and clustering methods for defining and calculating optimal or well-suited partitions for data sets are available. Perhaps the most difficult problem facing the user of cluster analysis techniques in practice is the objective assessment of the stability and validity of the clusters found by the numerical technique used. The problem of determining the "true" number of clusters has been called the fundamental problem of cluster validity. The aim of this paper is to compare three methods based on the hypervolume criterion with other well-known methods. To illustrate and compare their behaviour, these procedures for determining the number of clusters are applied to artificially constructed bivariate data containing various types of structure. To provide a variety of solutions six clustering methods are used. We finally conclude by pointing out the performance of each method and by giving some recommendations to help potential users of these techniques.

Keywords: Cluster analysis; Hypervolume criterion; Number of clusters

1. The clustering problem

We consider a set E of n m-dimensional observation points $x_1, x_2, \ldots, x_n$ in the Euclidean space R^m. We want to find a partition of the set of objects E into k clusters; k is supposed fixed. Let P_k denote the set of all partitions of E into k clusters.

For that problem to be mathematically well-stated, we associate, to each P in P_k, the value of a clustering criterion $W(P,k)$ which measures the quality of each partition into k clusters.

The clustering problem is then to find the partition P^* that maximizes or minimizes the criterion $W(P,k)$ over all partitions into k clusters.

PII S0167-9473(96)00022-9

2. The hypervolume method and the hypervolume criterion

We suppose that we have a clustering problem when the points we observe are a realization of a Poisson process in a set D, where D is the union of k disjoint domains D_i $(1 \leq i \leq k)$. $C_i \subset \{x_1, x_2, \ldots, x_n\}$ is the subset of observations belonging to D_i $(1 \leq i \leq k)$. The problem is to estimate the unknown subsets D_i in which the points are distributed.

Let x denote the sample vector $\{x_1, x_2, \ldots, x_n\}$. The likelihood function can be written as follows:

$$f_D(x) = \frac{1}{(m(D))^n} \prod_{i=1}^{n} I_D(x_i),$$

where $I_D(x_i)$ is the indicator function of the set D at the point x_i and $m(D)$, the sum of the Lebesgue measures of the k subsets D_i. The hypervolume criterion is deduced from this statistical model by the method of maximum likelihood estimation. For that problem to be statistically well-defined, we have to impose the convexity of the sets D_i (Ripley and Rasson, 1977). So the maximum likelihood estimation of the k subsets $D_1, D_2, \ldots, D_k$ is constituted by the k subgroups of points such that the sum of the Lebesgue measures of their disjoint convex hulls is minimum (Hardy and Rasson, 1982; Hardy, 1983).

So the hypervolume criterion can be written as follows:

$$W : P_k \to R^+ : P = \{C_1, C_2, \ldots, C_k\} \hookrightarrow W(P,k) = \sum_{i=1}^{k} m(H(C_i)),$$

where $H(C_i)$ is the convex hull of the points belonging to C_i and $m(H(C_i))$ is the m-dimensional Lebesgue measure of that convex hull. One-dimensional hypervolume partitions are necessarily contiguous because of the assumption of convexity of the clusters, i.e. each group corresponds to a single interval disjoint from all other groups. So if we measure only one variable on each object, the problem consists in finding the k intervals of points such that the sum of the lengths of the intervals is minimum. In a two-dimensional space (m=2), the optimal solution is constituted by the k subgroups of points such that the sum of the areas of their disjoint convex hulls is minimum. In an m-dimensional space we minimize the sum of the m-dimensional hypervolumes of the convex hulls of the clusters.

Let us insist on the fact that the only conditions we impose are the homogeneous Poisson process for the distribution of the points in D and the convexity of the clusters. We make no other restrictions on the shape, the size, the tightness or the orientation of the clusters.

The majority of clustering techniques are based on the calculation of a matrix of similarities or distances between entities. The originality of the criterion comes from the fact that we generalize the Lebesgue measure in R by the Lebesgue measure in R^m itself. The hypervolume method is a non-hierarchical procedure for determining non-overlapping convex clusters.

