

ハンス・アルプ 星座 1932

ファジイ・ クラスタリング

大 隅 昇

1. はしがき

クラスタ分析の変遷推移を追うと1960年代半ばから急速に発展したことがわかる。一方、ファジイ理論は同時期に Zadeh により提唱され、その後著しい進展をみた概念である。このように両者はほぼ同時期にそれぞれ独自に広い分野に浸透していったといういきさつがある。そして両者の間には実に密接な関係がみられるのだが（とくに後述の5節にみられるような種々の性質）、それより早く、すでに1966年に Bellman, Kalaba, Zadeh によるファジイ理論の分類問題への適用に関する示唆がみられる。

その後、ファジイ理論と統計理論とをたくみに取り入れた Ruspini の方法が現われ(1969)、これが“ファジイ・クラスタリング”のはしりとなった。これに続いて、Zadeh の精力的な研究の一部として、“ファジイ関係 (fuzzy relation)”や“ファジイ・グラフ”が現われ、この中に若干ではあるが“数値分類法”と関連した意見がみられる (1971)。これは系統樹 (デンドログラム) を作る分類手法である“階層的分類法”ときわめて密接な関係にあり、彼の示唆はクラスタ分析の研究上で大変重要な意味を持っている。これと同時期に田村らの類似の考え方が現われるが (1971)、実はこうした考え方が、数値分類法の分野で利用されてきた数多くの発見的な手法を体系的に説明する場合に大変便利な道具となるのである。さらに1970年代に入ってからいくつか目立った研究報告が現われる。Gitman らの“モード探索法”や、従来の反復分割最適化型手法とファジイ理論の考え方を結びつけた Bezdek, Dunn らの“ファジイ k -means 法”などがそれである。その後 Lee の染色体の分類問題への応用 (1975, 1976) や、Backer (1977), Rosen-

feld (1974), Halpern (1975) など多くの報告が現れている。

2. Ruspini の方法

前述のようにファジイ・クラスタリングのきっかけを作った1人が Ruspini である。彼の方法は次のようなものである。

いま n 個の分類対象 (個体と呼ぶことにする) を、

$$E = \{O_1, O_2, \dots, O_i, \dots, O_n\} \text{ または } \{1, 2, \dots, i, \dots, n\} \quad (1)$$

で表わす。また第 i 個体の p 変量観測ベクトルを、

$$\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip}, \dots, x_{ip})$$

で表わし多変量データ行列全体を、

$$X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n) \quad (2)$$

と書く。いま個体の集合 E を適当に分割して互いに排反な空でない g 個のクラスタ (つまりファジイ部分集合) が与えられたものとする。これを、

$$\tilde{I} = \{\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_j, \dots, \tilde{S}_g\} \quad (3)$$

で表わす。次にある \tilde{S}_j 内にある個体の割合の見積り (一種の事前確率) を $P(\tilde{S}_j)$ と書きその全体を、

$$\tau = \{P(\tilde{S}_1), P(\tilde{S}_2), \dots, P(\tilde{S}_j), \dots, P(\tilde{S}_g)\} \quad (4)$$

で表わす。この τ は次のようにして作る。いま、 \mathbf{x}_i の確率分布が与えられたものとしてそれを $p_i = P(\mathbf{x}_i)$ とかき、これをまとめて、

$$\mathbf{p} = (p_1, p_2, \dots, p_i, \dots, p_n) \quad (5)$$

とする。つぎに \mathbf{x}_i が \tilde{S}_j に所属する度合、すなわち所属関数を $\mu_{ij} = P(\tilde{S}_j | \mathbf{x}_i)$ で表わしその全体を行列で

$$\Pi = (\mu_{ij}) \quad (i=1, 2, \dots, n; j=1, 2, \dots, g) \quad (6)$$

とかく。このとき $\tau = \mathbf{p}\Pi$ (つまり $P(\tilde{S}_j) = \sum_{i=1}^n P(\mathbf{x}_i) \cdot P(\tilde{S}_j | \mathbf{x}_i)$) と与えられる。ところですべての \mathbf{x}_i と \tilde{S}_j に

対して、ベイズの定理から

$$P(\mathbf{x}_i|\tilde{S}_j) = P(\tilde{S}_j|\mathbf{x}_i) \cdot P(\mathbf{x}_i) / P(\tilde{S}_j) \quad (7)$$

となるので、これを行列 $Q=(q_{ij})$ と書く。このとき $\mathbf{p} = \tau Q$ がなりたつ。

次に個体間の距離 $d(\mathbf{x}_i, \mathbf{x}_k)$ を求め距離行列 $D = \{d(\mathbf{x}_i, \mathbf{x}_k)\}$ を作る。さらに次の重み関数を用意する。

$$m(\mathbf{x}_i) = \sum_k^n P(\mathbf{x}_k) d(\mathbf{x}_i, \mathbf{x}_k) \quad (8)$$

これはまとめて $\mathbf{m} = \mathbf{p}D$ とかける。さらに、

$$M_j(\mathbf{x}_i) = \sum_k^p P(\mathbf{x}_k|\tilde{S}_j) d(\mathbf{x}_i, \mathbf{x}_k) \quad (9)$$

を作る。ここで $m(\mathbf{x}_i)$ は \mathbf{x}_i のまわりの分布の平均密度であり、 $M_j(\mathbf{x}_i)$ は \mathbf{x}_i のまわりのクラスター \tilde{S}_j の平均密度と考えられる。 $M_j(\mathbf{x}_i)$ を η_{ij} とかきこれを新たに行列 M で表わす。つまり、 $M=(\eta_{ij})$ ($i=1, \dots, n$, $j=1, \dots, g$) と書くと $M=QD$ となり $\tau M = \mathbf{m}$ が成り立つ。以上の約束のもとに次の基準の最適化をはかる。

