Pattern Recognition Letters 4 (1986) 159–164
North-Holland

July 1986

Fuzzy measures in determining seed points in clustering

S.K. PAL and P.K. PRAMANIK
Electronics and Communication Sciences Unit, Indian Statistical Institute, Calcutta 700035, India

Received 3 October 1985

Abstract: Algorithms for automatic selection of seed points for clustering are described using the terms 'index of fuzziness', 'entropy' and 'π-ness' of a fuzzy set.

Two membership functions in $\mathbb{R}^n$ have been defined such that the fuzzy measures posses maximum values when the cross-over points/central points of the membership functions correspond to the points around which the data has a tendency to cluster. The effectiveness of the algorithm is demonstrated on a set of speech data.

Key words: Clustering, fuzzy measures.

1. Introduction

The aim of a clustering technique is partioning a feature space into some homogeneous groupings. The criteria of clustering may follow a heuristic scheme or may be based on the optimisation of a certain performance index or the combination of both. It has been observed that the overall performance of iterative algorithms such as ISODATA, DYNOC etc. depends greatly on the initial choice of cluster centres (seed points). Different methods of selecting such seed points include extreme point approach, graph-theoretic approach, thresholdings etc. [1].

When the clusters to be detected are not compact and well separated (i.e., boundaries are ill-defined the fuzzy set theoretic representation has been found to provide an usefull tool for cluster analysis. In such cases, it is more natural to assign each object to a cluster with a degree of cluster membership than it is done in classical set theory where each point may either belong to a cluster or not [2, 3].

The present work demonstrates an application of the theory of fuzzy sets in determining the initial seed points in clustering. These were achieved through the terms index of fuzziness, entropy, and π-ness, which measure the amount of fuzziness present in a set. In implementing these measures, two membership functions have been defined in $\mathbb{R}^n$.

The effectiveness of the algorithm is demonstrated on a set of 871 speech data.

2. Fuzzy sets

2.1. Definition of a fuzzy set

Definition 2.1. A fuzzy set A with its finite number n_0 of supporting elements $x_1, x_2, \ldots, x_{n_0}$ in the universe of discourse U is defined as

$$A = \{(\mu_A(x_i), x_i)\}, \quad i = 1, 2, \ldots, n_0,$$

where the membership function $\mu_A(x_i)$ having positive values in the interval [0, 1] denotes the degree to which an event x_i may be a member of A. x_i is said to be a cross-over point of A if $\mu_A(x_i) = 0.5$.

2.2. *Membership functions*

The standard S-function is defined [4] as

$$S(x; a, b, c) = 0, \qquad x \le a, \qquad (1a)$$

$$= 2\left(\frac{x-a}{c-a}\right)^2, \qquad b \ge x \ge a, \qquad (1b)$$

$$= 1 - 2\left(\frac{x-c}{c-a}\right)^2, \qquad c \ge x \ge b, \qquad (1c)$$

$$= 1, \qquad x \ge c. \qquad (1d)$$

The function π defined in terms of the S-function is

$$\pi(x; b, c) = S(x; c-b, c-b/2, c), \qquad x \le c, \qquad (2a)$$

$$= 1 - S(x; c, c+b/2, c+b), \qquad x \ge c. \qquad (2b)$$

In $S(x; a, b, c)$, b is the cross-over point, i.e., $S(b; a, b, c) = 0.5$. In $\pi(x; b, c)$, b is the bandwidth, i.e., the separation between the two cross-over points of the function π. c is the central point at which $\pi = 1$.

The functions π and S represent the compatibility functions corresponding to the fuzzy sets 'x is large' and 'x is c', respectively.

Let us now define two membership functions for $x \in \mathbb{R}^n$.

The first one is

$$\hat{S}(x; b, \lambda) = \tfrac{1}{2}(1 - \|x-b\|/\lambda)^2 \quad \text{or} \quad 1 - \tfrac{1}{2}(1 - \|x-b\|/\lambda)^2, \quad \|x-b\| \le \lambda, \qquad (3a)$$

$$= 0 \text{ or } 1, \quad \text{otherwise}, \qquad (3b)$$

where $\|\cdot\|$ denotes any norm in $\mathbb{R}^n$, $\lambda > 0$ is said to be the radius of $\hat{S}(x; b, \lambda)$ and b is the cross-over point. It is to be noted that equation (3) is a two-valued function (values being complementary).