The existing global algorithm is a dynamic programming procedure based on a multidimensional generalization of the Fisher algorithm (Hartigan, 1975; Hardy and Rasson, 1982; Hardy, 1983). The partition obtained is guaranteed optimal, but not necessarily unique. Nevertheless, according to the assumptions of the model (Poisson process), the solution is unique with probability one. That procedure can be implemented in a polynomially bounded time (Lengyel, 1979). So the exact algorithm is efficient (Barthélemy et al., 1992). Nevertheless, it is time consuming.

We proposed also a heuristic approach for the hypervolume method. It is a split and merge algorithm. It consists of two parts. The first one is a hierarchic divisive procedure which yields a partition of the set of objects into k clusters. The aim of the second part of the algorithm is to improve that partition according to the hypervolume criterion.

The local algorithm is faster than the global one. If the clusters are "well-separated" the first part of the algorithm yields the partition corresponding to the global optimum of the hypervolume criterion function. In almost all the other cases we obtain the exact partition after the second part of the local algorithm.

The local algorithm is faster than the global one, but it is still time consuming because it imposes the computation of convex hulls and hypervolumes of convex hulls in a m-dimensional space.

It is possible to avoid that computations by using classical approximations of convex hulls (Bentley et al., 1982, ...) or by computing depth contours (Ruts and Rousseeuw, 1994). That last approach has the advantage that replacing the convex hulls of the clusters by depth contours robustify the corresponding procedure.

Another approach: the concept of weak convex hulls and the corresponding algorithms (Schmitt and Mattioli, 1993, 1994) should also contribute to the improvement of the proposed algorithms for the hypervolume method.

The hypervolume procedure fulfils most of the admissibility conditions of Fisher and Van Ness (1971). It retrieves also the "natural" structure of the data, if any.

3. Methods to determine the number of clusters

The first three methods involve the hypervolume criterion (Hardy, 1992).

3.1. A classical geometric method (M1)

That well-known method consists in plotting the value of a clustering criterion W against k, the number of groups, and assessing the plot by eye, looking for discontinuities in slope. A sharp step in the curve indicates the number of classes; otherwise there is no justification for having more than one class. Unfortunately, that procedure can be unreliable; some clustering criteria can show large changes when analysing unstructured data. Nevertheless, that method associated with the hypervolume criterion gives interesting results.

3.2. Method based on the estimation of a convex set (M2)

That method is based on the solution of the following problem formulated by Professor D.G. Kendall: "Given a realization of a homogeneous planar Poisson process of unknown intensity within a compact convex set D, find D". If $x_1, x_2, \ldots, x_N$ denote the points in D where realizations are observed, then conditional on the value n of N, $x_1, \ldots, x_n$ are the values of n independent uniformly distributed random vectors on D. When the value n of N is known, the problem is reduced to the estimation of the contour of a compact convex set D given the position of n points uniformly drawn from it.

The solution analysed by Ripley and Rasson (1977) consists of a dilation of the convex hull $H(x)$ of $x = \{x_1, \ldots, x_n\}$ about its centroid. So that solution is expressed by $D' = g(H(x)) + cs(H(x))$, where x is the realization of the Poisson process in D, $H(x)$ the convex hull of x, $g(H(x))$ the centroid of $H(x)$ and $s(H(x)) = H(x) - g(H(x))$.

This solution is affine invariant, as is the problem itself.

The constant of dilation c is given by

$$c = \frac{n+1}{(n+1) - E(V_{n+1})},$$

where V_{n+1} is the number of vertices in the convex hull of $x_1, \ldots, x_{n+1}$ (i.e. as if there would be one more point drawn from D). Unfortunately, the determination of $E(V_{n+1})$ is difficult in practice. But it can be estimated by $c = \sqrt{n/(n - v_n)}$, where v_n is the number of vertices in the convex hull $H(x)$ (Moore, 1984).

The realization of a Poisson process within the sum of k subsets $D_1, D_2, \ldots, D_k$ can be considered as the realization of k Poisson processes of the same intensity within the k subsets $D_1, D_2, \ldots, D_k$ (Neveu, 1974).

Let us denote by $P^* = \{C_1^k, C_2^k, \ldots, C_k^k\}$ the optimal partition of E into k clusters and by $D_i'^k$ the estimate of the convex compact set D_i^k containing C_i^k. So $D_i'^k$ is the dilated version of D_i^k.