【基準 I】

$$\begin{aligned} P(\tilde{S}_j^*) &= \max_j P(\tilde{S}_j|\mathbf{x}_i) \\ &= \max_j \mu_{ij} \quad (i=1, 2, \dots, n) \end{aligned} \quad (10)$$

となる \tilde{S}_j を探しこれを各個体 i について集めたものが \tilde{S}_i^* である。このとき、

$$J = \sum_{i=1}^n \left[P(\mathbf{x}_i) \cdot \frac{P(\tilde{S}_i^*) \cdot M_i^*(\mathbf{x}_i)}{m(\mathbf{x}_i)} \right]^2 \Rightarrow (\text{最小化}) \quad (11)$$

となる行列 Π を求めること。ここで $M_i^*(\mathbf{x}_i)$ はある \mathbf{x}_i について(10)式により選ばれた j に対応する η_{ij} の値である。

この基準は、所属する可能性がもっとも強いクラスターまでの各点の平均距離の平均を最小化することに相当している。

【基準 II】

$$\begin{aligned} J &= \sum_{i=1}^n P(\mathbf{x}_i) \sum_{j=1}^g P(\tilde{S}_j) \cdot \frac{P(\tilde{S}_j) \cdot M_j(\mathbf{x}_i)}{m(\mathbf{x}_i)} \\ &\quad \times P(\tilde{S}_j|\mathbf{x}_i) \Rightarrow (\text{最小化}) \end{aligned} \quad (12)$$

にする Π をさがすこと。つまり、相対的にみた $P(\tilde{S}_j|\mathbf{x}_i) \cdot M_j(\mathbf{x}_i)$ (行列 $\Pi Q D$ の要素) を最小化することである。

実際には上の基準を最適化する算法と計算プログラムが必要となるが、ここでは上の考え方を試行錯誤的に追ってみる。

【数値例】

表 1 の 2 次元データ X が与えられこれを 2 群 ($g=2$) にクラスター化することを考えよう。

まず式(5)で

表 1

個体番号 i	x_{i1}	x_{i2}
1	0	0
2	0	1
3	5	4
4	5	5
5	3	4
6	1	0

$$\mathbf{p} = \left(\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6} \right)$$

とおく。次に式(6)で、

$$\Pi^{(0)} = \begin{bmatrix} 1 & 0.9 & 0.2 & 0.1 & 0.8 & 0.7 \\ 0 & 0.1 & 0.8 & 0.9 & 0.2 & 0.3 \end{bmatrix}'$$

と見積ってみる。式(10)をつかって、 \sim 印の値を選べば、 $P(\tilde{S}_1^*)=1$, $P(\tilde{S}_2^*)=0.9, \dots, P(\tilde{S}_6^*)=0.7$ となる。さらに $\tau = \mathbf{p}\Pi^{(0)} = [0.617, 0.383]$ となり式(7)からは、

$$Q = \begin{bmatrix} 0.270 & 0.243 & 0.054 & 0.027 & 0.216 & 0.189 \\ 0 & 0.044 & 0.348 & 0.392 & 0.087 & 0.131 \end{bmatrix}$$

を得る。次に距離行列 D を作り (平方ユークリッド距離とした) \mathbf{m}, M を求める。

$$D = \{d(\mathbf{x}_i, \mathbf{x}_k)\} = \begin{bmatrix} 0 & & & & & \\ 1 & 0 & & & & \\ 41 & 34 & 0 & & & \\ 50 & 41 & 1 & 0 & & \\ 25 & 18 & 4 & 5 & 0 & \\ 1 & 2 & 32 & 21 & 20 & 0 \end{bmatrix}$$

$$\mathbf{m} = \mathbf{p}D = [19.67, 16, 18.67, 19.67, 12, 16]$$

$$M = QD$$

$$= \begin{bmatrix} 9.396 & 7.479 & 26.271 & 28.566 & 15.255 & 7.911 \\ 36.218 & 29.732 & 6.428 & 5.338 & 6.764 & 29.036 \end{bmatrix}$$

となるから、 $\Pi^{(0)}$ から選出した $P(\tilde{S}_i^*)$ に対応させて M から \sim 印の要素を選べば、これが $M_i^*(\mathbf{x}_i)$ となる。よって式(11)から、

$$\begin{aligned} J^{(0)} &= \left[\left(\frac{1 \times 9.396}{19.67} \right)^2 + \left(\frac{0.9 \times 7.479}{16} \right)^2 \right. \\ &\quad + \left(\frac{0.8 \times 6.428}{18.67} \right)^2 + \left(\frac{0.9 \times 5.338}{19.67} \right)^2 \\ &\quad + \left(\frac{0.8 \times 15.255}{12} \right)^2 + \left(\frac{0.7 \times 7.911}{16} \right)^2 \left. \right] \\ &\quad \times \left(\frac{1}{6} \right)^2 \doteq 0.04708 \end{aligned}$$

を得る。続いて $\Pi^{(0)}$ の見積りを

$$\Pi^{(0)} = \begin{bmatrix} 1 & 0.8 & 0.2 & 0 & 0.4 & 0.9 \\ 0 & 0.2 & 0.8 & 1 & 0.6 & 1 \end{bmatrix}'$$

とかえて、上と同様の計算により $J^{(1)}$ を求めると、 $J^{(1)}$

≈ 0.01171 となる。つまり $\Pi^{(0)}$ より $\Pi^{(1)}$ の方が改良されたことになる。さらに、

$$\Pi^{(2)} = \begin{bmatrix} 1 & 0.9 & 0.2 & 0 & 0.2 & 0.9 \\ 0 & 0.1 & 0.8 & 1 & 0.8 & 0.1 \end{bmatrix}'$$

