Fuzzy dynamic clustering algorithm

Sankar K. PAL and Sushmita MITRA

Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta 700035, India

Abstract: A three-stage dynamic fuzzy clustering algorithm consisting of initial partitioning, a sequence of updating and merging by optimisation of a characterisation function based on measures of fuzziness in a set is described. Unlike the conventional detection of disjoint initial clusters, the algorithm can extract overlapping initial cluster boundaries when the feature space has ill-defined regions. The membership function in $\mathbb{R}^n$ involves the density of patterns at a point in addition to its Euclidean distance. The merging criterion involves the number of samples and the amount of fuzziness in the intersection of two clusters, and the disparity in their size. The effectiveness of the algorithm is demonstrated on the speech recognition problem.

Key words: Fuzzy clustering, measures of fuzziness, speech recognition.

1. Introduction

Clustering may be viewed as a problem of unsupervised pattern recognition. The objective is to partition the given data set into a certain number of natural and homogeneous sets where the elements of each set are as similar as possible and different from those of the other sets. In practice, the separation of clusters is a fuzzy notion and hence the concept of fuzzy subsets offers special advantages over conventional clustering [1]. In fuzzy clustering each element is assigned a finite membership to each of the clusters. The well-known fuzzy clustering algorithms include the fuzzy isodata [5], fuzzy C-means [6] and clustering by decomposition of induced fuzzy sets [3]. In the last two cases, the number of clusters is assumed to be known. Again, all these methods consider the initial clusters to start with to be disjoint.

The measures index of fuzziness, entropy and π-ness [1] (which provide an amount of difficulty in deciding whether a pattern is a member of a set or not) have been found to be successful in various pattern recognition and image processing problems, e.g., in segmenting an image [7], in defining a feature evaluation index [8], in determining

initial seed points [2], and in providing a quantitative measure for image enhancement [9].

The present work attempts to demonstrate another application of the aforesaid fuzzy measures in dynamic clustering of a data set. The technique involves a three-stage hierarchy. In the first stage, various fuzzy sets representing 'points clustered around some point, say b' are obtained. By optimizing measures of fuzziness over these sets, the seed points and the corresponding initial cluster boundaries of the feature space are extracted. Unlike conventional detection of initial clusters where the boundaries are made disjoint, this algorithm can extract overlapping initial clusters (boundaries) when the feature space has ill-defined regions.

In the second stage, membership values are assigned to the points in the feature space corresponding to each cluster. Besides the conventional use of Euclidean distance in measuring membership value [4], the density of patterns at a point is also considered here in this evaluation. This, in turn, makes use of the aforesaid fuzzy measures (in terms of the density of patterns at a point) in the process of evaluation. A sequence of cluster updating and membership assignment is repeated un-

til a local maximum value of a characterization function is obtained. The characterization function also includes the fuzzy measure as described above.

Finally, in the third stage a provision for merging is kept on the basis of an objective function. The objective function is dependent on three factors, namely, the number of points in the intersection of two clusters, fuzziness in the intersection of two clusters and the disparity in the size of two clusters.

The algorithm is able to generate the optimal number of clusters k_0 in the feature space both when k_0 is known and unknown. Effectiveness of the algorithm is demonstrated on speech recognition problems. Results of the individual stages are also highlighted.

2. Outline of the algorithm

Consider the feature space

$$\Omega_X = (l_1, u_1) \times (l_2, u_2) \times \cdots \times (l_n, u_n)$$

to be split into L^n grid points where $L = (u_i - l_i)/d$, and l_i, u_i are the lower and upper bounds of the ith property of the sample. d denotes the grid width.

Inputs

   (i) The n coordinates of the N pattern points in the n-dimensional feature space. (The upper and lower bounds u_i and l_i for $i = 1, 2, \ldots, n$ can then be obtained.)

   (ii) The grid width d.

   (iii) The radius λ of the $\hat{\pi}$-function.

Procedure

1. Determine the initial seedpoints and corresponding cluster boundaries.

2. Repeat steps 3 to 5 until a local maximum of a characterization function ψ occurs.

3. Assign membership values to each grid point.

4. Compute the function ψ.

5. Update the cluster centers.

Outputs

   (i) The k_0 cluster centers corresponding to op-

timal partitioning of Ω_X.

   (ii) The membership of L^n grid points to each of the k_0 clusters.

   (iii) The local maximum value of the characterization function ψ.

This algorithm may be run for suitable combinations of d and λ yielding a large number of initial clusters k. The optimization leads to minimum fuzziness among the k clusters in Ω_X. Now two conditions may arise.

Case 1. The number of optimal clusters k_0 is known, where $k > k_0$. At each stage, the pair of clusters having maximum fuzziness between them may be merged and the local maximum of ψ computed until $k = k_0$.

Case 2. The number of optimal clusters k_0 is unknown. A large k is first of all chosen. At each stage merging and maximization of ψ [as in Case 1] may then be repeated until a global maximum value of ψ is obtained. The corresponding set of k_0 clusters constitute the optimal partitioning of Ω_X.

In the following sections, the above-mentioned steps will be explained in detail.

