第4章　ファジィデータに対するL_1距離に基づくファジィクラスタリング

4.1　はじめに

最近，ファジィクラスタリング手法の中で代表的なファジィc-平均法において，ファジィデータを扱う研究がされている [28, 39, 49]．これらの研究では目的関数にユークリッド距離の二乗が含まれているが，ファジィc-平均法の目的関数に対する厳密な交互最適化をおこなっていない．

ファジィc-平均法では通常，データとクラスター中心との間の距離は，各成分の差の二乗をすべての成分に関して加え合わせることにより求められる．いいかえればユークリッド距離の二乗によって測定される．個体をユークリッド空間の１点としてとらえることは，座標軸の回転を許容することであり，その便利さのために最も頻繁に用いられている．しかしながら，多変量の統計ではつねにユークリッド空間が用いられるとは限らず，マンハッタン空間と呼ばれるL_1空間もしばしば用いられている．

ところで，個体１つ１つに不確定性があると考える場合，ここで用いる仮定のようにファジィ数の直積で表現されるファジィデータが最も自然である．ところが，このようなファジィデータでは，座標軸を回転すると，不確定性を含むデータはもはやファジィ数の直積ではなくなる．したがって，回転に対して個体データの性質が不変ではないので，回転を許容しない空間を考えるほうがむしろ適切である．すなわち，L_1空間は不確定性を含むデータを扱う場合に，その仮定の自然さ，適切さにおいてユークリッド空間に劣ることはない．

L_1距離を用いた不確定性を含まないデータに対するファジィc-平均法もこれまで研究されており [8, 34]，目的関数の交互最適化におけるクラスター中心に関する最

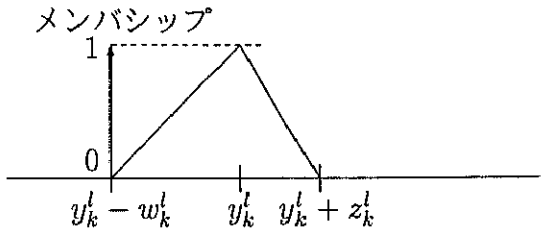

図 4.1: 三角ファジィ数 M_k^l の例

適化ステップの効率的なアルゴリズムが，宮本ら [43] によって提案されている．

クラスタリングは，データ間の類似度の大きさ (距離の近さ) をもとに分類を行っている．よって，ファジィデータをクラスタリングするためには，ファジィデータ間の距離を定めなければならない．ファジィ c-平均法では，クラスター中心とファジィデータとの間の距離を定めることになる．本論文では目的関数に含まれるクラスター中心とファジィデータ間の L_1 距離を，最長距離法，最短距離法を用いて定義する．目的関数を交互最適化で最適化する際に，クラスター中心に関する最適解は，従来のファジィ c-平均法と同様に求めることができない．そこで，新たにクラスター中心計算アルゴリズムを開発し，目的関数の厳密な交互最適化を実現する．また，実際の不確定性を含むデータを本手法で分類した結果と，同じデータが不確定性を含まないと仮定した場合の分類結果を比較する．

4.2　ファジィデータ

これまでにファジィデータをクラスタリングする手法として，各々の個体に超楕円体の不確定性を仮定した佐藤らの手法 [53, 54] が提案されている．しかしながら，一般に超楕円体を用いた解析は複雑で多くのデータを厳密に扱うには適さない．ここでは，単純なファジィ数の直積で表現できるようなファジィデータを取り扱う．

分類されるファジィデータを

$$M = \{M_1, M_2, \cdots, M_n\} \tag{4.1}$$

と表し，個体 M_k は，ファジィ数 M_k^l のデカルト積で表されるとする．

$$M_k = M_k^1 \times M_k^2 \times \cdots \times M_k^p. \tag{4.2}$$

ファジィデータの第 l 成分であるファジィ数 M_k^l が三角ファジィ数である場合，第 l 成分の軸を x とすると，メンバシップ関数 $\mu_{M_k^l}$ を次のように表現できる．

$$\mu_{M_k^l} = \begin{cases} 0 & (x \leq y_k^l - w_k^l). \\ (x - (y_k^l - w_k^l))/w_k^l & (y_k^l - w_k^l \leq x \leq y_k^l). \\ ((y_k^l + z_k^l) - x)/z_k^l & (y_k^l \leq x \leq y_k^l + z_k^l). \\ 0 & (y_k^l + z_k^l \leq x). \end{cases} \tag{4.3}$$

図 4.1 は三角ファジィ数の例を示している．この三角ファジィ数は，個体の l 成分が不確定性を含み，$[y_k^l - w_k^l, y_k^l + z_k^l]$ の範囲内に存在し，y_k^l に存在する可能性が高く $y_k^l - w_k^l$ もしくは $y_k^l + z_k^l$ に存在する可能性は低いことを表している．

このようなファジィデータを L_1 距離を用いたファジィ c-平均法でクラスタリングすることを試みる．

4.3　ファジィデータに対する目的関数

既存の L_1 距離に基づく標準的なファジィ c-平均法とエントロピー正則化を用いたファジィ c-平均法の目的関数は，

$$J_m(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m \parallel x_k - v_i \parallel_1$$

$$J^\lambda(U,V) = \sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik} \parallel x_k - v_i \parallel_1 + \lambda^{-1}\sum_{i=1}^{c}\sum_{k=1}^{n} u_{ik}\log u_{ik}$$

であり，クラスタリングすべき個体 x_k は不確定性を含まないと仮定している．それぞれの目的関数では，L_1 距離 $\parallel x_k - v_i \parallel_1$ が含まれている．不確定性を含まない個体 x_k のかわりにファジィ数の直積で表される個体 M_k を扱うには，L_1 距離を新たに定義する必要がある．