とすると、 $J^{(2)} = 0.008$ を得る。以上をくり返して微少な ε に対して $|J^{(k+1)} - J^{(k)}| < \varepsilon$ となって J に改良がみられなくなるまで Π を更新する。

上で得られた $\Pi^{(2)}$ を、データ X の散布図の中に書き込むと図1がえられる。こうして各個体が2つのクラスターにそれぞれどの程度の度合で所属しているかを見積ることができる。

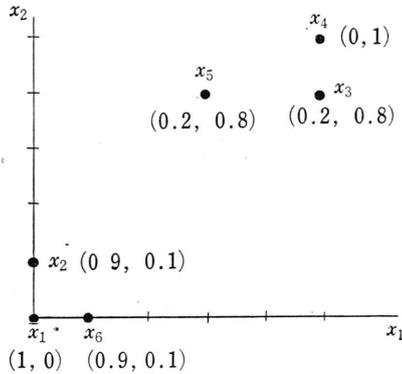


図 1

この方法では、初期クラスターの与え方、初期の事前確率の与え方、 J を最適化する方法などが問題となるが、これについては Ruspini が2次計画法などを使った算法や種々の実験結果を報告している (1970, 1973)。

3. ファジィ k -means 法

従来よく利用されてきた分割最適化型の分類手法に k -means 法がある。これはクラスターの等質性基準としてクラスター内平方和を用い、これを適当な算法で最小化するという方法である (詳しくは [26] などを見るとよい)。あるいは“クラスター内分散最小化方式 (minimum variance approach)” や “クラスター内平方ユークリッド距離最小化方式” もこれに類する方法である。こうした方式を“重みつき最小二乗法”の問題として扱い、その重みとして適当な所属関数を工夫するというのが Bezdek, Dunn らの考えたファジィ k -means 法である。

いま式(3)のような g 個の分割に対して、各個体が各クラスターに所属する度合を次の行列 (membership matrix) で表わす。

$$U = (u_{ji}) \quad (j=1, 2, \dots, g; i=1, 2, \dots, n) \quad (13)$$

ここで、 $u_{ji} \in [0, 1]$, $\sum_{j=1}^g u_{ji} = 1$ である。

つまり u_{ji} が所属関数でありこれで個体 i がクラスター S_j に所属する度合を示す。このとき次の関数の最適化を考える。

$$J_m(U, \mathbf{v}) = \sum_{i=1}^n \sum_{j=1}^g (u_{ji})^m \|\mathbf{x}_i - \mathbf{v}_j\|^2 \quad (1 \leq m < \infty) \quad (14)$$

ここで $1 < m < \infty$ に対して、

$$u_{ji} = \frac{1}{\sum_{k=1}^g \left(\frac{\|\mathbf{x}_i - \mathbf{v}_k\|^2}{\|\mathbf{x}_i - \mathbf{v}_j\|^2} \right)^{1/(m-1)}} \quad \begin{matrix} (1 \leq j \leq g) \\ (1 \leq i \leq n) \end{matrix} \quad (15)$$

$$\mathbf{v}_j = \frac{\sum_{i=1}^n (u_{ji})^m \mathbf{x}_i}{\sum_{i=1}^n (u_{ji})^m} \quad (1 \leq j \leq g) \quad (16)$$

であり、 \mathbf{v}_j はクラスター S_j の平均ベクトルである。いま $m=1$, $u_{ji} \in \{0, 1\}$ と考えるといわゆる通常の k -means 法となり J_1 は平方和基準そのものである。 $m=1, 2$ に対して式(14)の J_m を最小化する重み係数が式(15), (16)の形により与えられることは典型的な極値問題としてラグランジュ未定係数法などを使って示すことができる。これについては Dunn の報告がある。そしてこれを一般の m にまで拡張をはかったのが Bezdek である。実際の算法は k -means 法と良く類似している。つまり k -means 法に上の U と \mathbf{v} を更新するルーチンをつけ加えればよい。

(手順1) クラスター数 g を定め、 U の初期条件 $U^{(0)}$ を適当に与える。

(手順2) 式(16)により初期クラスターの平均ベクトル $\mathbf{v}_j^{(0)} (j=1, 2, \dots, g)$ を求める。

(手順3) 式(15)により $U^{(0)}$ を $U^{(1)}$ に更新する。これを演算子 T を使って、 $U^{(1)} = T_m^{(1)}(U^{(0)})$ と書く。これを反復する。すなわち、

$$U^{(k+1)} = T_m^{(k+1)}(U^{(k)}) \quad (k=0, 1, 2, \dots)$$

(手順4) 適当な収束判定値 ε を与えて、 $|U^{(k+1)} - U^{(k)}| \leq \varepsilon$ となったとき計算終了。さもなければ上の計算を繰り返す。

こうして最終的に得られた \hat{U} の要素 \hat{u}_{ji} によって個体 i がクラスターに所属する度合を知る、というのがこの手法の要である。通常の k -means 法であれば求めたクラスター集合 $\Gamma = \{S_1, S_2, \dots, S_j, \dots, S_g\}$ に対し、ある個体 i が S_j に所属している ($u_{ji}=1$) か、所属していないか ($u_{ji}=0$) を知るだけであるが、この方式によると各クラスターへの所属の度合を知ることができるというところが面白い。この行列 U を使って、さらに別の指標を考えることもできる。

$$S_g(U) = \frac{1}{n} U U' \quad (U' \text{は } U \text{ の転置行列}) \quad (17)$$

この行列は“パラメータ行列”と名づけられている。行列の (j, k) 要素を s_{jk} で表わすと、 $s_{jk} = \sum_{i=1}^n u_{ji} u_{ik} / n$ であるが対角要素 s_{jj} はクラスター \tilde{S}_j の相対的な“まとまりの良さ”を示し s_{jk} は \tilde{S}_j と \tilde{S}_k の間の“重複の程度（あるいは接近度）”を示していると考えられる。