3. Extraction of initial clusters

Let $X = \{X_1, X_2, \ldots, X_N\}$ be a set of N pattern points in an n-dimensional ($n \geq 2$) feature space Ω_X. The fuzzy set associated with X may be defined as [2]

$$X(b, \lambda) = \{\mu_{X(b,\lambda)}(X_i), X_i\}, \quad i = 1, 2, \ldots, N \quad (1)$$

where

$$\mu_{X(b,\lambda)}(X_i) = \hat{S}(X_i; b, \lambda) \text{ or } \hat{\pi}(X_i; b, \lambda)$$
$$\text{and} \quad X_i \in \mathbb{R}^n.$$

b corresponds to the cross-over point for the function $\hat{S}$ and the central point for the function $\hat{\pi}$ [2].

Here the $\hat{S}$-function is defined as

$$\hat{S}(X_i; b, \lambda) = (1 - \|X_i - b\|/\lambda)^2/2 \quad \text{or}$$
$$1 - (1 - \|X_i - b\|/\lambda)^2/2,$$
$$\text{when } \|X_i - b\| \leq \lambda,$$
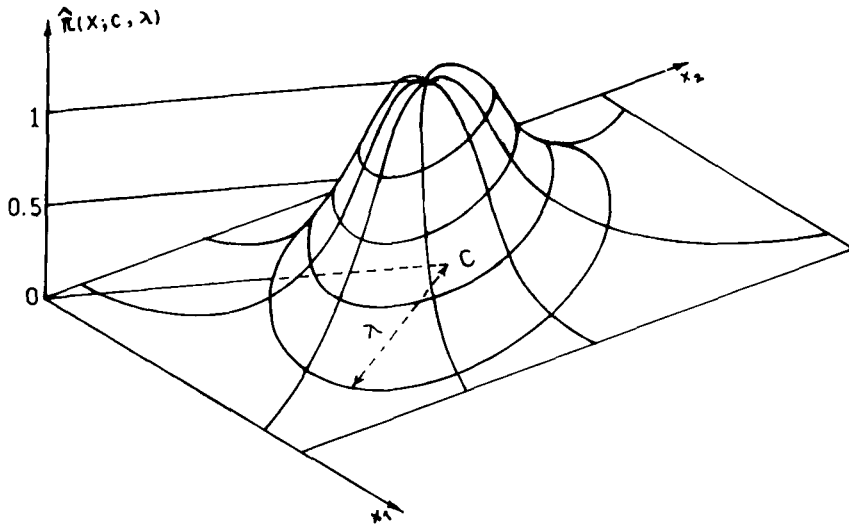$$= 0 \text{ or } 1, \quad \text{otherwise} \quad (2)$$

Figure 1. $\hat{\pi}$ function when $X \in \mathbb{R}^n$.

where $\| \cdot \|$ denotes the Euclidean distance in $\mathbb{R}^n$ and $\lambda > 0$ is the radius. The $\hat{\pi}$-function is defined as

$$\hat{\pi}(X_i; b, \lambda) = 2(1 - \|X_i - b\|/\lambda)^2,$$
$$\lambda/2 \le \|X_i - b\| \le \lambda,$$
$$= 1 - 2(\|X_i - b\|/\lambda)^2,$$
$$0 \le \|X_i - b\| \le \lambda/2. \tag{3}$$

This is shown in Figure 1 where $X_i \in \mathbb{R}^2$. Note that the central point b of equation (3) is designated as c in the figure.

Considering $\mu_{X(b,\lambda)}(X_i) = \hat{\pi}(X_i; b, \lambda)$, equation (1) can be viewed as a fuzzy set $X(b, \lambda)$ of 'points clustered around b' so that $\mu_{X(b,\lambda)}(X_i)$ denotes the degree of belonging of X_i to such a set. Keeping λ constant and changing b, we can therefore generate a class of such fuzzy sets.

Let l_i and u_i be the lower and upper bounds of the ith property of the sample. The feature space $(l_1, u_1) \times (l_2, u_2) \times \cdots \times (l_n, u_n)$ is split into L^n grid points where $L = (u_i - l_i)/d$ for a fixed i. Here d is some pre-assigned positive constant called grid width. Let b_i, $i = 1, 2, \ldots, L^n$, be the grid points.

Choosing λ suitably, calculate the amount of fuzziness in the set $X(b_i, \lambda)$ using any of the following measures.

Index of fuzziness

$$\gamma(X(b_i, \lambda)) = \frac{2}{N} \sum_{j=1}^{N} \min(\mu_{X(b_i,\lambda)}(X_j)) \tag{4}$$

Entropy

$$H(X(b_i, \lambda)) = \frac{1}{N \ln 2} \sum_{j=1}^{N} S_n(\mu_{X(b_i,\lambda)}(X_j)) \tag{5}$$

where S_n is the Shannon function defined as

$$S_n(\mu_{X(b_i, \lambda)}(X_j))$$
$$= -(\min \mu_{X(b_i, \lambda)}(X_j)) \ln(\min \mu_{X(b_i, \lambda)}(X_j))$$
$$- (\max \mu_{X(b_i, \lambda)}(X_j)) \ln(\max \mu_{X(b_i, \lambda)}(X_j))$$

and $\min \mu_{X(b_i, \lambda)}(X_j)$, $\max \mu_{X(b_i, \lambda)}(X_j)$ refer to the minimum and the maximum of the two μ-values at the point X_j of the $\hat{S}$-function.