The second function may be defined in terms of equation (3) as

$$\hat{\pi}(x; c, \lambda)$$

$$= \min \hat{S}(x; y, \lambda/2), \qquad \lambda/2 < \|x-c\| < \lambda, \qquad (4a)$$

$$= \max \hat{S}(x; y, \lambda/2), \qquad 0 < \|x-c\| \le \lambda/2, \qquad (4b)$$

where $\|y-c\| = \lambda/2$, and $\min \hat{S}(x; y, \lambda/2)$ implies the minimum of the two values of the $\hat{S}$-function (equation (3)) at the point x. Similarly, $\max \hat{S}(x; y, \lambda/2)$ implies the maximum of the two values of the $\hat{S}$-function at the point x. c is the central point, i.e., $\hat{\pi}(c; c, \lambda) = 1$ and λ is the bandwidth. This is shown in Figure 1, where $x \in \mathbb{R}^2$. By simplification, (4) reduces to

$$\hat{\pi}(x; c, \lambda)$$

$$= \tfrac{1}{2}(1 - 2\|x-y\|/\lambda)^2, \qquad \lambda/2 \le \|x-c\| \le \lambda, \qquad (5a)$$

$$= 1 - \tfrac{1}{2}(1 - 2\|x-y\|/\lambda)^2, \qquad 0 < \|x-c\| < \lambda/2, \qquad (5b)$$

with $\|y-c\| = \lambda/2$.

Considering the Euclidean norm,

$$\|x-y\| = \|x-c\| - \lambda/2 \quad \text{if } \lambda/2 \le \|x-c\| \le \lambda,$$
$$= \lambda/2 - \|x-c\| \quad \text{if } 0 \le \|x-c\| \le \lambda/2. \qquad (6)$$

Using equation (6) we can further reduce equation (5) to

$$\hat{\pi}(x; c, \lambda)$$

$$= 2(1 - \|x-c\|/\lambda)^2, \qquad \lambda/2 < \|x-c\| < \lambda, \qquad (7a)$$

$$= 1 - 2\frac{\|x-c\|^2}{\lambda^2}, \qquad 0 < \|x-c\| < \lambda/2. \qquad (7b)$$

2.3. *Fuzzy measures*

2.3.1. *Index of fuzziness*

The index of fuzziness of A having n_0 supporting points is defined as [5]

$$\gamma(A) = \frac{2}{n_0^{1/k}} d(A, \underline{A}) \qquad (8)$$

where $d(A, \underline{A})$ denotes the distance between A and its nearest ordinary set $\underline{A}$ such that $\mu_{\underline{A}}(x_i) = 0$ if $\mu_A(x_i) \le 0.5$ and 1 if $\mu_A(x_i) > 0.5$. $k = 1$ for linear distance and 2 for Euclidean distance.

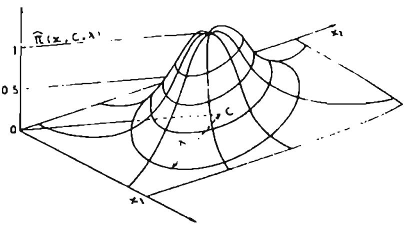

Figure 1. The function $\hat{\pi}$ in $\mathbb{R}^2$.

When d is linear, we have

$$d(A, \underset{\sim}{A}) = \sum_{i=1}^{n_0} |\mu_A(x_i) - \mu_{\underset{\sim}{A}}(x_i)|$$

and, accordingly, the linear index of fuzziness $\gamma_l(A)$ can be written as

$$\gamma_l(A) = \frac{2}{n_0} \sum_i |\mu_A(x_i) - \mu_{\underset{\sim}{A}}(x_i)| \tag{9a}$$

$$= \frac{2}{n_0} \sum_i \mu_{A \cap \bar{A}}(x_i) \tag{9b}$$

$$= \frac{2}{n_0} \sum_i \min(\mu_A(x_i), 1 - \mu_A(x_i)). \tag{9c}$$

Extension. Considering equation (3) for $x_i \in \mathbb{R}^n$, we may write

$$\gamma_l(A) = \frac{2}{n_0} \sum_i \min \mu_A(x_i) \tag{10}$$

where $\min \mu_A(x_i)$ implies the minimum of the two μ-values at the point x_i of $\hat{S}$ (equation (3)).