さらに $S_g(U)$ のトレースをとって、

$$\begin{aligned} \text{tr } S_g(U) &= F_g(U) \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^g (u_{ji})^2 \end{aligned} \quad (18)$$

を作ると $g^{-1} \leq F_g(U) \leq 1$ であって、これを U の“分割係数 (partitioning coefficient)”という。これは (ファジイ) 分割 \tilde{I} 内の任意のクラスターの対の間の平均的な連結度 (対でみた相対的なファジイネス) を表わす指標と考えられる。

この方法は J_m の最適化を反復法で行うのでその解は局所最適解となるが、 k -means 法やそれに類した方法の計算プログラムの簡単な手直しによって計算ができるという利点をもっている。ここで Bezdek の示した例のうちの1つを取り上げよう。

[数値例]

表2の2次元データがある。この散布図は図2のようになりいわゆる“touching cluster”を作っている。このデータに対して、クラスター数を $g=2$ と与えて通常の k -means 法を適用すると個体番号1~13が1つのクラスターを構成し残りの14~25が第2のクラスターを作る (この計算は筆者作成のプログラムによる。具体的な算法については[26]を参照)。こうして通常の k -means 法による各個体のクラスターへの所属決定ができるが、さらに u_{ji} を見積ると表3が得られる。これを観察すると図2の布置でくびれた部分にある2つの個体13, 14の

表 2

個体番号 i	x_{i1}	x_{i2}	個体番号 i	x_{i1}	x_{i2}
1	2	11	14	10	10
2	3	7	15	12	7
3	3	13	16	12	9
4	4	15	17	12	12
5	4	10	18	14	6
6	5	8	19	14	8
7	5	12	20	13	11
8	6	14	21	14	13
9	6	6	22	16	7
10	7	11	23	16	10
11	8	8	24	17	12
12	8	13	25	18	9
13	9	11			

表 3

個体番号 i	u_{1i}	u_{2i}	個体番号 i	u_{1i}	u_{2i}
1	0.932	0.068	14	0.414	0.586
2	0.865	0.135	15	0.141	0.859
3	0.932	0.068	16	0.079	0.921
4	0.874	0.126	17	0.186	0.814
5	0.978	0.022	18	0.105	0.895
6	0.911	0.089	19	0.023	0.977
7	0.984	0.016	20	0.055	0.945
8	0.890	0.110	21	0.136	0.864
9	0.760	0.240	22	0.070	0.930
10	0.944	0.056	23	0.036	0.964
11	0.710	0.290	24	0.101	0.899
12	0.801	0.199	25	0.089	0.911
13	0.663	0.337			

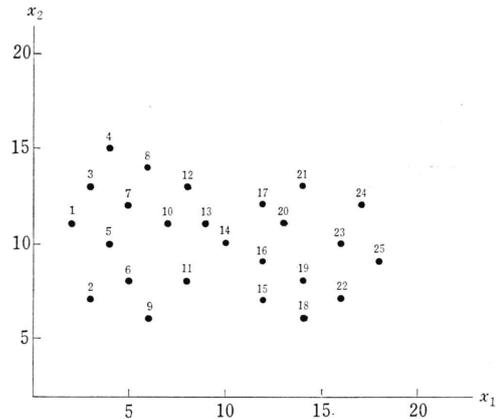


図 2

所属関数の値が他のもののそれにくらべてきわだった特徴を持っていることがわかる。つまりこの2つの個体が残りのものにくらべて2つのクラスターへの所属判定があいまいであることを知ることができる。

さらに式(17), (18)を使って $S_2(U)$, $F_2(U)$ を求めると、それぞれ、

$$S_2(U) = \begin{pmatrix} s_{11} & s_{12} \\ s_{21} & s_{22} \end{pmatrix} = \begin{pmatrix} 0.4056 & 0.1016 \\ 0.1016 & 0.3913 \end{pmatrix}$$

$$F_2(U) = 0.7969$$

となる。 s_{11} と s_{22} の値は近いので2つのクラスターのまとまりの程度は似ていることがわかる。また s_{12} の値は s_{11}, s_{22} にくらべて極端に小さいというほどではないから2つのクラスターの接近度が若干あるということもわかる。

Bezdek の方法は計算手続きとしては重みつき最小二乗法の応用であるが、重み係数に所属関数に相当するものを工夫して取り入れたところに特徴がある。もちろん J_m の性質から考えてクラスターの形状に若干の制約が

生まれることはまぬがられない。つまりクラスターのまとまりの度合を平方和ないしはそれに類した指標で測るのであるからそれが有効に働くような形状を探る場合に適した方法とみられる。たとえば図3の(a), (d)のようなデータに向いており(b), (c)のような布置に対しては後述のグラフ的な方法が適している(こうしたいい方はその布置をみるのが難かしい多次元データを扱う一般の場合に対して矛盾するかもしれない。しかし上のような最適化基準を設定する背景には、クラスターは塊状のものとの前提があり、こうしたクラスターを“生成”する方法となっているということである)。