Pi-ness

$$I(X(b_i, \lambda)) = \frac{1}{N} \sum_{j=1}^{N} \hat{\pi}(X_j; b_i, \lambda) \tag{6}$$

The expressions (4)–(6) represent an average amount of difficulty (ambiguity) in deciding a point X_j as a member of the set $X(b_i, \lambda)$. It follows that γ and H increase monotonically in the interval $[0, 0.5]$ and decrease monotonically in $[0.5, 1]$ with a maximum value 1 at $\mu = 0.5$. On the other hand, I (eq. (6)) has maximum value 1 at $\mu = 1$.

3.1. Criteria for seed point / boundary point extraction

We observe that the contributions towards $\gamma(X(b, \lambda))$ or $H(X(b, \lambda))$ or $I(X(b, \lambda))$ are mostly

from the points around b and it decreases as the points move away from b. Hence, if the number of points around b is more, there will be a greater number of points X_i having $\mu \approx 0.5$ (1) while using the $\hat{S}$-function ($\hat{\pi}$-function) [resulting in γ, H and $I \approx 1$] and a less number of points having $\mu \approx 0$ or 1 (0) [resulting in γ, H and $I \approx 0$]; thus increasing the value of $\gamma(X(b,\lambda))$ or $H(X(b,\lambda))$ or $I(X(b,\lambda))$. Therefore b can be considered as a seed point (center of an initial cluster). In other words, the higher the value of these fuzzy measures, the greater is the density of patterns clustered around b.

Similarly, points along the boundary of the cluster would have minimum γ or H or I values as the number of points around b would be less due to the sparse pattern distribution there. So corresponding to each cluster center the cluster boundaries can be extracted by detecting the locus of points with minimum $\gamma(X(b,\lambda))$ or $H(X(b,\lambda))$ or $I(X(b,\lambda))$ values surrounding the seed point.

This suggests that modification of the cross-over point/central point b will result in different fuzzy sets having various fuzzy measures. The grid points $\{b_i\}$ for which the corresponding fuzzy measures are locally maximum may be taken as the initial seed points. For each such seed point, the surrounding locus of grid points having minimum values of fuzzy measures constitute the corresponding cluster boundary. The algorithm for extracting seed points has been reported in [2]. The algorithm for generating the initial cluster boundaries is stated below.

3.2. Algorithm

Inputs

(i) The b_i, $i = 1, 2, ..., L^n$, grid points and the corresponding measures of fuzziness.

(ii) The initial seed points (already detected).

Procedure [considering $n = 2$] to detect the boundary corresponding to each seed point

Main

1. Proceed horizontally along the row (along axis 2, say) on both sides of the seed point.

2. Call subroutine FNDPTS to detect end points

(rightmost and leftmost points of the cluster boundary) along this axis.

3. For each grid point along this axis do steps 4 to 5.

4. Proceed vertically along the column (along axis 1) on both sides of this point.

5. Call subroutine FNDPTS to detect end points (constituting part of the cluster boundary) above and below this point.

Subroutine FNDPTS

1. Proceed to the next grid point along the desired direction.

2. Compare the measure of fuzziness at this point with that of the previous point.

3. If either a minimum is obtained or the decrease is very low (i.e., a valley is reached or the slope is very gentle compared to the adjacent points), then continue to the next step; otherwise, go to step 1.

4. If this point is nearer or equidistant to any other seed point, then choose this point as one of the end points; otherwise, go to step 1.

Output

A listing of the cluster boundary for each seed point. Here F_i, $i = 1, 2, ..., k$, designates the ith cluster associated with the ith seed point.

If the feature space has overlapping regions, that will be reflected by the output boundaries obtained by the above-mentioned algorithm. In other words, the initial cluster boundaries so obtained by the above-mentioned algorithm will be overlapping conforming to the notion that each point may have finite membership to more than one cluster. It is to be noted that this is unlike the conventional detection of initial clusters where the boundaries are made disjoint.

3.3. Variation of d and λ

As λ of the $\hat{\pi}$-function (Figure 1) decreases, the $\mu_{X(b,\lambda)}$ plane representing the fuzzy sets 'points clustered around b' would have more intensified contrast around the cross-over point b resulting in decrease of ambiguity (γ or H or I value) in $\mu_{X(b,\lambda)}$. As a result, the possibility of detecting some undesirable seed point (representing the

spurious maxima in the feature space) increases. Similar is the case with decrease in the value of the grid width d.

On the other hand, increase in value of λ or d results in a higher value of fuzziness and this leads towards the possibility of losing some of the weak maxima.

4. Assignment of membership value

Fuzzy clustering uses iterative optimization of an objective function based on a weighted similarity measure between the pattern points in the feature space and each of the cluster centers. A local extremum of this objective function indicates an optimal clustering of the input data.