クラスター中心は $V = \{v_1, v_2, \cdots, v_c\}$ であり，各々の中心は (2.16) と同様，

$$v_i = (v_i^1, v_i^2, \cdots, v_i^p)^T \tag{4.4}$$

とする．クラスター中心 v_i は，不確定性を含まないと仮定する．メンバシップ行列 $U = (u_{ik})$ の制約条件は (2.17) と同じである．

$$\mathcal{M} = \{(u_{ik}) \mid u_{ik} \in [0,1], \sum_{i=1}^{c} u_{ik} = 1, k = 1, 2, \cdots, n\}. \tag{4.5}$$

標準的なファジィc-平均法とエントロピー法の目的関数は，

$$J_m(U,V)=\sum_{i=1}^{c}\sum_{k=1}^{n}(u_{ik})^m D_{ik} \tag{4.6}$$

$$J^{\lambda}(U,V)=\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}D_{ik}+\lambda^{-1}\sum_{i=1}^{c}\sum_{k=1}^{n}u_{ik}\log u_{ik} \tag{4.7}$$

のように新たに定義する．D_{ik} は不確定性を含むデータ M_k とクラスター中心 v_i との L_1 距離を示しており，各々の成分の和で表現される．

$$D_{ik}=\parallel M_k-v_i\parallel_1=\sum_{l=1}^{p}\parallel M_k^l-v_i^l\parallel_1 . \tag{4.8}$$

この距離を最長距離と最短距離を用いて定義する．

まず，ファジィデータの α カットはファジィ数の α カットの直積，つまり区間の直積で表現される．

$$(M_k)_\alpha=(M_k^1)_\alpha\times(M_k^2)_\alpha\times\cdots\times(M_k^p)_\alpha. \tag{4.9}$$

ここで，ファジィデータの α カット $(M_k)_\alpha$ と中心 v_i との間の L_1 距離 $\parallel(M_k)_\alpha-v_i\parallel_1$ を求める．L_1 距離は，各々の成分 $\parallel(M_k^l)_\alpha-v_i^l\parallel_1$ を加え合わせたものである．

$$\parallel(M_k)_\alpha-v_i\parallel_1=\sum_{l=1}^{p}\parallel(M_k^l)_\alpha-v_i^l\parallel_1 . \tag{4.10}$$

ここで $(M_k^l)_\alpha$ は区間であることに注意する．L_1 距離の1つの成分 $\parallel(M_k^l)_\alpha-v_i^l\parallel_1$ を最長距離，最短距離を用いて求める．集合 K と集合 L の間の最長距離，最短距離 $d(K,L)$ は

最長距離法 (Farthest Neighbor)

$$d(K,L)=\max\{d(x,y):\ for\ all\ x\in K,y\in L\}$$

最短距離法 (Nearest Neighbor)

$$d(K,L)=\min\{d(x,y):\ for\ all\ x\in K,y\in L\}$$

である．これらの距離を用いて $\| (M_k^l)_\alpha - v_i^l \|_1$ を求める．区間 $(M_k^l)_\alpha=[f_{k1}^l, f_{k2}^l]$ とおくと，最長距離法を用いた場合は，

$$\| (M_k^l)_\alpha - v_i^l \|_1 = \max\{|f_{k1}^l - v_i^l|, |f_{k2}^l - v_i^l|\}$$

最短距離法を用いた場合は，

$$\| (M_k^l)_\alpha - v_i^l \|_1 = \begin{cases} 0, & (f_{k1}^l \le v_i^l \le f_{k2}^l) \\ \min\{|f_{k1}^l - v_i^l|, |f_{k2}^l - v_i^l|\}, & (otherwise) \end{cases}$$

のように求められる．このようにして (4.10) 式の $\| (M_k)_\alpha - v_i \|_1$ を求めることができる．目的関数 (4.6), (4.7) に含まれる L_1 距離 (4.8) はファジィデータの α カットとクラスター中心との L_1 距離 $\| (M_k)_\alpha - v_i \|_1$ を用いて，

$$\begin{aligned} D_{ik} &= \int_0^1 \| (M_k)_\alpha - v_i \|_1 \, d\alpha \\ &= \sum_{l=1}^{p} \int_0^1 \| (M_k^l)_\alpha - v_i^l \|_1 \, d\alpha \end{aligned} \tag{4.11}$$

と定義する．つまり，ファジィデータとクラスター中心との間の距離は，ファジィデータの α カットとクラスター中心との間の距離の α に関する積分で定義される．そして，ファジィデータの α カット (区間の直積) とクラスター中心との間の距離は最長距離もしくは最短距離で求める．これで最長距離，最短距離を用いて，ファジィデータ M_k に対するファジィ c-平均法の目的関数 (4.6)，(4.7) が定義された．

データが三角ファジィ数 (4.3) の直積で表現されると仮定した場合，(4.11) における積分の値は次のように計算される．

最長距離法を用い，$z_k^l \le w_k^l$ の場合には，

$$\int_0^1 \| (M_k^l)_\alpha - v_i^l \|_1 = \begin{cases} y_k^l - v_i^l + \dfrac{1}{2} z_k^l, (v_i^l \le \dfrac{2y_k^l - w_k^l + z_k^l}{2}). \\ 2\dfrac{(v_i^l - y_k^l)^2}{w_k^l - z_k^l} + v_i^l - y_k^l + \dfrac{1}{2} w_k^l, \\ \quad (\dfrac{2y_k^l - w_k^l + z_k^l}{2} \le v_i^l \le y_k^l, w_k^l \ne z_k^l). \\ v_i^l - y_k^l + \dfrac{1}{2} w_k^l, (y_k^l \le v_i^l). \end{cases}$$

最長距離法を用い，$w_k^l \leq z_k^l$ の場合には，

$$\int_0^1 \| (M_k^l)_\alpha - v_i^l \|_1 = \begin{cases} y_k^l - v_i^l + \frac{1}{2} z_k^l, (v_i^l \leq y_k^l). \\ 2\frac{(y_k^l - v_i^l)^2}{z_k^l - w_k^l} + y_k^l - v_i^l + \frac{1}{2} z_k^l, \\ \quad (y_k^l \leq v_i^l \leq \frac{2y_k^l - w_k^l + z_k^l}{2}, w_k^l \neq z_k^l). \\ v_i^l - y_k^l + \frac{1}{2} w_k^l, (\frac{2y_k^l - w_k^l + z_k^l}{2} \leq v_i^l). \end{cases}$$