We propose the following decision rule for estimating k; checking for $t = 2, 3, \ldots$:

- if, for all $\{i,j\} \subset \{1,2,\ldots,t\}$, $i \neq j$: $D_i'^t \cap D_j'^t = \emptyset$, and if, for any integer s with $2 \leq s < t$ and for all $\{i,j\} \subset \{1, 2, \ldots, s\}$, $i \neq j$: $D_i'^s \cap D_j'^s = \emptyset$, then we conclude that the natural partition contains at least t clusters and we examine the partition into $(t+1)$ clusters;
- if there exists $\{i,j\} \subset \{1,2,\ldots,t\}$, $i \neq j$: $D_i'^t \cap D_j'^t \neq \emptyset$, and if for any integer s with $2 \leq s < t$, and for all $\{i,j\} \subset \{1,2,\ldots,s\}$, $i \neq j$: $D_i'^s \cap D_j'^s = \emptyset$, then we conclude that the data set contains exactly $t-1$ natural clusters;
- if $D_1'^2 \cap D_2'^2 \neq \emptyset$, then we conclude that there is no clustering structure in the data.

3.3. A likelihood ratio test for clusters (M3)

Because of the existence of an explicit model associated with the hypervolume method, we can formulate a likelihood ratio test for the number of clusters.

Let $x_1, x_2, \ldots, x_n$ be a random sample from a Poisson process on k disjoint convex sets $D_1, D_2, \ldots, D_k$ in an m-dimensional Euclidean space.

For a given integer $k \geq 2$, we test whether a subdivision into k clusters is significantly better than a subdivision into $k-1$ clusters, i.e. $H_0: t = k$ against the alternative $H_1: t = k-1$.

Let us denote by

- $C = \{C_1, C_2, \ldots, C_k\}$ the optimal partition into k clusters, with respect to the hypervolume criterion;
- $D = \{D_1, D_2, \ldots, D_{k-1}\}$ the optimal partition into $k-1$ clusters.

The likelihood ratio takes the form (Hardy, 1992):

$$Q(x) = \frac{\sup_D f_D(x; t = k-1)}{\sup_C f_c(x; t = k)} = \frac{1/(m(D))^n}{1/(m(C))^n} = \left(\frac{W(P,k)}{W(P,k-1)}\right)^n.$$

We have $Q(x) \in [0,1]$. Thus, we will reject H_0 iff:

$$S = \frac{W(P,k)}{W(P,k-1)} > l,$$

where $l \in R$. Unfortunately, we do not know the sampling distribution of the statistic S (as it is often the case with statistics derived from other clustering methods).

Nevertheless, for practical purposes, we can use the following rule: reject H_0 if S takes large values i.e. if S is close to 1. Practically, we will apply the test in a sequential way: if k_0 is the smallest value of $k \geq 2$ for which we reject H_0, we will consider $k_0 - 1$ as the appropriate number of natural clusters.

The four other methods are well-known in the scientific literature and were chosen because they are available in the cluster analysis software Clustan (Wishart, 1978).

3.4. Wolfe's test (M4)

That test, proposed by Wolfe, (1970), is based on the assumption of multivariate normality. It is a likelihood ratio criterion to test the hypothesis of k clusters against $k-1$ clusters.

3.5. The upper tail rule (M5)

The upper tail rule, proposed by Mojena (1977), is a test for a significant number of clusters in a hierarchical clustering sequence. That statistical stopping rule selects the partition associated to the first stage j in the cluster sequence $j = 2, 3, \ldots, N-2$ which satisfy the condition

$$z_{j+1} > \bar{z} + ks_z,$$

where $\bar{z}$ and s_z are the mean and unbiased standard deviation of the distribution of $N-1$ values; k is the standard deviate.

So z_{j+1} lies in the upper tail of the distribution of criterion values $\bar{z}$. The best number of clusters will be $N-j$ for the first z_{j+1} satisfying the condition. The conclusion is that there is no significant classification if no z_{j+1} satisfies the condition.