Let $X' = \{b_i; i = 1, 2, \ldots, L^n\}$ be the set of grid points in Ω_X. Let the fuzzy measure (index of fuzziness or entropy or pi-ness) computed at point x, where $x \in X'$, be denoted by z_x. As seen in Section 3, z_x measures an average amount of difficulty in deciding whether a pattern can be considered a member of the set 'points clustered around x' or not. The higher the value of z_x, the greater is the density of patterns clustered around x.

A way of computing membership $\mu_{F_i}(x)$ of a point x to a cluster F_i is given below.

Case 1. When $\|x - v_i\| \neq 0$ and $\|z_x - z_{v_i}\| \neq 0$ for all i, then

$$\mu_{F_i}(x) = \left(\sum_j \left[\frac{\|x - v_i\|}{\|x - v_j\|} \left(\frac{\|z_x - z_{v_i}\|}{\|z_x - z_{v_j}\|} \right)^\beta \right]^\delta \right)^{-1}$$

(7)

where $0 < \beta \leq 1$, $\sum_i \mu_{F_i}(x) = 1$, $i, j = 1, 2, \ldots, k$, and $\mu_{F_i}(x) \in [0, 1]$.

The center of cluster F_i is denoted by v_i. Note that during the iterative updating (to be explained in Section 5), v_i is to be considered as the initial ith seed point (extracted in Section 3) at the first iteration. The positive constant δ controls the fuzziness in a set. β is a weight associated with the pattern density measure.

The expression (7) incorporates a measure of the difference in density of pattern distribution between the point x and the corresponding cluster center, in addition to their Euclidean distance. The significance of inclusion of the fuzzy measures in

this expression can be visualized by considering the speech recognition problem, as an example. Here dialectic and other such variations may lead to the generation of a good amount of samples of the same vowel having coordinates (features) that are far apart in the feature space as measured by the distance metric. Considering only the Euclidean distance as a criterion for evaluating the membership may yield low values in such cases. However, by considering z_x as a factor we give due importance to the density of patterns around a point in determining its membership to a cluster. So different versions of the same vowel, lying far apart and yet having considerable pattern density (number of occurrence), are assigned higher membership values as compared to that which would have been assigned using Euclidean distance only. Since $\beta \leq 1$, the factor z_x has, of course, less importance than the distance factor in providing membership value.

It is to be mentioned here that the work in [4] considered a membership function involving only Euclidean distance.

Case 2. When $\|x - v_i\| \neq 0$ and $\|z_x - z_{v_i}\| = 0$ for any i, then it implies that x is not the seed point but has equal amount of pattern density as the seed point has. In that case use of equation (7) will result in infinite μ value which is impractical. In order to circumvent this, we use equation (7) with $z_x = (z_x)_{av}$ where $(z_x)_{av}$ denotes the average value of z_x computed over its four neighbours. It is expected that $\|(z_x)_{av} - z_{v_i}\| \neq 0$.

Case 3. When $\|x - v_i\| = 0$ for $i = i_0$, then (obviously, $\|z_x - z_{v_i}\| = 0$)

$$\mu_{F_i}(x) = 1 \quad \text{for } i = i_0,$$

(8)

$$= 0 \quad \text{otherwise,}$$

such that $\sum_i \mu_{F_i}(x) = 1$.

The membership of each grid point to each of the K clusters can therefore be evaluated using equation (7) or (8).

5. Optimised partitioning

Since an optimum clustering corresponds to minimum overlap among the clusters $F_1, F_2, \ldots, F_k$,

the notion of fuzzy set intersection therefore becomes an important criterion for measuring fuzziness (ambiguity) in clustering. The amount of fuzziness present in clusters F_j and F_i ($j \neq i$) may be viewed to be equivalent to the amount present in the subset ($F_j \cap F_i$).

The fuzziness in $F_j \cap F_i$ is defined as

$$L(F_j \cap F_i) = \frac{1}{|X'|} \sum_{x \in X'} \min(\mu_{F_j}(x), \mu_{F_i}(x)) \quad (9)$$

where $|X'|$ refers to the number of grid points in the feature space Ω_X, i.e., $|X'| = L^n$.

Given a k-collection of fuzzy sets $\{\mu_{F_j}(x), \forall x \in X'$ and $j = 1, 2, \ldots, k\}$ satisfying $\sum_j \mu_{F_j}(x) = 1$, the measure of average pairwise fuzzy set separability is [3]

$$\psi = 1 - \frac{2}{k-1} \sum_j \sum_i L(F_j \cap F_i) \quad (10)$$

where $j = 1, 2, \ldots, k-1$ and $i = j+1, \ldots, k$.

The constant term $2/(k-1)$ appears in order to make the characterization function ψ lie between 0 and 1.

Since there are $k(k-1)/2$ nonzero terms, the maximum fuzziness ($\psi = 0$) is attained when

$$\mu_{F_j}(x) = 1/k \quad \text{for all } x \in X', j = 1, 2, \ldots, k, \quad (11)$$

and minimum fuzziness (hard partition, i.e., $\psi = 1$) when

$$\mu_{F_j}(x) = 0 \text{ or } 1 \quad \text{for all } x \in X', j = 1, 2, \ldots, k. \quad (12)$$

ψ can therefore be used as an evaluation index of partitioning.