2.3.2. *Entropy*

The entropy is defined for the set A as [6]

$$H(A) = \frac{1}{n_0 \ln 2} \sum_i S_n(\mu_A(x_i)),$$

$$i = 1, 2, \ldots, n_0, \tag{11a}$$

where

$$S_n(x) = -\mu_A(x) \ln \mu_A(x)$$
$$-(1 - \mu_A(x)) \ln(1 - \mu_A(x)). \tag{11b}$$

Extension. For $x_i \in \mathbb{R}^n$, we may write considering equation (3),

$$H(A) = \frac{1}{n_0 \ln 2} \sum_i S_n(\mu_A(x_i)) \tag{12a}$$

where

$$S_n(\mu_A(x_i)) = -(\min \mu_A(x_i)) \ln(\min \mu_A(x_i))$$
$$-(\max \mu_A(x_i)) \ln(\max \mu_A(x_i)). \tag{12b}$$

$\min \mu_A(x_i)$ and $\max \mu_A(x_i)$ imply the minimum and maximum of the two μ-values at the point x_i of the function $\hat{S}$ respectively.

Equations (9) and (11) measure the average amount of difficulty (ambiquity) that arises when one has to take the decision whether x_i would be considered a member of A or not. These values lie in the interval [0, 1] such that

$\gamma(A)$ or $H(A) = 0$ (minimum)

when $\mu_A(x_i) = 0$ or 1 for all i, (13a)

$\gamma(A)$ or $H(A) = 1$ (maximum)

when $\mu_A(x_i) = 0.5$ for all i, (13b)

$$\gamma(A) = \gamma(\bar{A}) \quad \text{and} \quad H(A) = H(\bar{A}), \tag{13c}$$

$$\gamma(A^*) \le \gamma(A) \quad \text{or} \quad H(A^*) \le H(A), \tag{13d}$$

where A^* is a sharpened version of A such that $\mu_{A^*}(x_i) \ge \mu_A(x_i)$ for $\mu_A(x_i) \ge 0.5$ and $\mu_{A^*}(x_i) \le \mu_A(x_i)$ for $\mu_A(x_i) \le 0.5$. It follows that $\gamma(A)$ or $H(A)$ increases monotonically in the interval [0, 0.5] and decreases monotonically in [0.5, 1] with maximum value 1 at $\mu = 0.5$.

2.3.3. *π-ness* [7]

The π-ness is defined for the set A as

$$I(A) = \frac{1}{n_0} \sum_{i=1}^{n_0} \pi(x_i; b, c) \tag{14}$$

where $\pi(x_i; b, c)$ is defined by equation (2).

For $x \in \mathbb{R}^n$, the π-ness is defined as

$$I(A) = \frac{1}{n_0} \sum_{i=1}^{n_0} \hat{\pi}(x_i; c, \lambda). \tag{15}$$

2.4. *Concept of fuzzy sets in extracting seed points*

Let $X = \{X_1, X_2, \ldots, X_N\}$ be the set of N pattern points in the n-dimensional feature space ($n \ge 2$). We define the fuzzy set associated with the set X as

$$X(b, \lambda) = \{\mu_{X(b,\lambda)}(X_i), X_i\}, \quad i = 1, 2, \ldots, N \tag{16}$$

where

$$\mu_{X(b,\lambda)}(X_i) = \hat{S}(X_i; b, \lambda) \text{ or } \hat{\pi}(X_i; b, \lambda).$$