3.6. The moving average quality control rule (M6)

It is also a test for a significant number of clusters in a hierarchical sequence. That statistical stopping rule, also proposed by Mojena (1977), chooses the partition associated to the first stage j in the partial cluster sequence $j = r, r+1, \ldots, N-2$ which satisfy the condition:

$$z_{j+1} > \bar{z} + L_j + b_j + ks_z,$$

where r is the number of items in the moving average; $\bar{z}$ is the moving mean in stage j; L_j is the correction for trend lag in stage j; b_j is the moving least squares slope in stage j; s_z is the moving unbiased estimate of the standard deviation in stage j. As in the preceeding rule, the conclusion is that there is no significant classification if no z_{j+1} satisfies the condition.

3.7. Marriot's test (M7)

Marriot (1971) suggests that a possible procedure for assessing the number of clusters is to take that value of k for which $k^2|W|$ is a minimum where W is the matrix of within-group dispersion.

4. Results

In order to make the announced comparison, we have chosen six well-known clustering procedures (nearest neighbour, furthest neighbour, centroid, Ward, K-means and hypervolume) and six artificial data sets illustrating the principal difficulties of cluster analysis (bridges between classes, absence of any group structure, elongated clusters, non-linearly separable groups, unequal-size hyperspherical-shaped clusters and non convex clusters). We have then applied seven methods to determine the best number of clusters to the results so obtained.

We have chosen two-dimensional data because it is easier to point out the performances and properties of statistical procedures on data that can be represented graphically and directly assessed by eye. The majority of the results and conclusions remains valid if we analyse m-dimensional data sets where m is greater or equal to three.

4.1. First set of data: bridges between clusters

If the clusters are compact and well separated then almost any method will succeed in finding the obvious clusters. In our first example (Fig. 1), two hyperspherical-shaped clusters are linked together by chains of intermediate objects (bridges). We have obtained the following results for k with our seven procedures.

In the first column of each table we find the names of the clustering methods. When a procedure reveals the "true" classification, a "+" appears in the second column of the table. The nearest-neighbour method suffers from what is generally called

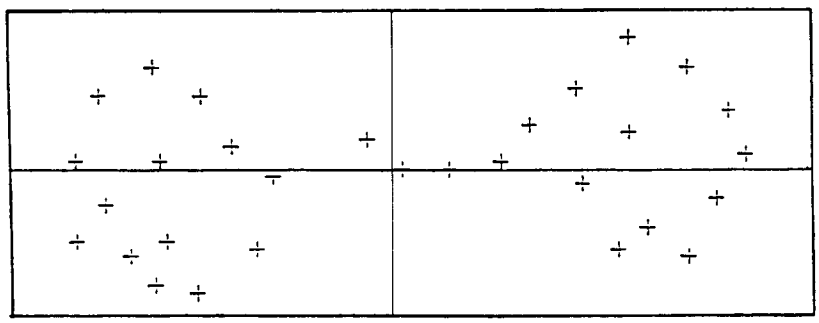

Fig. 1. Bridges between clusters.

Table 1
Bridges between clusters

Bridges between clusters		M1	M2	M3	M4	M5	M6	M7
Nearest neighbour	–	*	–	–	2	2	2	10
Furthest neighbour	+	3	–	–	2	2	2	10
Centroid	+	*	–	–	2	2	4	10
Ward	+	*	–	–	2	2	4	10
K-means	+	*	–	–	2	–	–	10
Hypervolume	+	2	2	2	–	–	–	10

a "chaining effect". Here the resulting grouping is meaningless. This is expressed by a "–" in the second column of Table 1. The seven last columns show the results given by the seven methods for the determination of the optimal number of clusters.

For example, the application of the upper tail rule (M5) to the results given by the furthest neighbour clustering procedure leads to the conclusion that there are two clusters in the data of Fig. 1.

Here, the graphical procedure (M1) is not very helpful; the decision over whether such plots contain the necessary "sharp step" is sometimes too subjective. Nevertheless, method M1 applied to the groupings given by the hypervolume method, determines the correct value of k. When the result given by a method is not clear enough to conclude, a "$*$" appears in the corresponding case of the table. The two other methods based on the hypervolume criterion (M2 and M3) as well as the upper tail rule (M5) and Wolfe's test (M4) give also the expected result. But the moving average control rule (M6) applied to the results given by the centroid and the Ward methods and the Marriot's test (M7) do not give the appropriate result; here the curve $k^2 \det(W)$ has its minimum value for $k^* = 10$.