The updating of the partition forms an essential part of the optimization procedure. A relocation is done only if it results in an improvement in the process of maximization of ψ. The iterative relocation process continues until convergence on some local maximum of ψ occurs.

At the first iteration, the cluster centers $\{v_i; i = 1, 2, \ldots, k\}$ correspond to the k initial seed points as determined in Section 3. However, in succeeding iterations the cluster center v_i is updated as [4]

$$v_i = \frac{\sum_{x \in X'} [\{\mu_{F_i}(x)\}^w * x]}{\sum_{x \in X'} \{\mu_{F_i}(x)\}^w} \quad (13)$$

where $w \geq 1$ and $i = 1, 2, \ldots, k$.

The sequence of membership evaluation and cluster updating is repeated until a local maximum value of ψ is obtained. This stage corresponds to the minimization of fuzziness in the resulting clusters and leads to an optimal partitioning of the feature space.

In order to have nonfuzzy (may be overlapping) output one may generate suitable α-cuts from the resulting μ-plane. That is, an element x can be said to belong to cluster F_i if and only if $\mu_{F_i}(x) \geq \alpha$ where $0 < \alpha \leq 1$.

6. Merging

The necessity of merging two clusters has been explained in Section 2. When the ambiguity between a pair of clusters is high, they can be merged to result in a further maximization of ψ in the feature space. For this, a measure of ambiguity between each pair of clusters may be determined by a number of factors as explained below.

(i) As mentioned in Section 5, the amount of fuzziness in the intersection between a pair of clusters F_i and F_j indicates a measure of the ambiguity (overlapping) between them. This is given by equation (9). Let it be denoted here as

$$J = L(F_j \cap F_i). \quad (14)$$

(ii) The sum of the fuzzy measures z_x in a region of intersection between a pair of clusters indicates the total pattern density in the ambiguous region. In other words, this may be viewed as some measure of the total number of pattern points that belong to both clusters. An α-cut determines the cluster boundaries as mentioned in Section 5. Therefore

$$M = \sum_{x \in X'} z_x \quad (15)$$

where $\min(\mu_{F_j}(x), \mu_{F_i}(x)) \geq \alpha$, can be regarded as another ambiguity measure for merging.

(iii) Again, for the α-cut plane, the sum of the fuzzy measures z_x within a cluster (called within-class fuzzy measure) is proportional to the total number of pattern points in that cluster. If there is a large disparity between the within-class fuzzy measures (i.e., large disparity in the number of supports or samples) of two intersecting clusters F_i

and F_j, then they can also be considered for merging. A measure of this disparity is given as

$$D = \left| \sum_{x \in F_i} z_x - \sum_{x \in F_j} z_x \right| \qquad (16)$$

where $\min(\mu_{F_j}(x), \mu_{F_i}(x)) \geq \alpha$.

For each pair of clusters F_i and F_j, a combined product

$$P = J * M * D \qquad (17)$$

may then be computed. This is chosen as an objective measure such that the pair of clusters generating the maximum value of P may be merged, if desired.

7. Implementation and results

The above-mentioned algorithm was implemented on a set of 871 Indian Telugu vowel sounds in a Consonant–Vowel–Consonant context uttered by three male speakers in the age group 30 to 35 years. The ten vowel classes $(\partial, a, i, i:, u, u:, e, e:, o, o:)$, including the shorter and longer categories, have been used. Figure 2 shows the feature space

of ten vowel classes in the F_1–F_2 plane where F_1 and F_2 correspond to the first and second vowel formant frequencies obtained through spectrum analysis of the speech data. The algorithm was implemented in Fortran-77 and run on a PDP-11 computer.

The experiment has been undertaken for various d and λ combinations. As mentioned in Section 3, the number of seed points (and hence initial clusters) increases with decrease in either d or λ. This has been described earlier in [2].