b corresponds to the cross-over point for the function $\hat{S}$ and the central point for the function $\hat{\pi}$. If we keep λ constant and change b we get different fuzzy sets, i.e., changing b we can generate a class of fuzzy sets. From the expressions (3), (4), (10), (12), (15) and (13) we see that the contributions towards $\gamma(X(b, \lambda))$ or $H(X(b, \lambda))$ or $I(X(b, \lambda))$ are mostly from those points which are around b and it decreases as the points move away from b. In other words, if the number of points around b is more there will be a greater number of points X_j having $\mu=0.5/=1$ while using the function $\hat{S}$/function $\hat{\pi}$ (resulting in γ, H and $I \subseteq 1$) and a less number of points having $\mu=0$ or $1/=0$ while using the function $\hat{S}$/function $\hat{\pi}$ (resulting in γ, H and $I=0$) thus increasing the value of $\gamma(X(b,\lambda))$ or $H(X(b,\lambda))$ or $I(X(b,\lambda))$. Therefore the more points from fuzzy set $X(b,\lambda)$ are compact around b, the greater would be its γ or H or I value and b can be considered as a seed point (centre of an initial cluster).

This suggests that modification of the cross-over point/central point b will result in variation of the measures $\gamma(X(b,\lambda))$, $H(X(b,\lambda))$ and $I(X(b,\lambda))$ and so a set of seed points $\{b\}$ may be estimated for which the corresponding fuzzy measures are locally maximum.

3. Algorithm for determining seed points using fuzzy measures

Let $X_1, X_2, \ldots, X_N$ be the N pattern points each of them having n properties, i.e., they are samples from the n-dimensional feature space. Let l_i, u_i be the lower and upper bounds of the i-th property of the sample. Let us split the space $(l_1,u_1)\times(l_2,u_2)\times\cdots\times(l_n,u_n)$ into L^n grid points where L is given by $(u_i-l_i)/d$ for a fixed i and d is some preassigned positive constant called grid width. Let b_i, $i=1,2,\ldots,L^n$, be the grid points.

Choosing λ suitably, calculate the fuzzy measures using any of the equations given below:

$$\gamma_i(X(b_i,\lambda)) = \frac{2}{N}\sum_{j=1}^{N} \min(\mu_{X(b_i,\lambda)}(X_j)), \tag{17a}$$

$$H(X(b_i,\lambda)) = \frac{1}{N\ln 2}\sum_{j=1}^{N} S_n(\mu_{X(b_i,\lambda)}(X_j)), \tag{17b}$$

$$I(X(b_i,\lambda)) = \frac{1}{N}\sum_{j=1}^{N} \hat{\pi}(X_j, b_i, \lambda), \quad i=1,2,\ldots,L^n. \tag{17c}$$

The grid points $\{b_i\}$ for which the corresponding fuzzy measures are locally maximum may be taken as initial seed points.

4. 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 speakers in the age group of 30 to 35 years [8]. Figure 2 shows the feature space of six vowels

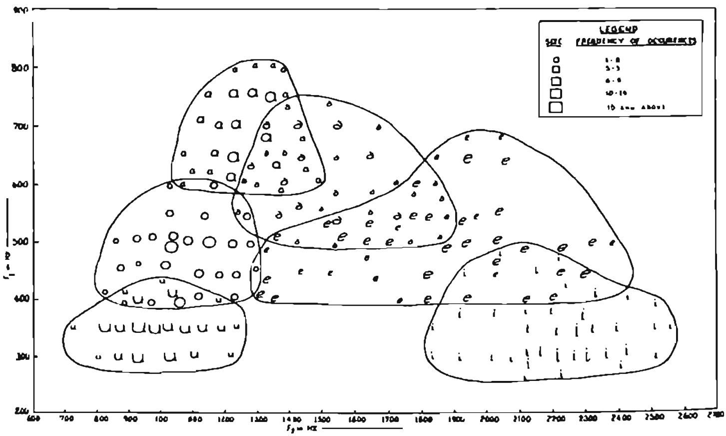

Figure 2. Vowel classes in the (F_1-F_2) plane.

($\partial, a{:}, i, u, e, o$) corresponding to F_1 and F_2. F_1 and F_2 denote the first and second vowel formant frequencies which were obtained through spectrum analysis of the speech data.