The sign "–" in one of the seven last columns of a table means that the assumptions of the method are not fulfilled. For example, methods M2 and M3 are based on the hypervolume criterion; so they are only applicable to the results given by the hypervolume clustering procedure. M5 and M6 are valid only for hierarchic clustering methods. The results of Table 1 show that we must be very careful; for example, Wolfe's test (M4) yields the correct number of clusters when applied to the results of

the nearest neighbour procedure ... but the classification obtained is not the "natural" one.

4.2. *Second set of data: absence of any group structure*

Here we have simulated a Poisson process into a single convex set in the plane; so the 150 points are independently and uniformly distributed in this set (Fig. 2). Let us recall that we are interested here in classification, and not in dissection. Many of the methods find groups when in fact none are present. Results of Table 2 show that the three methods based on the hypervolume clustering criterion are very efficient when the problem is to test if there is any grouping structure in a data set; in this case the hypervolume criterion decreases monotonically with k. The other methods are not very efficient. Furthermore, Marriot's test (M7) is not

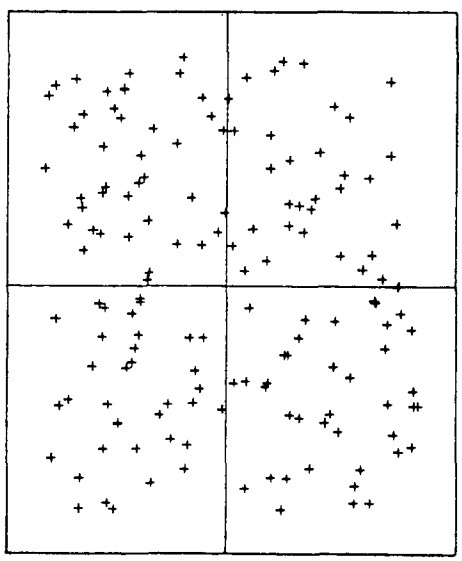

Fig. 2. Absence of any group structure.

Table 2
Absence of any group structure

Absence of any group structure	M1	M2	M3	M4	M5	M6	M7
Nearest neighbour	1	–	–	1	2	2	X
Furthest neighbour	*	–	–	2	2	4	X
Centroid	*	–	–	2	3	3	X
Ward	4	–	–	3	2	4	X
K-means	4	–	–	2	–	–	X
Hypervolume	1	1	1	–	–	–	X

applicable here: it is only valid when the number of clusters is greater or equal to two.

4.3. Third set of data: two elongated clusters

In this example, the natural structure consists of two elongated clusters (Fig. 3). The results obtained are given in Table 3.

Only the nearest neighbour and the hypervolume clustering methods reproduce the "natural" classification of the data. Possibly chaining of a clustering method is often regarded as a defect, but for example here, the very tendency of the nearest neighbour to form chains can be advantageous if the clusters are elongated or possess elongated limbs. The furthest neighbour, centroid, Ward and K-means are biaised towards finding spherical clusters. So here these methods fail to find the correct clustering when we fix the number of clusters to two. The three methods based on the hypervolume criterion, and methods M1, M4, M5 and M6 applied to the results given by the nearest neighbour method, give the correct number of "natural" clusters. Marriot's test is not very performing.

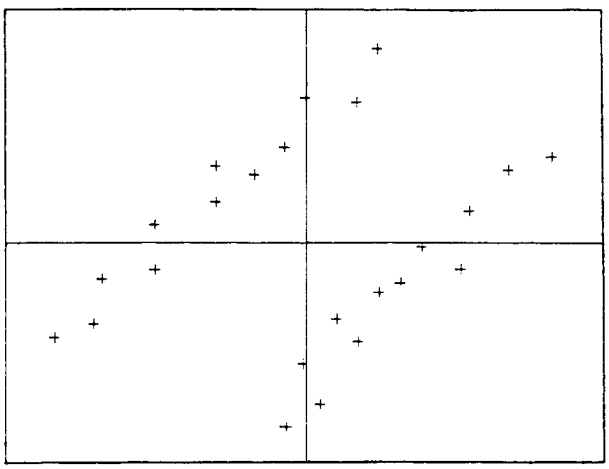

Fig. 3. Two elongated clusters.