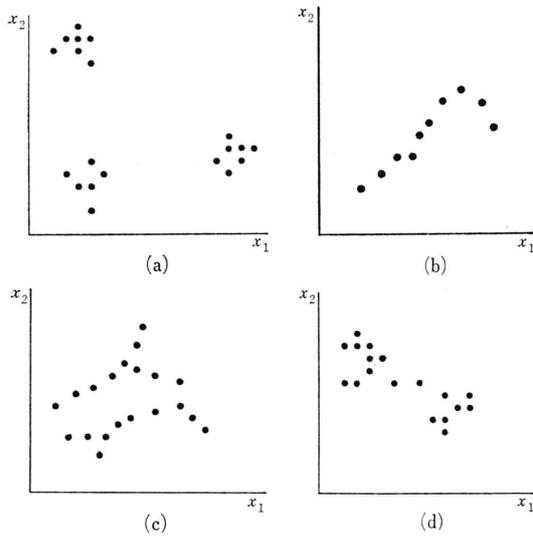


図 3

以上の2つの方法にみられるような、統計的手法とファジイ理論を結びつけた方法ばかりでなく別のアプローチもある。

4. モード探索法

次に Gitman の“モード探索法 (mode seeking technique)”をみよう。いま、ある個体 i の近傍 (たとえば i の観測ベクトル \mathbf{x}_i から半径 α 内の領域) に入る i 以外の個体の数を $f(\mathbf{x}_i)$ で表わすとこれを1つの所属関数とみなすことができる。すなわち、

$$A_i = \{\mathbf{x} | d(\mathbf{x}_i, \mathbf{x}) \leq \alpha\} \quad (i=1, 2, \dots, n) \quad (19)$$

を各個体について調べ $f(\mathbf{x}_i) = \text{card } A_i$ を求める (ここで $\text{card } A_i$ は A_i 内の要素数)。なお $d(\mathbf{x}_i, \mathbf{x})$ は個体間の (メトリックな) 距離とする。そしてこの $f(\mathbf{x}_i)$ で特徴づけられるファジイ集合 B を考え、この B の性質として次の2つの種類のファジイ集合を考える。

① B 内にある任意の個体 i, k に対して、

$$d(\mathbf{x}_i, \mu) \leq d(\mathbf{x}_k, \mu) \iff f(\mathbf{x}_i) \geq f(\mathbf{x}_k) \quad (20)$$

を満たすファジイ集合を“対称なファジイ集合 (symmetric fuzzy set)”という。ここで μ は $f(\mu) = \sup_{\mathbf{x} \in B} [f(\mathbf{x})]$ となる、つまり B 内において最大の所属関数値を与える \mathbf{x} のことである。

② 次に B 内のすべての個体 i に対して連結となる (connected) ような集合 $\tilde{I}_{\mathbf{x}_i} = \{\mathbf{x} | f(\mathbf{x}) \geq f(\mathbf{x}_i)\}$ を考え、これを満たす B はファジイの意味で“単峰的 (unimodal)”であると約束する (いわゆる level set を作ったとき、それが互いに素とならぬ集合を構成し、 B 内のすべての個体が必ずその同一集合の中に含まれているということ)。

以上の準備のもとに与えられた個体集合 E を次の手順で単峰的なファジイ部分集合に分割するという方法がモード探索法である。

【手順1】 まず X あるいは E を①の条件を満たすように区分して (対称な) ファジイ部分集合 $\tilde{S} = \{\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_m\}$ を作る。そしてある \tilde{S}_j 内の最大の $f(\mathbf{x}_i)$ をとる \mathbf{x}_i を \tilde{S}_j の“モード” μ_i とする。すなわち $f(\mu_i) = \max_{i \in \tilde{S}_j} \{f(\mathbf{x}_i)\}$ である。

【手順2】 つぎに各 \tilde{S}_j を調べて μ_i が②を満たすような“極大値”となるかどうかを検証する。もし μ_i が極大値となり得るときにはそれを改めて ν_k で書く。こうしてえられた ν_k があわせて q 個あったとする。こうして、 $\{\nu_1, \nu_2, \dots, \nu_q\}$ が生成される。

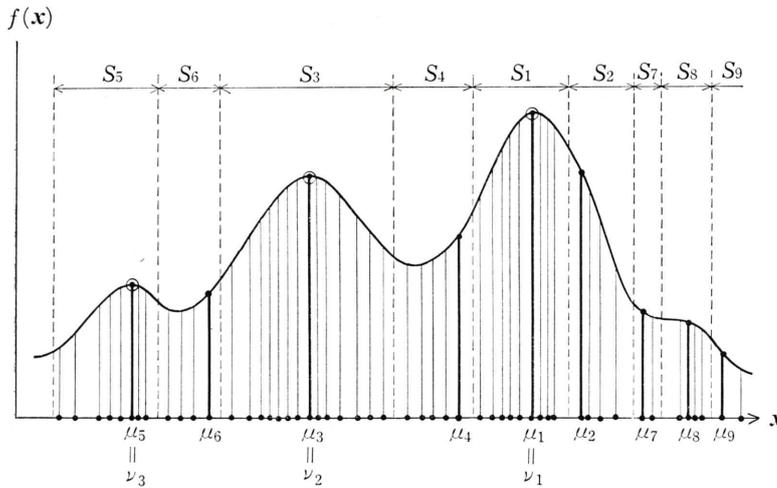
【手順3】 そして $\nu_k (k=1, 2, \dots, q)$ を除くすべての個体の観測ベクトルを“最近隣法”に類似した方式で最も近い ν_k に逐次配置する。こうしてすべての個体の所属が確定するのでこれを最終クラスターとして採用する。

図4は上の手順を1次元のデータについて模式化したものである。まず式(20)を満たすファジイ集合 $\tilde{S} = \{\tilde{S}_1, \tilde{S}_2, \dots, \tilde{S}_9\}$ を作る。次にモード $\mu_i (i=1, 2, \dots, 9)$ を求める。この中から②の条件を満たすような極大値となる μ_1, μ_2, μ_3 の3つのモードを残しこれを順に ν_1, ν_2, ν_3 と置く ($f(\nu_1) > f(\nu_2) > f(\nu_3)$)。ここで改めて ν_1, ν_2, ν_3 を除く残りのすべての個体をこの3つの極値の最も近いもの、つまり $\min_j \{d(\mathbf{x}, \nu_j)\}$ を満たす \tilde{S}_j に \mathbf{x} をわりふる (こうして3個のクラスターがえられる)。