Table 1 shows the (F_1, F_2)-values for which the fuzzy measures γ, H and I were found to be locally maximum when d and λ were considered to be 50 and 100 respectively. A set of 5 to 7 maxima is observed for different measures. The seed points as given by these (F_1, F_2)-values are found to agree well with those of the classes in Figure 2.

In order to demonstrate the effect of the parameters d and λ on the selection of seed points, we have considered only the measure 'entropy', and Tables 2 and 3 illustrate such an effect when d and λ were chosen to be 25, 50 and 100, and 100, 150 and 200 respectively.

At low values, viz. $d = 25$, $\lambda = 100$ (Table 2a), the number of seed points is found to be 9 and as we increase λ some of the weak local maxima get lost leaving behind only the strong ones. A similar effect is also found from Tables 2(a), 1(b) and 3 to occur when one increases the grid width for a fixed value of λ.

Table 3
Seed point and the corresponding H-values for $d = 100$ and $\lambda = 100$

Seed point (F_1, F_2)	$H \times 10^{-2}$
(350, 2123)	2.430
(450, 985)	5.797
(550, 1838)	1.697
(750, 1269)	2.960

5. Conclusions

A method has been outlined using the concept of fuzzy sets whereby a satisfactory choice of initial seed points for clustering a given data set may be determined.

Table 1
Seed points and corresponding fuzzy measures when $d = 50$ and $\lambda = 100$, (a) γ_I, (b) H, (c) I

(a)		(b)		(c)	
Seed point (F_1, F_2)	$\gamma \times 10^{-2}$	Seed point (F_1, F_2)	$H \times 10^{-2}$	Seed point (F_1, F_2)	$I \times 10^{-2}$
(350, 2265)	2.152	(350, 2265)	4.172	(350, 2265)	3.709
(400, 985)	3.316	(450, 985)	5.797	(500, 985)	5.464
(500, 985)	3.750	(500, 1981)	2.022	(500, 1981)	1.885
(500, 1981)	1.237	(550, 1554)	1.478	(550, 1554)	1.267
(600, 1411)	0.079	(750, 1269)	2.960	(600, 1838)	1.512
(600, 1838)	0.091			(750, 1269)	2.742
(750, 1269)	1.651				

Table 2
Seed points and corresponding H-values for $d = 25$, (a) $\lambda = 100$, (b) $\lambda = 150$, (c) $\lambda = 200$

(a)		(b)		(c)	
Seed point (F_1, F_2)	$H \times 10^{-2}$	Seed point (F_1, F_2)	$H \times 10^{-2}$	Seed point (F_1, F_2)	$H \times 10^{-2}$
(300, 1483)	0.106	(350, 2265)	7.301	(350, 2265)	10.200
(350, 2265)	4.172	(425, 985)	10.898	(425, 985)	16.167
(475, 985)	5.997	(525, 1838)	3.800	(500, 1910)	6.258
(500, 1838)	1.795	(550, 1554)	2.619	(675, 1269)	8.016
(500, 1981)	2.022	(700, 1269)	5.384		
(525, 1554)	1.498				
(600, 1838)	1.685				
(650, 1198)	2.794				
(725, 1269)	3.158				

References

[1] Tou, J.T. and R.C. Gonzalez (1974). *Pattern Recognition Principles*. Addison-Wesley, London.

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

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

[4] Zadeh, L.A. (1975). Calculus of fuzzy restrictions. In: L.A. Zadeh et al., Eds., *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*. Academic Press, London, 1-39.

[5] Kaufmann, A. (1975). *Introduction to the Theory of Fuzzy Subsets*, Vol. 1, Academic Press, New York.

[6] De Luca, A. and S. Termini (1972). A definition of non-probabilistic entropy in the setting of fuzzy sets theory. *Inf Control*, 20, 301-312.

[7] Pal, S.K. (1982). Fuzzy set theoretic approach - A tool for speech and image recognition. In: J. Kittler et al., Eds., *Pattern Recognition Theory and Applications*, Reidel, Dordrecht, Holland, 103-117.

[8] Pal, S.K. and D. Dutta Majumder (1977). Fuzzy sets and decision making approaches in vowel and speaker recognition. *IEEE Trans. Syst. Man Cybernet.* 7, 625-629