Table 3
Two elongated clusters

Two elongated clusters		M1	M2	M3	M4	M5	M6	M7
Nearest neighbour	+	2	–	–	2	2	2	9
Furthest neighbour	-	2	–	–	*	2	2	9
Centroid	-	*	–	–	2	2	4	9
Ward	-	4	–	–	2	2	4	9
K-means	-	*	–	–	*	–	–	9
Hypervolume	+	2	2	2	-	-	-	9

4.4. *Fourth set of data: non-linearly separable clusters*

In this example, artificial data were generated to contain three elongated clusters such that no one is linearly separable from the two others. Many clustering methods fail to detect these three groups. The results obtained are given in Table 4.

Only the nearest neighbour and the hypervolume clustering methods are efficient in recovering the underlying structure into three clusters. Let us remark that Wolfe's test (M4) applied to the results of the furthest neighbour, centroid, Ward and K-means procedures give the correct number of clusters ... but the classifications obtained by these clustering methods are not "natural" ones. Here also M7 does not give a good result.

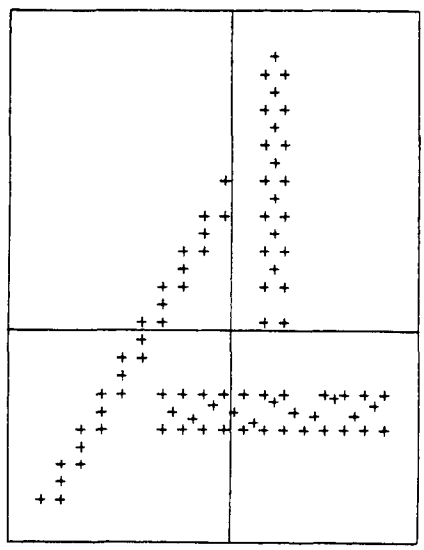

Fig. 4. Non-linearly separable clusters.

Table 4
Non-linearly separable clusters

Non linearly separable clusters		M1	M2	M3	M4	M5	M6	M7
Nearest neighbour	+	3	–	–	3	2	3	10
Furthest neighbour	–	*	–	–	3	2	2	10
Centroid	–	*	–	–	3	2	2	10
Ward	–	3	–	–	3	2	3	10
K-means	–	3	–	–	3	–	–	10
Hypervolume	+	3	*	3	-	-	-	10

4.5. Fifth set of data: unequal-size hyperspherical-shaped clusters

The assumption about the "shape" and size of groups can be implicit in the definition of a clustering criterion; for example, the sum of squares criterion will tend to produce equal-size hyperspherical-shaped clusters. The "–" in the second column of Table 5 means that the Ward method does not produce the "natural" clusters.

In the preceeding examples, Wolfe's test (M4) performed well; here it fails in recognizing the "true" number of natural clusters. The three methods based on the hypervolume criterion perform well. The upper tail rule (M5) and the moving average control rule (M6) yield the correct number of clusters when they are applied to the results given by the nearest neighbour, furthest neighbour and centroid methods.

4.6. Sixth set of data: non-convex clusters

In this example, it appears clear that there are two groups. One of them is not convex and is not linearly separable from the other group.

Here only the nearest neighbour method retrieves the "natural" structure of the data when we fix the number of clusters to two. The hypervolume clustering method is

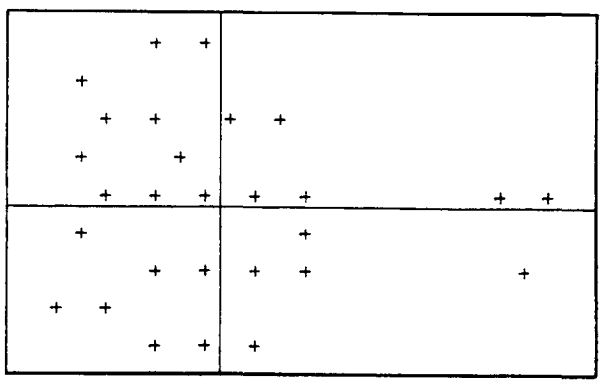

Fig. 5. Unequal-size clusters.