最短距離法を用いた場合には，

$$\int_0^1 \| (M_k^l)_\alpha - v_i^l \|_1 = \begin{cases} (y_k^l - w_k^l) - v_i^l + \frac{w_k^l}{2}, (v_i^l \leq y_k^l - w_k^l). \\ \frac{(y_k^l - v_i^l)^2}{2w_k^l}, (y_k^l - w_k^l \leq v_i^l \leq y_k^l, w_k^l \neq 0). \\ \frac{(v_i^l - y_k^l)^2}{2z_k^l}, (y_k^l \leq v_i^l \leq y_k^l + z_k^l, z_k^l \neq 0). \\ v_i^l - (y_k^l + z_k^l) + \frac{z_k^l}{2}, (y_k^l + z_k^l \leq v_i^l). \end{cases}$$

4.4　クラスター中心に対する最適化

ファジィc-平均法の解は目的関数 (4.6), (4.7) を最小化することにより求められる．これらの目的関数は2つの変数 U と V を含んでおり，アルゴリズム **FCM** を用いて最小化を行う．ステップ **FCM2** において，V を固定し，$J(U, \bar{V})$ の最小化を行うが，ファジィデータを扱う場合も既存のファジィc-平均法 (2.21),(2.22) 同様に最適解 U を求めることができる．標準的なファジィc-平均法のステップ **FCM2** における解は，

$$\bar{u}_{ik} = \left[\sum_{j=1}^{c} \left(\frac{D_{ik}}{D_{jk}} \right)^{\frac{1}{m-1}} \right]^{-1} \tag{4.12}$$

エントロピー正則化を用いたファジィc-平均法のステップ **FCM2** における解は，

$$\bar{u}_{ik} = \frac{e^{-\lambda D_{ik}}}{\sum_{j=1}^{c} e^{-\lambda D_{jk}}} \tag{4.13}$$

のように表される．ステップ **FCM3** における最適解を求めるには新たな解法が必要となる．本論文では厳密な最適解を求めるアルゴリズムを提案する．まず，メン

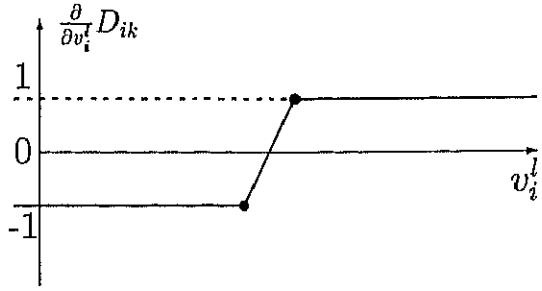

図 4.2: ある k についての $\frac{\partial}{\partial v_i^l} D_{ik}$ の例示 (最長距離法)

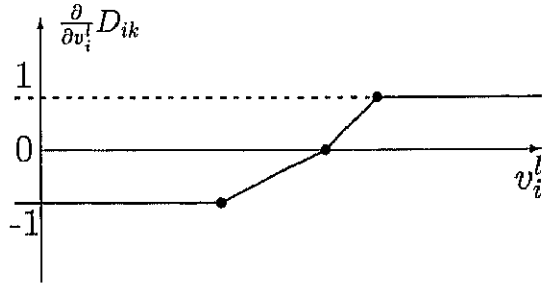

図 4.3: ある k についての $\frac{\partial}{\partial v_i^l} D_{ik}$ の例示 (最短距離法)

バシップ U を固定した場合の標準的なファジィ c-平均法の目的関数 (4.6) の形状を考えてみよう．目的関数のクラスター中心 V に関する偏導関数は

$$\frac{\partial}{\partial v_i} J_m(U,V) = \sum_{k=1}^{n} (u_{ik})^m \frac{\partial}{\partial v_i} D_{ik} \tag{4.14}$$

と表される．目的関数 (4.7) に対する偏導関数は (4.14) 式において $m = 1$ とおけばよい．したがって，以下 $J_m(U,V)$ の偏導関数のみを扱う．

(4.14) 式における $\frac{\partial}{\partial v_i} D_{ik}$ の第 l 成分 $\frac{\partial}{\partial v_i^l} D_{ik}$ は図 4.2，図 4.3 のように区分的に線形であり，次のようになる．

最長距離を用いて $z_k^l \leq w_k^l$ のとき，

$$\frac{\partial}{\partial v_i^l} D_{ik} = \begin{cases} -1, (v_i^l \leq \dfrac{2y_k^l - w_k^l + z_k^l}{2}). \\ 4\dfrac{v_i^l - y_k^l}{w_k^l - z_k^l} + 1, (\dfrac{2y_k^l - w_k^l + z_k^l}{2} \leq v_i^l \leq y_k^l, w_k^l \neq z_k^l). \\ 1, (y_k^l \leq v_i^l). \end{cases}$$

最長距離を用いて $w_k^l \leq z_k^l$ のとき，

$$\frac{\partial}{\partial v_i^l} D_{ik} = \begin{cases} -1, (v_i^l \leq y_k^l). \\ 4\dfrac{v_i^l - y_k^l}{z_k^l - w_k^l} - 1, (y_k^l \leq v_i^l \leq \dfrac{2y_k^l - w_k^l + z_k^l}{2}, w_k^l \neq z_k^l). \\ 1, (\dfrac{2y_k^l - w_k^l + z_k^l}{2} \leq v_i^l). \end{cases}$$

最短距離を用いて

$$\frac{\partial}{\partial v_i^l} D_{ik} = \begin{cases} -1, (v_i^l \leq y_k^l - w_k^l). \\ \dfrac{v_i^l - y_k^l}{w_k^l}, (y_k^l - w_k^l \leq v_i^l \leq y_k^l, w_k^l \neq 0). \\ \dfrac{v_i^l - y_k^l}{z_k^l}, (y_k^l \leq v_i^l \leq y_k^l + z_k^l, z_k^l \neq 0). \\ 1, (y_k^l + z_k^l \leq v_i^l). \end{cases}$$