The feature space is split into a number of grid points and the fuzzy measures are computed around each such point with a suitable radius λ of the $\hat{\pi}$-function. The seed points are obtained by detecting the grid points b_i, for which the associated fuzzy set has maximum ambiguity. These correspond to the initial cluster centers. The locus of points of minimum ambiguity around each cluster center determine the initial cluster boundaries. As a typical illustration, the overlapping regions obtained in the process of extracting initial clusters (using $d = 50$ and $\lambda = 100$) are shown in Figure 3. The fuzzy measure selected was the index of fuz-
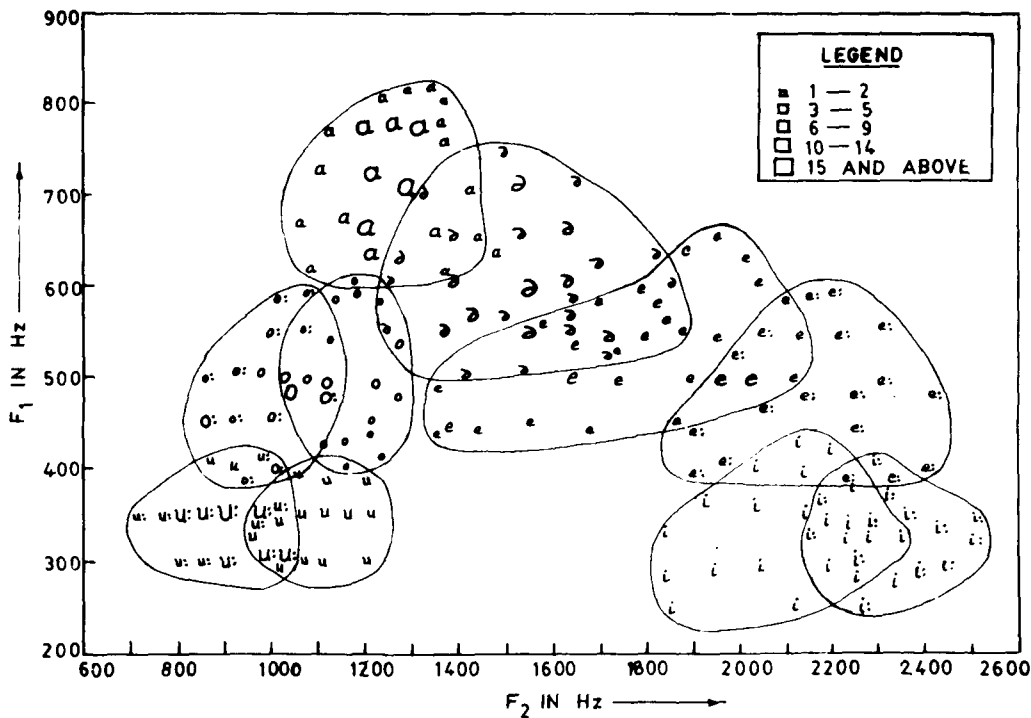

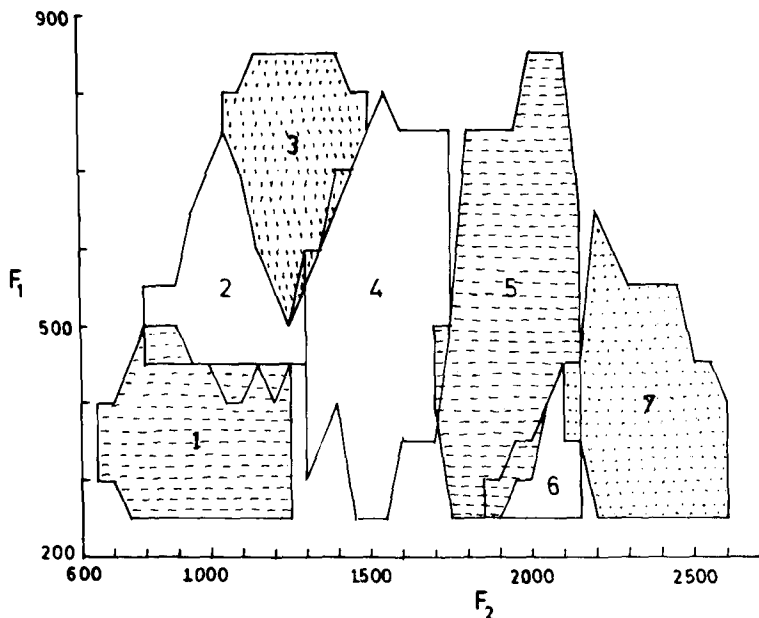Figure 2. Feature space in F_1–F_2 plane.

Figure 3. Overlapping initial clusters for $d = 50$ and $\lambda = 100$.

Table 1
Initial seed points and characterization function

Initial seed points (F_1, F_2)	Characterization function
(400, 1000)	
(500, 1000)	
(750, 1300)	
(550, 1500)	0.722
(500, 2000)	
(300, 2100)	
(350, 2250)	

ziness γ. Initially $k = 7$ clusters are obtained. The 7 initial seed points and the resulting characterization function ψ at this stage are shown in Table 1.

Table 2 depicts the updated optimum (in the sense of maximization of ψ) cluster centers (F_1, F_2) and the corresponding local maximum values of the characterization function ψ. For example, the first row of Table 2 shows the 7 cluster centers obtained from the initial seed points (Table 1) after a series of updatings. The value of ψ (0.76) is the maximum value obtained in the process of updating. If the optimum number of clusters k_0 is known, then the process is terminated at the row corresponding to k_0 clusters and the local maximum value of ψ is obtained. The

optimal cluster centers at this stage are given by the first column of this row. The pair of clusters having maximum P value (eq. (17)) are merged, when required, and the resulting clusters are correspondingly updated. The cluster pairs to be merged are shown in the second column of this table. For example, clusters 6 and 7 are merged to yield six clusters, whose updated centers and the locally maximized ψ value are shown in the second row of Table 2.

When k_0 is unknown, the process is repeated until a global maximum value of ψ is obtained. The fourth column of the corresponding row indicates the optimal number of clusters and the first column gives the resulting cluster centers.

Figure 4 shows the variation of ψ with the number of iterations for initial clusters $k = 7$. The curve depicts the behavior of ψ at each stage of merging and updating. A global maximum of ψ is seen to be obtained at $k_0 = 3$. As a typical example, consider the case with $k_0 = 7$. Here six updatings are needed to reach the local maximum value of ψ as given in the first row of Table 2. Similarly the variation of ψ for $k_0 = 6$, 5, 4, 3, 2 are shown in Figure 4. It is seen that each stage requires a different number of updatings to yield a corresponding local maximum value of ψ.