Table 5
Unequal-size clusters

Unequal-size clusters		M1	M2	M3	M4	M5	M6	M7
Nearest neighbour	+	2	–	–	1	2	2	10
Furthest neighbour	+	2	–	–	1	2	2	10
Centroid	+	2	–	–	1	2	2	10
Ward	-	3	–	–	1	2	3	10
K-means	+	*	–	–	1	–	–	10
Hypervolume	+	2	2	2	–	–	–	10

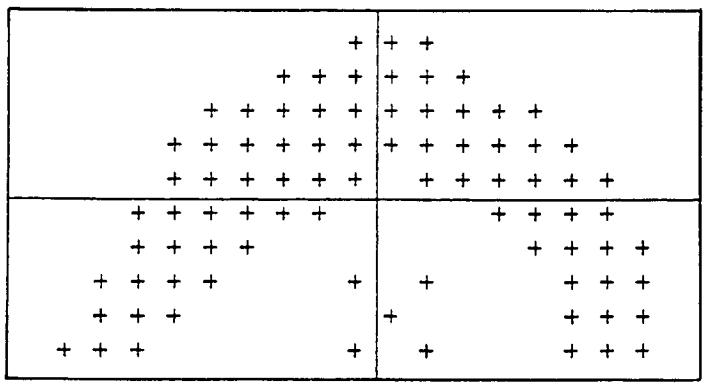

Fig. 6. Non convex clusters.

Table 6
Non convex clusters

Non convex clusters		M1	M2	M3	M4	M5	M6	M7
Nearest neighbour	+	*	–	–	3	2	2	10
Furthest neighbour	–	5	–	–	3	2	3	10
Centroid	–	5	–	–	3	2	3	10
Ward	–	6	–	–	2	2	3	10
K-means	–	*	–	–	3	–	–	10
Hypervolume	–	1	1	1	–	–	–	10

not applicable because the assumption of convexity is not fulfilled. Wolfe's test does not give the correct number of clusters. Only methods M5 and M6 associated with the classification given by the nearest neighbour procedure lead to the conclusion that there are two clusters in the data set illustrated in Fig. 6.

5. Conclusions

The aim of this paper was to investigate the performance of seven methods for determining the "true" number of clusters, and particularly to compare three methods based on the hypervolume criterion with four other well-known methods.

But we have seen that a big problem when we have to find groups in data is to choose a good clustering procedure; most of the methods make implicit assumptions about the type of structure present; when these assumptions fail to be met spurious solutions are likely to be obtained; it is not the case of the hypervolume method; we impose the convexity of the clusters, but we make no other restrictions on the shape, the size, the tightness or the orientation of the clusters.

Concerning the number of clusters, we have to be very careful: some methods give the "correct" number of clusters ... based on an "bad" classification. So we have to take into account the a priori information we have on the clustering

methods and the underlying assumptions of each of them: the chaining effect of the nearest neighbour method; the furthest neighbour, centroid, Ward and K-means methods favour hyperspherical-shaped clusters; the hypervolume method requires the convexity of the clusters, ...

When they are applicable, the three methods based on the hypervolume criterion perform well, even in the case of absence of any group structure.

The graphical method (M1) is in general very subjective. The decision over whether such plots contain the necessary "discontinuity in slope" is likely to be very subjective. Nevertheless, it gives the most interesting and clear results when we plot the hypervolume criterion against the number of groups.

Method M2 works usually very well but it may fail to obtain good results in presence of elongated clusters containing a small number of points. It should also be interesting and useful to find a more appropriate value for the constant of dilation c when we use it to detect the optimal number of clusters.

The likelihood ratio test (M3) is one of the most interesting method. Unfortunately, on a theoretical point of view, the sampling distribution of the test statistic is not known. But in practice it gives clear and relevant results.