これらの偏導関数 $\frac{\partial}{\partial v_i^l} D_{ik}$ は図 4.2，図 4.3 のような区分線形な単調増加関数である．このことから (4.14) 式の第 l 成分

$$\frac{\partial}{\partial v_i^l} J_m(U, V) = \sum_{k=1}^{n} (u_{ik})^m \frac{\partial}{\partial v_i^l} D_{ik} \tag{4.15}$$

は区分線形な単調増加関数であり，$J_m(U, V)$ は v_i^l に関して凸関数である．そこで偏導関数 (4.15) 式が 0 となるような v_i^l の値を線形探索するようなアルゴリズムを用いて目的関数 $J_m(U, V)$ の最適解を求める．

4.5　最適化アルゴリズム

ファジィデータの第 l 成分は (4.3) 式のように表現されるとする．目的関数の v_i^l に関する偏導関数 (4.15) 式は区分線形な単調増加関数である．(4.15) 式の区分点を探索し，クラスター中心 v_i の最適解を出力するアルゴリズムを以下に示す．

最長距離法を用いた場合は，$\dfrac{2y_k^l - w_k^l + z_k^l}{2} (k = 1, \cdots, n)$, $y_k^l (k = 1, \cdots, n)$ の $2n$ 個の値を小さい順に並べ替え，$X_1^l, X_2^l, \cdots, X_{2n}^l$:

$$X_1^l \leq \cdots \leq X_j^l \leq \cdots \leq X_{2n}^l$$

とおく．

Algorithm (Searching v_i^l in Farthest Neighbor)

begin
$U := 0$;
$S := -\sum_{k=1}^{n}(u_{ik})^m$;
$j := 1$;
while (1) **do begin**
if $(z_k^l < w_k^l)$ **then**
if $X_j^l = \dfrac{2y_k^l - w_k^l + z_k^l}{2}$
then $U := U + \dfrac{4(u_{ik})^m}{w_k^l - z_k^l}$;
if $X_j^l = y_k^l$
then $U := U - \dfrac{4(u_{ik})^m}{w_k^l - z_k^l}$;
if $(w_k^l = z_k^l)$**then**
if $X_j^l = y_k^l$
then $S := S + (u_{ik})^m$;
if $X_j^l = \dfrac{2y_k^l - w_k^l + z_k^l}{2}$
then $S := S + (u_{ik})^m$;
if $S > 0$ **then** $\bar{v}_i^l := X_j^l$; **break**;
if $(w_k^l < z_k^l)$**then**
if $X_j^l = y_k^l$
then $U := U + \dfrac{4(u_{ik})^m}{z_k^l - w_k^l}$;
if $X_j^l = \dfrac{2y_k^l - w_k^l + z_k^l}{2}$
then $U := U - \dfrac{4(u_{ik})^m}{z_k^l - w_k^l}$;
$S := S + U(X_{j+1}^l - X_j^l)$;
$j := j + 1$;
if $S > 0$ **then** $\bar{v}_i^l := X_j^l - \frac{S}{U}$; **break**;
end;
output $\bar{v}_i^l$ as the l-th coordinate of the
cluster center v_i
end.

最短距離法を用いた場合，$y_k^l - w_k^l (k = 1, \cdots, n)$, $y_k^l (k = 1, \cdots, n)$, $y_k^l + z_k^l (k = 1, \cdots, n)$ の $3n$ 個の値を小さい順に並べ替え，$X_1^l, X_2^l, \cdots, X_{3n}^l$:

$$X_1^l \leq \cdots \leq X_j^l \leq \cdots \leq X_{3n}^l$$

のようにおく．

Algorithm (Searching v_i^l in Nearest Neighbor)

begin
  $U := 0$;
  $S := -\sum_{k=1}^{n}(u_{ik})^m$;
  $j := 1$;
  while (1) **do begin**
    if $X_j^l = y_k^l - w_k^l$ **then**
      if $w_k^l = 0$ **then**
        $S := S + (u_{ik})^m$;
        if $S > 0$ **then** $\bar{v}_i^l := X_j^l$; **break**;
      else then
        $U := U + \dfrac{(u_{ik})^m}{w_k^l}$;
    if $X_j^l = y_k^l$ **then**
      if $w_k^l = 0$ **and** $z_k^l \neq 0$ **then**
        $U := U + \dfrac{(u_{ik})^m}{z_k^l}$;
      if $w_k^l = 0$ **and** $z_k^l = 0$ **then**
        $U := U$;
      if $w_k^l \neq 0$ **and** $z_k^l = 0$ **then**
        $U := U - \dfrac{(u_{ik})^m}{w_k^l}$;
      if $w_k^l \neq 0$ **and** $z_k^l \neq 0$ **then**
        $U := U + (-\dfrac{1}{w_k^l} + \dfrac{1}{z_k^l})(u_{ik})^m$;
    if $X_j^l = y_k^l + z_k^l$ **then**
      if $z_k^l = 0$ **then**
        $S := S + (u_{ik})^m$;
        if $S > 0$ **then** $\bar{v}_i^l := X_j^l$; **break**;
      else then
        $U := U - \dfrac{(u_{ik})^m}{z_k^l}$;
    $S := S + U(X_{j+1}^l - X_j^l)$;
    $j := j + 1$;
    if $S > 0$ **then** $\bar{v}_i^l := X_j^l - \frac{S}{U}$; **break**;
  end;
  output $\bar{v}_i^l$ as the l-th coordinate of the
  cluster center v_i
end.