この方法はきわめて発見的な方法で原理が簡単でありしかも計算が容易であるがいくつかの問題もある。たとえば、

- i) 個体の近傍のとり方、つまり α の定め方をどうするか。

図 4
モード探索法の模式図



- ii) $f(x_i)$ ができるだけ滑らかに変化することが望ましいがかなり大量のデータを使ってもこれを満たすことは難しい。
- iii) ii) と関連して、 $f(x_i)$ に同じ値のものがあるべくないほうがよいが実験するとこれが多く現われる (Gitman らはこれを逃れるいわゆる “tie-breaking” の方法を工夫してはいる)。

などである。

ところでこの方法は D. Wishart の “モード法” (1969) や Astrahan の “hyperphoneme method” (1970) に大変良く似ている。各個体の近傍を調べて個体密度 (つまり $f(x_i)$ のこと) を求め、この値がある値以上になった個体だけを密度の濃い領域として取り出し、残りの部分にある個体をこの取り出した領域のいずれかに最近隣方式で配置するというのがこれらの方法の要点であるが、上の方法と比較すると明らかに類似性がみられる。また Jarvis らの提案した方法 (1973) もその一部がこれらの方法と似ている。いずれも、上で指摘のような問題を抱えているのが欠点である。

5. ファジイ関係と階層的分類法

いままで紹介してきた方法はどちらかというと計量的なデータに向けた方法で、そこで使われる距離や類似度もメトリックな性質を満たしていることが望ましい。いま個体 \bar{x} と \bar{y} との差異を示す非類似度を考えることにしこれを $d(x_i, x_j)$ または簡単に d_{ij} と書く。このとき個体間の関連は非類似度行列

$$D = (d_{ij}) \quad (i, j = 1, 2, \dots, n) \quad (21)$$

で与えられる。しかしこの d_{ij} は必ずしもメトリックとは限らずむしろ現実にはノン・メトリックである場合

が多い。すなわち反射律、対称律は満たすが推移律 (三角不等式) を満たすとは限らない。ときには対称律の成立が危ぶまれる場合すらおこる。こうしたきわめて柔らかな情報しか与えられない場合でも個体を分類したいという要請が起こる。

このようなとき上の D を一つの “ファジイ関係 (fuzzy relation)” あるいは “ファジイ・グラフ” と考え d_{ij} を個体間の関連性をはかる所属関数とみることは自然である。つまり D は一つの “ファジイ非類似関係 (fuzzy dessemblance relation)” である。この D に対して次の演算を用意する。

$$\begin{cases} D^{(0)} \equiv D & (n \times n \text{ 次行列}) \\ D^{(l+1)} = D^{(l)} \otimes D^{(0)} & (l = 0, 1, 2, \dots) \end{cases} \quad (22)$$

ここで \otimes は、

$$c_{ij} = \min_k \{ \max \{ a_{ik}, b_{jk} \} \} \quad (23)$$

という演算を行うことを表わす。また a_{ik}, b_{jk}, c_{ij} はそれぞれ行列 $D^{(0)}, D^{(l)}, D^{(l+1)}$ の要素である。こうした演算 (これをファジイ min-max 合成という) を順次繰り返す。一般に l は高々 $l \leq n$ であり多くの場合ある $l' (< n)$ で $D^{(l')} = D^{(l'+1)} = D^{(l'+2)} = \dots$ となる。つまりベキ等関係に到達することが知られている。この性質は D が $\{0, 1\}$ 要素だけからなるブール行列 (あるいは通常のグラフ) の場合に成り立つ性質の拡張である。こうしたファジイ合成で得た行列について、さらに演算 \wedge を適用して

$$\tilde{D} = D \wedge D^{(1)} \wedge D^{(2)} \wedge \dots \wedge D^{(l)} \quad (l \leq n) \quad (24)$$

を作る。ここで \wedge は行列の要素単位に minimum を作用させる演算を示している (ファジイ演算における共通 (intersection) である)。こうして求めた関係 \tilde{D} を “ファジイ (min-max) 閉包” という。 \tilde{D} の (i, j) 要素を

$\tilde{\delta}_{ij}$ で示すと演算の性質から明らかに $\tilde{\delta}_{ij} \leq d_{ij}$ が成り立っている。この \tilde{D} は Zadeh のいう“ファジイ同値関係”であるが、その要素 $\tilde{\delta}_{ij}$ は次の推移律を満たしている。

$$\tilde{\delta}_{ij} \leq \min_k \{ \max(\tilde{\delta}_{ik}, \tilde{\delta}_{kj}) \} \quad (\forall i, j, k \in E) \quad (25)$$

この \tilde{D} は1つの階層構造を生成するが(つまり $\tilde{\delta}_{ij}$ の値の小さい要素から順次連結すると系統樹が作られる)。その要素 $\tilde{\delta}_{ij}$ は実は“ultrametric”の性質を持つことがわかる。ultrametric とは非類似度 δ_{ij} について反射律 ($\delta_{ij}=0$), 対称律 ($\delta_{ij}=\delta_{ji}$) の他に推移律として個体 i, j, k の間に、

$$\delta_{ij} \leq \max\{\delta_{ik}, \delta_{kj}\} \quad (\forall k \in E) \quad (26)$$

が成り立つような場合をいう。これは階層的分類法のきわめて重要な性質であるが上の $\tilde{\delta}_{ij}$ が(26)を満たし、しかも(25), (26)を比べるとこの両者が同じ内容を示していることがわかる。