Table 2
Cluster centers and characterization function (for $k = 7$)

Cluster centers (F_1, F_2)	Clusters to be merged	Characterization function ψ	Number of clusters
(350, 950)			
(500, 950)			
(700, 1250)			
(500, 1500)	6, 7	0.76	7
(650, 1900)			
(400, 2000)			
(400, 2400)			
(350, 950)			
(500, 950)			
(700, 1250)	4, 5	0.766	6
(500, 1500)			
(550, 1950)			
(400, 2350)			
(350, 950)			
(500, 950)			
(700, 1300)	2, 3	0.763	5
(550, 1800)			
(400, 2300)			
(400, 950)			
(600, 1300)	2, 3	0.77	4
(550, 1800)			
(400, 2300)			
(450, 1000)			
550, 1600)	2, 3	0.78	3
400, 2250)			
(450, 1050)	—	0.757	2
(500, 1950)			

Figure 5 depicts the movement of the cluster centers (only for $k_0 = 7$) in the feature space, during the process of updating, leading to a local maximum value of ψ. A total of six iterations are required in the process, as observed from Figure 4. Note that different cluster centers undergo different amounts of movement in the feature space and all cluster centers do not move simultaneously. The initial seed points (Table 1) and the final updated cluster centers (first row of Table 2) are shown by the starting points and terminating points respectively of the arrows in Figure 5. It is seen that cluster center 6 undergoes a maximum of four updatings while cluster centers 1, 2 and 4 undergo a single updating each.

The vowel data has six classes (considering longer and shorter categories as the same). The optimal cluster centers obtained corresponding to $k_0 = 6$ (second row of Table 2) are seen to conform well to the vowel diagram.

8. Conclusion and discussion

A three-stage hierarchical fuzzy dynamic clustering algorithm consisting of initial clustering, updating and merging based on various characterization functions has been presented incorporating the measures of fuzziness (e.g., index
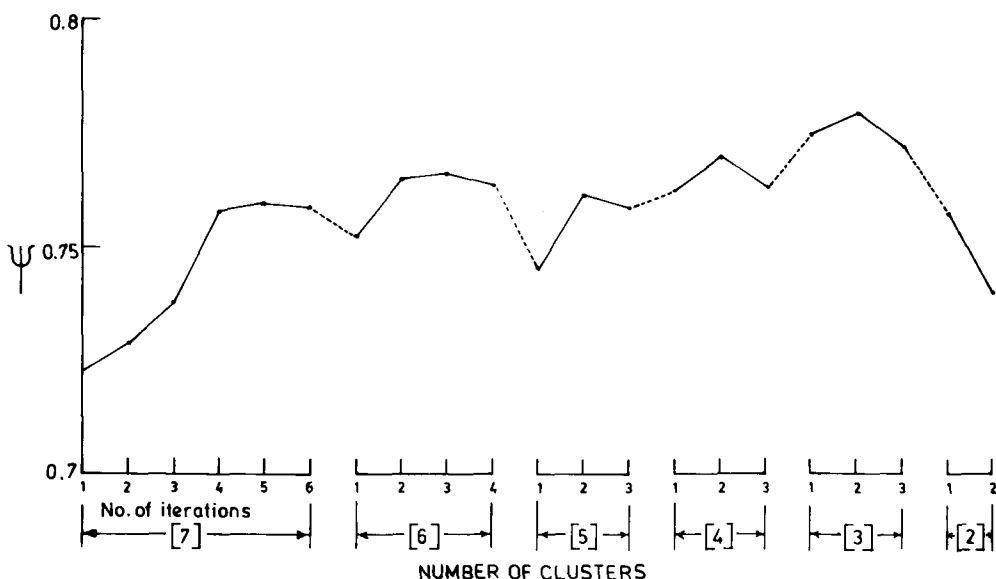

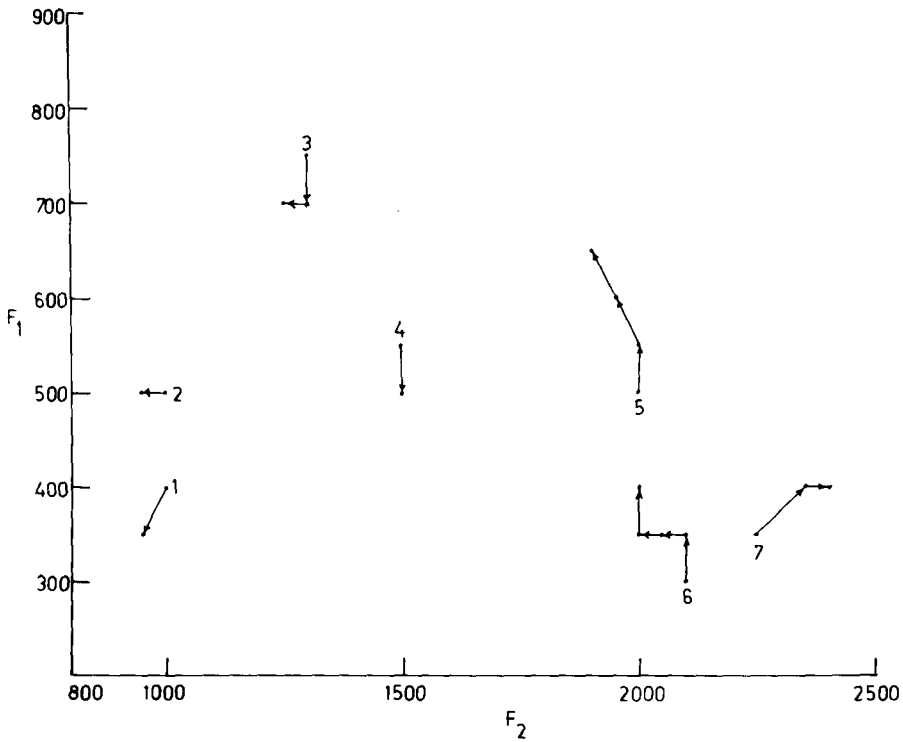Figure 4. Variation of ψ with iteration.