偏導関数 (4.15) 式は区分線形な単調増加関数なのでこれらのアルゴリズムでは並べ替えられた $X_1^l \leq X_2^l \leq X_3^l \leq \cdots$ を順に探索していく．探索中，

$$\sum_{k=1}^{n}(u_{ik})^m \frac{\partial}{\partial v_i^l} \parallel M_k^l - v_i^l \parallel_1$$

において v_i^l に X_j^l を代入した値が正となると探索を終了し，クラスター中心を出力する．上記のアルゴリズムにおける S は，

$$\sum_{k=1}^{n}(u_{ik})^m \frac{\partial}{\partial v_i^l} \parallel M_k^l - v_i^l \parallel_1$$

において v_i^l に X_j^l を代入した値を示しており，U は，

$$\sum_{k=1}^{n}(u_{ik})^m \frac{\partial^2}{\partial v_i^{l2}} \parallel M_k^l - v_i^l \parallel_1$$

における v_i^l に $X_j^l + 0$ を代入した値を示している．

また，これらのアルゴリズムでは，最大で $2n$ 個，もしくは $3n$ 個の区分点を探索するので，ソーティングを除いた計算量は，$O(n)$ である．既存のファジィ c-平均法で，クラスター中心を求めるステップ (2.5)，(2.11) での計算量も $O(n)$ である．

これらのアルゴリズムを用いることによりステップ **FCM3** におけるクラスター中心に関する厳密な最適解が得られ，ファジィデータに対する L_1 距離に基づく目的関数を厳密な交互最適化によって最適化することができる．

ここではファジィ数は三角ファジィ数に限定し，L_1 距離を用いている．非線形なファジィ数を用いた場合には，偏導関数も非線形となり，より複雑な解法が必要となる．また，L_1 距離のかわりにユークリッド距離の二乗を用いた場合もまた，偏導関数は非線形となり，より複雑な解法が必要となる．ここではこれらの解法に触れない．

4.6　数値例とその結果

まず，人工的に作成されたバタフライデータと呼ばれるデータに，ファジィの不確定性を仮定したもの，それから，人物の印象に対する幅をもった回答からなるデータに対し数値実験を行った．それぞれの個体が不確定性を含まないと仮定した場合の従来のファジィ c-平均法の結果と，不確定性を含んだまま，ここで提案した手法でクラスタリングした結果を比較する．

バタフライデータに対する数値実験

バタフライデータと呼ばれるデータ (図4.4 におけるそれぞれの長方形の中心点からなるデータ) がある．この不確定性を含まないデータを，既存のファジィ c-平均法

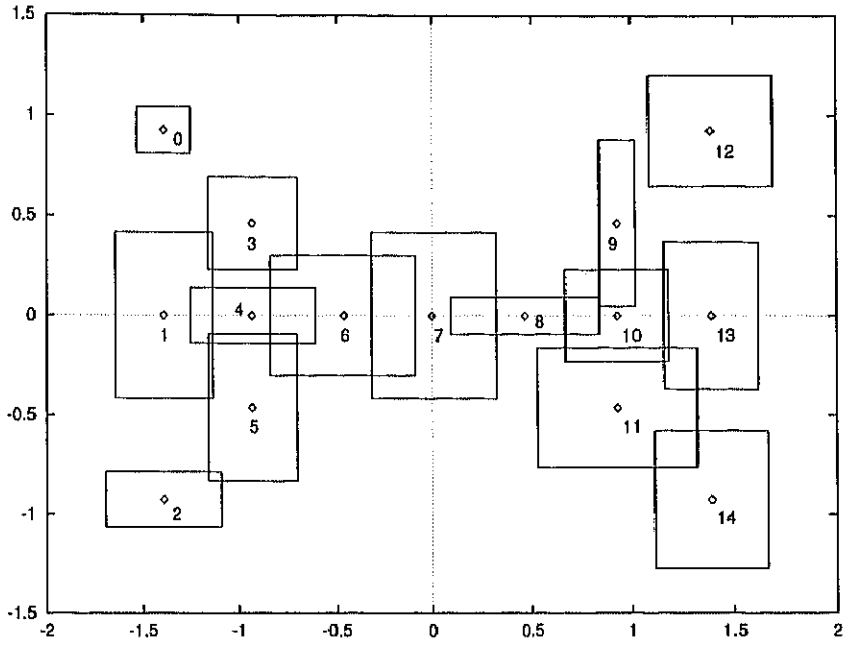

図 4.4: バタフライデータ (不確定性あり)

でクラスタリングを行い，結果を得る．また，同じデータに中心から両側に対称に不確定性を仮定したファジィデータ (図4.4 の長方形を底辺とし，長方形の中心点を頂点とする四角錐) を本手法で 2 つにクラスタリングした結果を示す．

クラスタリング手法は，標準的なファジィc-平均法 (パラメータ $m = 2.0$) と，エントロピー法 (パラメータ $\lambda = 0.7$) を用いた．アルゴリズム FCM における収束判定基準は，各ループでのメンバシップの変化の最大値がある値より小さいことを使った．

$$\max_{i,k} \mid u_{ik} - \bar{u}_{ik} \mid < \varepsilon, \quad \varepsilon = 1.0 \times 10^{-6}. \tag{4.16}$$

図4.5，図4.6 は標準的なファジィc-平均法とエントロピー法を用いて，クラスタリングした結果である．横軸は個体番号で，縦軸は右側のクラスターに対するメンバシップ値を表している．区間データに対する最短距離法，最長距離法を用いた結果はそれぞれ×，＊で表される．また，見やすくするために点線で結んでいる．不確定性を含まないデータに対する結果は＋で表され，実線で結んでいる．

図4.5 にみられるように，標準的なファジィc-平均法では，同じデータでも不確定性を含まない場合と不確定性を仮定した場合はかなり違った結果を示す．また最長距離法を用いたメンバシップと最短距離法を用いたメンバシップは異なっており，不