そして、こうした ultrametric の性質を備える数多くの分類手法のうち single linkage 法 (nearest neighbor, minimum method ともいう) によって生成される系統樹の上に現われる ultrametric は \tilde{D} と一致することがわかる(証明は省くが確かにこれが成立する)。

こうした関係を本誌 No. 181 (1978) に掲載の例を用いて眺めることにする。

【数値例】

つぎの行列 D はハワイにおける相異なる8つの人種間の結婚の割合を示す数で、値が大きいほどその人種間の組み合わせに対する結婚の割合は低くなるように、つまり差異を表わすように尺度化してある。

$$D = \begin{matrix} & O_1 & O_2 & O_3 & O_4 & O_5 & O_6 & O_7 & O_8 \\ \begin{matrix} O_1 \\ O_2 \\ O_3 \\ O_4 \\ O_5 \\ O_6 \\ O_7 \\ O_8 \end{matrix} & \begin{bmatrix} 0 & 6 & 21 & 30 & 18 & 27 & 33 & 32 \\ & 0 & 12 & 21 & 14 & 16 & 23 & 23 \\ & & 0 & 20 & 22 & 24 & 24 & 20 \\ & & & 0 & 19 & 37 & 41 & 37 \\ & & & & 0 & 30 & 30 & 28 \\ & & & & & 0 & 24 & 21 \\ & & & & & & 0 & 20 \\ & & & & & & & 0 \end{bmatrix} \end{matrix}$$

ここで8つの人種はそれぞれつぎの通りである。

- O_1 : Hawaiian (ハワイ人)
- O_2 : Part-Hawaiian (ハワイ系混血)
- O_3 : Caucasian (白人)
- O_4 : Puerto Rican (プエルト・リコ人)
- O_5 : Filipino (フィリピン人)
- O_6 : Chinese (中国人)
- O_7 : Japanese (日本人)
- O_8 : Korean (韓国人)

さて式(22)により $D^{(1)}, D^{(2)}, D^{(3)}, \dots$ と順次計算すると $D^{(3)}=D^{(4)}=\dots$ となり $l=3$ でベキ等に到達する。さらに \tilde{D} は、

$$\tilde{D} = (\tilde{\delta}_{ij}) = \begin{matrix} & O_1 & O_2 & O_3 & O_4 & O_5 & O_6 & O_7 & O_8 \\ \begin{matrix} O_1 \\ O_2 \\ O_3 \\ O_4 \\ O_5 \\ O_6 \\ O_7 \\ O_8 \end{matrix} & \begin{bmatrix} 0 & 6 & 12 & 19 & 14 & 16 & 20 & 20 \\ & 0 & 12 & 19 & 14 & 16 & 20 & 20 \\ & & 0 & 19 & 14 & 16 & 20 & 20 \\ & & & 0 & 19 & 19 & 20 & 20 \\ & & & & 0 & 16 & 20 & 20 \\ & & & & & 0 & 20 & 20 \\ & & & & & & 0 & 20 \\ & & & & & & & 0 \end{bmatrix} \end{matrix}$$

となる。ここで確かに $\tilde{\delta}_{ij} \leq d_{ij}$ となっている。またこれとは別に single linkage 法により系統樹を作りこの上から個体間の結合時の非類似度を順に読みとり1つの行列を作る(作り方は[26], 本誌 No. 181などを参照)。この結果が次の行列 D^* である。

$$D^* = \begin{matrix} & O_1 & O_2 & O_3 & O_4 & O_5 & O_6 & O_7 & O_8 \\ \begin{matrix} O_1 \\ O_2 \\ O_3 \\ O_4 \\ O_5 \\ O_6 \\ O_7 \\ O_8 \end{matrix} & \begin{bmatrix} 0 & 6 & 12 & 19 & 14 & 16 & 20 & 20 \\ & 0 & 12 & 19 & 14 & 16 & 20 & 20 \\ & & 0 & 19 & 14 & 16 & 20 & 20 \\ & & & 0 & 19 & 19 & 20 & 20 \\ & & & & 0 & 16 & 20 & 20 \\ & & & & & 0 & 20 & 20 \\ & & & & & & 0 & 20 \\ & & & & & & & 0 \end{bmatrix} \end{matrix}$$

こうして確かに D^* と \tilde{D} の一致が確かめられる。またいずれも ultrametric の性質を満たすことがわかる。

また次の性質をみることができる。いま \tilde{D} (または D^*) と D の要素をくらべて $\tilde{\delta}_{ij}=d_{ij}$ となる要素だけを取りだしてみよう(上の \tilde{D}, D^* で太字の要素がそれである)。そしてそのときの個体番号 (i, j) とその非類似度 δ_{ij} の値を小さいものから順に並べかえる。すると、次の表4がえられる。これをもとに各個体を順に連結すると図5が現われる。これはいわゆる“Minimal Spanning Tree (MST)”であり、このことからファジイ閉包 \tilde{D} は MST を生成する方法であることがわかる。

表 4

順位	i	j	結合距離 d_{ij}
1	1	2	6
2	2	3	12
3	2	5	14
4	2	6	16
5	4	5	19
6	3	8	20
7	7	8	20

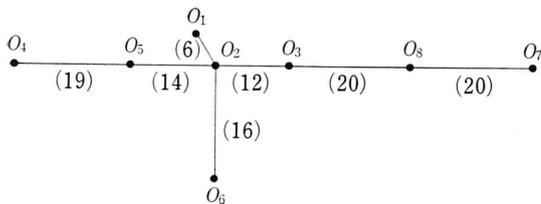


図5 \tilde{D} または D^* から生成した MST

このように階層的分類法とファジイ関係は密接な関係にある。上に述べた性質の他にさらに多くの面白い結果を得ることができるが、とくに同一データから得た複数個の系統樹間の比較、系統樹を切断してえられる分割集合(クラスター)の評価などの問題にこのファジイ関係の概念が活躍するのである(余談であるが、ファジイ関係も含めてファジイ理論全体を広く知りたい向きは Kaufmann[13]を見るとよい。この本には実に懇切丁寧な解説がなされている)。