Figure 5. Movement of the cluster centers for $k_0 = 7$.

of fuzziness, entropy and π-ness) at every stage. Unlike the conventional detection of disjoint initial clusters, the algorithm is able to extract the hard overlapping initial cluster boundaries (as shown in Figure 3) for the ill-defined vowel regions. Membership function in $\mathbb{R}^n$ involves both Euclidean distance and density of patterns at a point. The merging criterion involves the number of points and the amount of fuzziness in the intersection of two clusters, and the disparity in their size. Varying α creates overlapping output partitions. The algorithm is able to generate an optimal number of clusters k_0 both when k_0 is known and unknown. Results at every stage are shown to demonstrate the effectiveness of the algorithm.

In this connection, mention must be made of the work of Diday & Simon [11, 12] who have used the concept of cross-partition to generate strong and weak cluster patterns in their dynamic clustering algorithm. A cross-partition is obtained by repeated intersections of k_0-partitions, resulting in a set of disjoint subsets of the pattern space. A fuzzy characteristic function based on an ultrametric

distance is used to determine the degree of similarity between two strong cluster patterns. Each weak cluster pattern consists of a lumping of a set of strong cluster patterns that are nearest to each other. Initially, the kernels are so chosen that the partitions are realized around pattern points with high density. The algorithm involves computation of probability density functions. The objective function (based on distance measure) minimizes the inertia of each cluster versus its kernel, when the number of clusters k_0 is known, in order to obtain disjoint optimum clusters.

Interestingly, the concept of overlapping clusters and the fuzziness involved has not been touched upon in their treatment. It mainly considered the hard domain of clustering. These points have also been noted by Diday & Simon [12, p. 92].

The proposed algorithm, on the other hand, takes these factors into account in all the three stages, viz., initial partitioning, membership evaluation and updating, and merging, considering k_0 unknown (or known). Both initial clusters and final output generated can be overlapping, the out-

put being characterized by the membership function or α-cut.

The fuzzy measures used here incorporate the amount of difficulty in taking a decision based on an individual sample. The recent development on higher order entropy of a fuzzy set [10], which involves various combinations of samples, may be used as a measure of fuzziness in a set to result in an improved performance.

Acknowledgement

References

[1] Pal, S.K. and D. Dutta Majumdar (1986). *Fuzzy Mathematical Approach to Pattern Recognition*. Wiley (Halsted), New York.

[2] Pal, S.K. and P.K. Pramanik (1986). Fuzzy measures in determining seed points in clustering. *Pattern Recognition Letters* 4, 159–164.

[3] Backer, E. (1978). *Cluster Analysis by Optimal Decomposition of Induced Fuzzy Sets*. Delftse Univ. Pers, Delft.

[4] Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York.

[5] Dunn, J.C. (1973). A fuzzy relative of the Isodata process and its use in detecting compact well-separated clusters. *J. Cybernet.* 3, 32–57.

[6] Bezdek, J.C. (1973). *Fuzzy Mathematics in Pattern Classification*. Ph.D. Dissertation, Cornell Univ., Ithaca, NY.

[7] Pal, S.K. and A. Rosenfeld (1988). Image enhancement and thresholding by optimisation of fuzzy compactness. *Pattern Recognition Letters* 7. 77–86.

[8] Pal, S.K. and B. Chakraborty (1986). Fuzzy set theoretic measure for automatic feature evaluation. *IEEE Trans. Syst. Man Cybernet.* 16, 754-760.

[9] Pal, S.K. (1982). A note on the quantitative measure of image enhancement through fuzziness. *IEEE Trans. Pattern Anal. Machine Intell.* 204–208.

[10] Pal, N.R. and S.K. Pal. Higher order fuzzy entropy and hybrid entropy of a set. *Inform. Sci.*, communicated.

[11] Diday, E. (1974). Optimization in non-hierarchical clustering. *Pattern Recognition* 6, 17–33.

[12] Diday, E. and J.C. Simon (1976). Clustering analysis. In: K.S. Fu, ed., *Digital Pattern Recognition*. Springer, New York, 47–94.