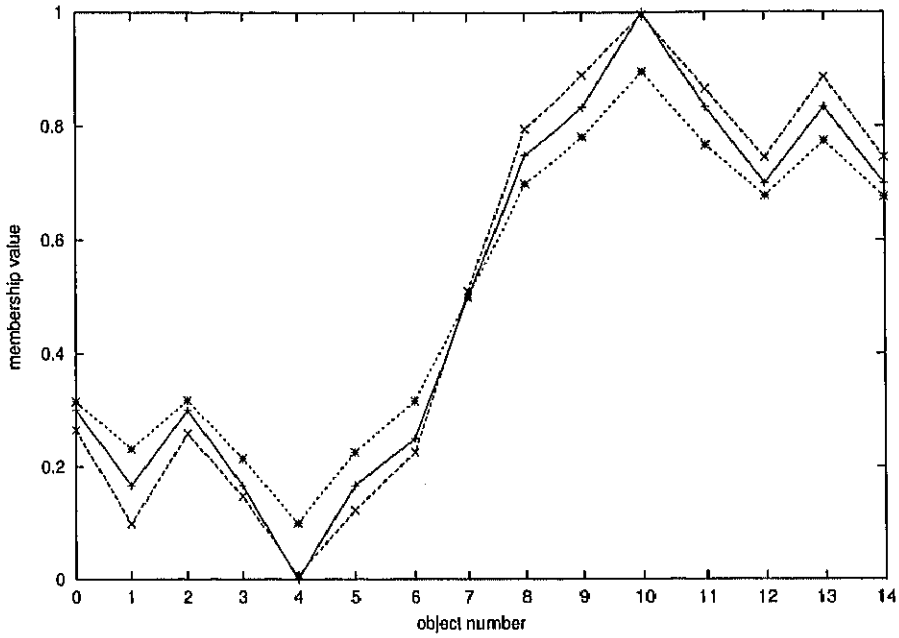

図 4.5: バタフライデータに対するクラスタリング結果 (個体番号 k に対するメンバシップ u_{1k}，標準的なファジィ c-平均法)

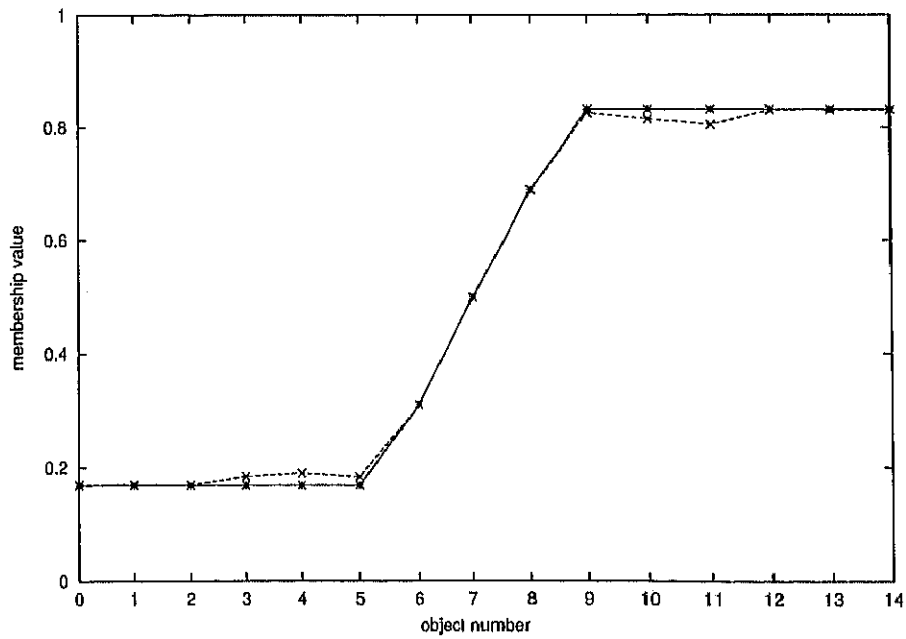

図 4.6: バタフライデータに対するクラスタリング結果 (個体番号 k に対するメンバシップ u_{1k}，エントロピー法)

確定性を含まない場合のメンバシップを両側からはさんでいる．ただし，この不確定性を含まない場合のメンバシップを不確定性を含む場合のメンバシップがはさんでいることについては，このデータに限った場合である．実際，次に述べる実データにおけるクラスタリング結果では，このメンバシップの関係が成り立たない場合がある．また，エントロピー法における結果は，不確定性を仮定しない場合と仮定する場合で，同じようなメンバシップの値が得られた．

標準的なファジィc-平均法を用いた場合に，不確定性を含まないと仮定した場合，不確定性を含むデータに対し最長距離，最短距離を用いた場合で，それぞれメンバシップが異なる．このように同じデータに対して異なる結果が得られた場合には結果が安定しているとはいえず，それらの結果を比較検討する必要がある．また，同じ結果が得られた場合には，結果は安定しているといえる．クラスター中心の位置に関しては，それぞれの場合で差異がほとんどみられなかった．

人物の印象に関する実データに対する数値実験

実際に調査されたデータに対する数値実験の結果を示す．被験者は私立K女子大学およびK女子短期大学学生400名であった．質問紙は，京都市内のK女子大学およびK女子短期大学で行われた俳優M氏の講演会へ出席する意図と，M氏についてのイメージに関するものであった．なお，M氏の講演会は1990年10月16日であり，調査時点は1990年7月中旬であった [61]．

この調査の中のM氏のイメージに関する回答を本数値実験で用いた．データは，あたたかい－つめたい，ばかだ－かしこい，やさしい－いじわるだ，おもしろい－つまらない，魅力がない－魅力がある，嫌いだ－好きだ，の6つの成分からなる．被験者はそれぞれの成分に関し幅を持って回答し，またその代表値も回答した (図4.7)．それぞれの回答は$[0,1]$に入るように標準化されている．この回答がそれぞれ非対称な三角ファジィ数であると仮定し (図4.8)，個体は三角ファジィ数の直積で表されると仮定する．

データは6つの成分からなるが，そのうちの2つの成分 (あたたかい－つめたい，おもしろい－つまらない) の代表値は図4.9のように分布している．数値実験は6つの成分全てを用いて行った．図4.10, 図4.11は代表値の回答を用いて既存のファジィc-平均法を行った結果 (+) と，幅を持った回答を非対称な三角ファジィ数とみなし (図4.8) 本手法でクラスタリングした結果 (最長距離法$*$，最短距離法$\times$) である．