6. むすび

ファジイ・クラスタリングの一端を垣間見たわけであるが歴史が浅いこともあって甲論乙駁の感がないでもない。結局どれがよいとか悪いとかいうよりも、どんな問題を相手にして分析を進めねばならないかという観点から、その問題に即応した方法を取捨選択することであろう。そしてクラスター分析におけるファジイ理論の役割りは、手法や算法の構築よりもむしろ手法間の比較や評価といった、従来クラスター分析の研究でアキレス腱であった部分に切り込む有力な道具として期待してよいのではなかろうか。

参考文献

[1] Astrahan, M. M.; Speech Analysis by Clustering, or the Hyperphoneme Method, Stanford Artificial Intelligence Project Memo AIM-124, 1970.
 [2] Backer, E.; On the Application of Fuzzy Set Theory to Cluster Analysis, in *Pattern Recognition Theory and Application*, (eds.) Fu, K. S. and Winston, A. B., NATO Advanced Study Institutes Series E: No. 22, 1977.
 [3] Bellman, R., Kalaba, R. and Zadeh, L.; Abstraction and Pattern Classification, *J. Math. Anal. Appl.* **13**, 1966.
 [4] Bezdek, J. C.; Numerical Taxonomy with Fuzzy Sets, *Journal of Mathematical Biology*, **1**, 57-71, 1974.
 [5] Bezdek, J. C.; Cluster Validity with Fuzzy Sets, *Journal of Cybernetics*, **3**, 3, 58-73, 1974.
 [6] Bezdek, J. C.; A Physical Interpretation of Fuzzy ISODATA, *IEEE Transactions on Systems, Man, and Cybernetics*, May, 1976.
 [7] Bezdek, J. C.; Prototype Classification and Feature Selection with Fuzzy Sets, *IEEE Transaction on*

Systems, Man, and Cybernetics, SMC-7, 2, 1977.
 [8] Dunn, J. C.; A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters, *Journal of Cybernetics*, **3**, 32-57, 1974.
 [9] Dunn, J. C.; A Graph Theoretic Analysis of Pattern Classification via Tamura's Fuzzy Relation, *IEEE Transaction on Systems, Man, and Cybernetics*, May, 1974.
 [10] Gitman, I. and Levine, M. D.; An Algorithm for Detecting Unimodal Fuzzy Sets and Its Application as a Clustering Techniques, *IEEE Transaction on Computer*, C-19, 7, 1970.
 [11] Halpern, J.; Set Adjacency Measures in Fuzzy Graphs, *Journal of Cybernetics*, **5**, 4, 1975.
 [12] Jarvis, R. A. and Patrick, E. A.; Clustering Using a Similarity Measure Based on Shared Near Neighbors, *IEEE Transaction on Computers*, C-22, 11, 1973.
 [13] Kaufmann, A.; *Introduction à la théorie des sous-ensemble flous*, Masson.
 Tome I.—*Elément théoriques de base* (1973).
 Tome II.—*Applications à la linguistique, à la logique et à la sémantique* (1975).
 Tome III.—*Applications à la classification et à la reconnaissance des formes, aux automates et aux systèmes, aux choix des critères* (1975).
 Tome IV.—*Complément et nouvelles applications* (1977).
 [14] Lee, E. T.; Shape-Oriented Chromosome Classification, *IEEE Transaction on Systems, Man, and Cybernetics*, 1975.
 [15] Lee, E. T.; An Application of Fuzzy Sets to the Classification of Geometric Figures and Chromosome Images, *Information Sciences*, **10**, 1976.
 [16] Lee, E. T.; Shape-Oriented Storage and Retrieval of Geometrical Figures and Chromosome Images, *Information Processing and Management*, **12**, 1976.
 [17] Negoită, C. V. and Ralescu, D. A.; *Applications of Fuzzy Sets to Systems Analysis*, Birkhäuser, 1975.
 [18] Ragade, R. K.; Fuzzy Interpretive Structural Modeling, *Journal of Cybernetics*, **6**, 1976.
 [19] Rosenfeld, A.; Fuzzy Graphs, Tech. Rep. (NTIS, PB-237469), 1974.
 [20] Ruspini, E. H.; A New Approach to Clustering, *Information and Control*, **15**, 1969.
 [21] Ruspini, E. H.; Numerical Methods for Fuzzy Clustering, *Information Sciences*, **2**, 1970.
 [22] Ruspini, E. H.; New Experimental Results in Fuzzy Clustering, *Information Sciences*, **6**, 1973.
 [23] Wishart, D.; Mode Analysis: A Generalization of Nearest Neighbour which Reduces Chaining Effects, in *Numerical Taxonomy*, (ed.) Cole, A. J., Academic Press, 1969.
 [24] Zadeh, L., Fu, K. S., Tanaka, K. and Shimura, M.; *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*, Academic Press, 1975.
 [25] Zadeh, L.; Similarity Relations and Fuzzy Orderings, *Information Sciences*, **3**, 177-200, 1971.
 [26] 大隅昇; クラスタ分析, 現代数学, **10**, 9, 1977.
 [27] 大隅昇; クラスタ分析—SL法をめぐる, 数理科学, No. 181, JULY, 1978.
 [28] 大隅昇; クラスタ分析はどう使われるか, 数理科学, No. 190, 1979.

(おおすみ・のぼる, 統計数理研究所)