Certainly, the fact that different clustering criteria and methods for determining the best number of clusters can suggest different results when applied to the same data set should make investigators wary about accepting uncritically the results of a single clustering method associated with a single method for the number of clusters. So what we can recommand, when we are confronted with data to classify, is to use several cluster analysis techniques and methods for determining the optimal number of clusters and to analyse all the results in order to have more informations about the clusters: size, shape, convexity, tightness, separation, ... and to take into account that information to choose the best classification and to interpret it carefully.

References

Anderberg, M.R., *Cluster analysis for applications* (Academic Press, New York, 1975).

Barthelemy, J.P., J.P. Cohen and A. Lobstein, *Complexité algorithmique et problèmes de communications* (Masson, Paris, 1992).

Baufays, P. and J.P. Rasson, Discriminant analysis with geometric tools, *Comput. Statist. Quat.*, **3** (1986) 1–20.

Bentley, J.L., M.G. Faust and F.P. Preparata, Approximation Algorithms for Convex hulls, *Commun. ACM*, **25** (1982) 64–68.

Bock, H.H., On some significance tests in cluster analysis, *J. Classification*, **2** (1985) 77–108.

Diday, E. and Collaborateurs, *Optimisation en classification automatique* (INRIA, Paris, 1979).

Duda, R.O. and P.E. Hart, *Pattern classification and scene analysis* (Wiley, New York, 1973).

Engelman, L. and J.A. Hartigan, Percentage points of a test for clusters, *J. Am. Stat. Assoc.*, **64** (1969) 1647–1648.

Everitt, B.S., Unresolved problems in cluster analysis, *Biometrics*, **35** (1979) 169–181.

Everitt, B.S., *Cluster analysis* (Halsted Press, London, 1980).

Fisher, L. and J.W. Van ness, Admissible clustering procedures, *Biometrika*, **58** (1971) 91–104.

Gordon, A.D., *Classification* (Chapman & Hall, London, 1981).

Hardy, A. and J.P. Rasson, Une nouvelle approche des problèmes de classification automatique, *Statistique et Analyse des données*, **7** (1982) 41–56.

Hardy, A., Statistique et Classification Automatique. Un modèle - Un nouveau critère - Des algorithmes - Des applications, Ph.D Thesis, Unité de Statistique, Département de Mathématique, Facultés Universitaires de Namur, Namur, Belgium, 1983.

Hardy, A, Criteria for determining the number of groups in a data set based on the hypervolume criterion, Technical Report, Unité de Statistique, Département de Mathématique, Facultés Universitaires de Namur, Namur, Belgium, 1992.

Hartigan, J.A., *Clustering algorithms* (Wiley, New York, 1975).

Lengyel, T., A note on the number of clustering. Technical Report IV/5, Computer and Automation Institute, Budapest, December 1979.

Marriot, F.H.C., Practical problems in a method of cluster analysis, *Biometrics*, **27** (1971) 501–514.

Mojena, R., Hierarchical grouping methods and stopping rules: an evaluation, *The Comput. J.*, **20** (1977) 359–363.

Mojena, R. and D. Wishart, Stopping rules for Ward's clustering method, *Compstat Proc.* (Physica Verlag, 1980) 426–432.

Moore, M., On the estimation of a convex set, *The Ann. Statist.*, **12** (1984) 1090–1099.

Neveu, J., Processus ponctuels, Technical Report, Laboratoire de Calcul des Probabilités, Université Paris VI, 1974.

Ripley, B.D. and J.P. Rasson, Finding the edge of a Poisson forest, *J. Appl. Probab.*, **14** (1977) 483–491.

Ruts, I. and P. Rousseeuw, Computing depth contours of bivariate point clouds, *Comput. Statist. Data Anal.*, This issue.

Schmitt, M. and J. Mattioli, *Morphologie Mathématique* (Masson, Paris, 1994).

Schmitt, M. and J. Mattioli, Strong and weak convex hulls in non-Euclidean metric: theory and application, to appear in: *Pattern Recognition Lett.*

Wishart, D., *Clustan User Manual*, 3rd edn. (Program Library Unit, University of Edimburgh, 1978).

Wolfe, J.H., Pattern clustering by multivariate mixture analysis, *Multivariate Behavioral Res.*, **5** (1970) 329–350.