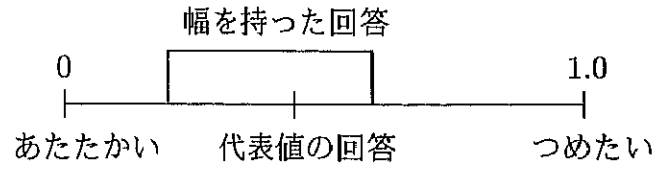

図 4.7: 1つの成分の回答例 ([61] による)

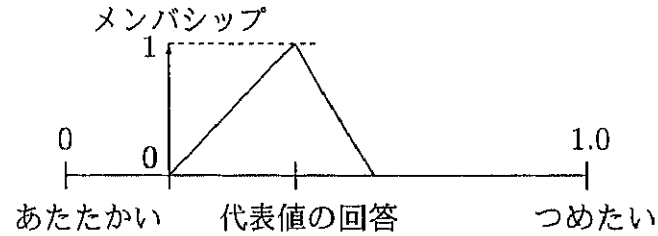

図 4.8: 得られた回答から仮定する三角ファジィ数

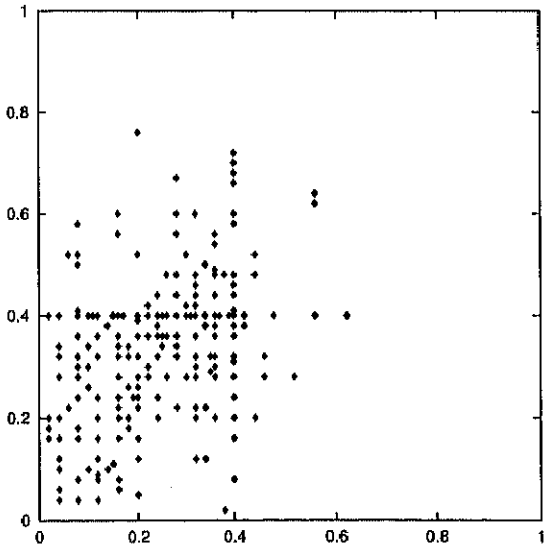

図 4.9: M 氏の印象 6 次元データの 2 成分 (横軸:あたたかいーつめたい，縦軸:おもしろいーつまらない)

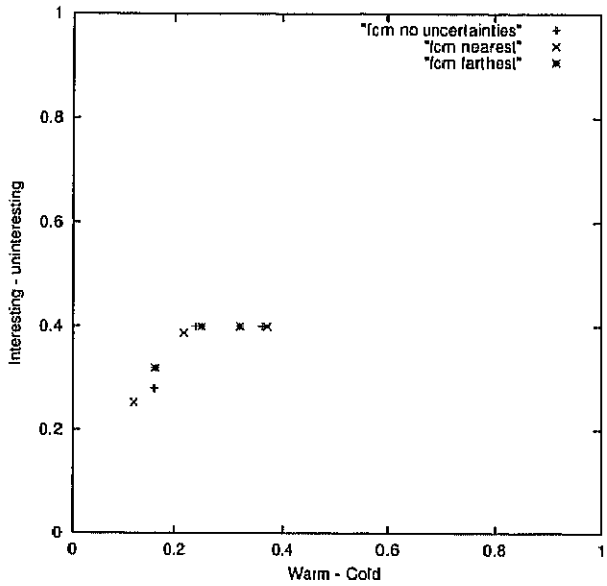

図 4.10: M 氏の印象 6 次元データのクラスタリング (標準的なファジィ c-平均法，3 分類) 結果の 2 成分 (横軸:あたたかい－つめたい，縦軸:おもしろい－つまらない)

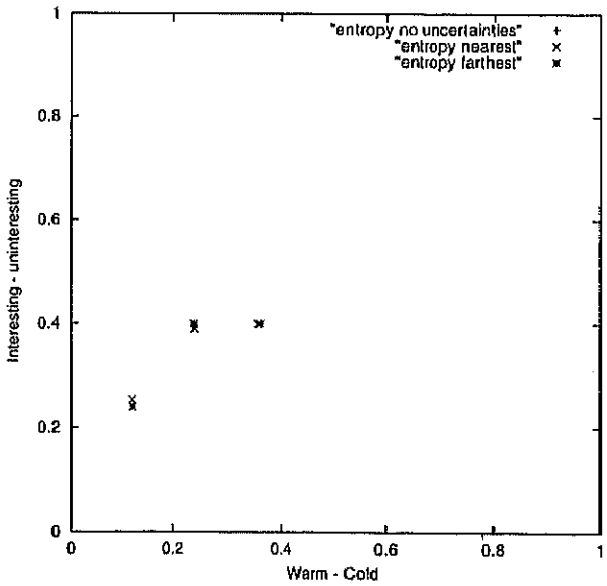

図 4.11: M 氏の印象 6 次元データのクラスタリング (エントロピー法，3 分類) 結果の 2 成分 (横軸:あたたかい－つめたい，縦軸:おもしろい－つまらない)

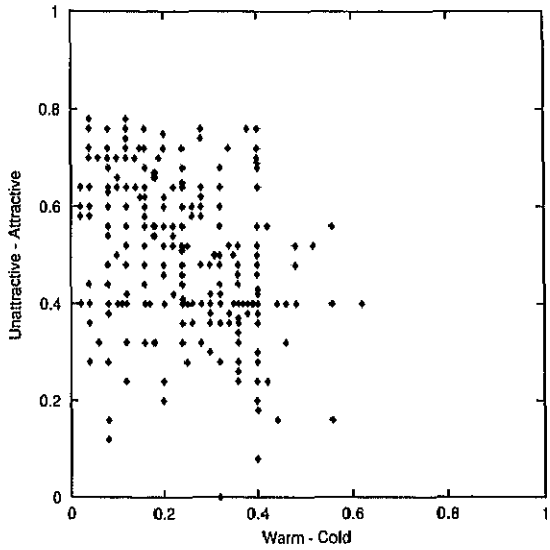

図 4.12: M 氏の印象 6 次元データの 2 成分 (横軸:あたたかい－つめたい，縦軸:魅力がない－魅力がある)

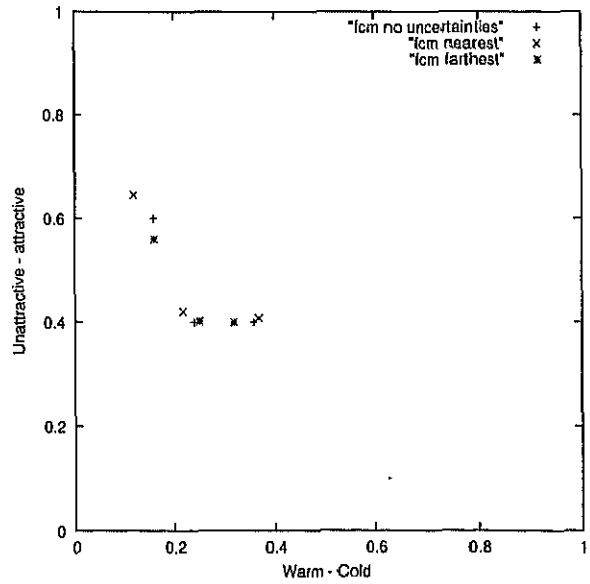

図 4.13: M 氏の印象 6 次元データのクラスタリング (標準的なファジィ c-平均法，3 分類) 結果の 2 成分 (横軸:あたたかい－つめたい，縦軸:魅力がない－魅力がある)

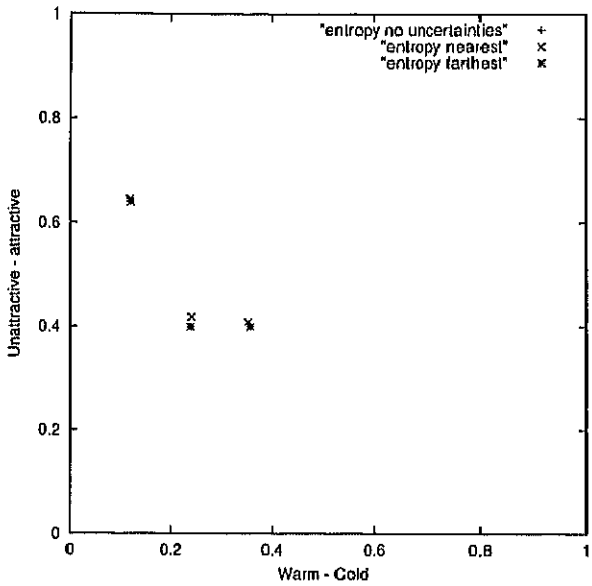

図 4.14: M 氏の印象 6 次元データのクラスタリング (エントロピー法，3 分類) 結果の 2 成分 (横軸:あたたかい－つめたい，縦軸:魅力がない－魅力がある)

ここでは 3 分類の結果を示しており，それぞれクラスター中心の位置を示している．標準的なファジィc-平均法ではパラメータ $m = 2.0$，エントロピー法ではパラメータ $\lambda = 5.0$ とした．横軸は (あたたかい－つめたい)[0,1]，縦軸は (おもしろい－つまらない)[0,1] である．ファジィc-平均法のアルゴリズム FCM における収束判定基準は，各ループでのメンバシップの変化の最大値がこの値より小さいことを使った．

$$\max_{i,k} \mid u_{ik} - \bar{u}_{ik} \mid < \varepsilon, \quad \varepsilon = 1.0 \times 10^{-6}. \tag{4.17}$$

標準的なファジィc-平均法を用いた場合，クラスター中心の位置が手法によって異なっている．エントロピー法を用いた場合にはクラスター中心の位置はどの手法でも似通っている．その理由を考えてみると，L_1 距離に基づくエントロピー法は，クラスター中心付近のデータは，ファジィc-平均法のアルゴリズムのメンバシップに関する最適化ステップにおいてほとんど同じ値を得る．このことにより，クラスター中心を計算するステップでは，不確定性を含まないデータの場合，最長，最短距離を用いたファジィデータの場合で，似通ったクラスター中心が得られる．このようなステップを繰り返した結果，最終的に似通ったクラスター中心が得られたと考えられる．

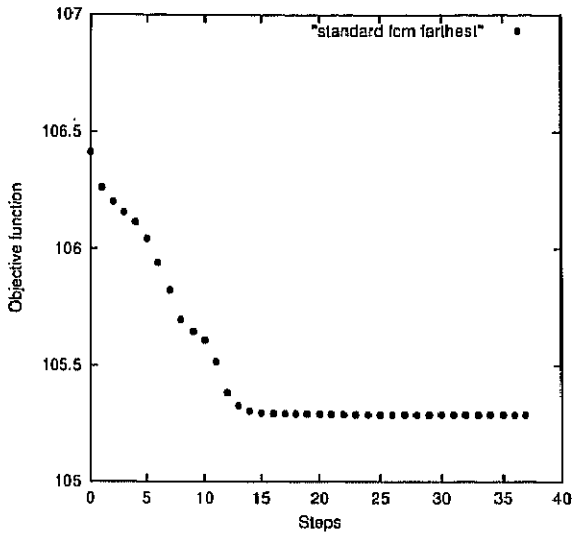

図 4.15: 目的関数が単調減少している例

また，図4.10の結果によく表れているように，よりあたたかく感じる人はより魅力があると感じる傾向と，よりつめたく魅力がないと感じる傾向の他に，比較的あたたかく感じているけれど，あまり魅力がないと感じている第3の傾向が存在することを示している．

また，同様に (あたたかい－つめたい，魅力がない－魅力がある) の2成分に関して，代表値は図4.12のように分布している．図4.13，4.14は代表値を用いて既存のファジィc-平均法を行った結果 (+) と，幅を持った回答を本手法でクラスタリングした結果 (最長距離法＊，最短距離法×) である．それぞれクラスター中心の位置を示している．標準的なファジィc-平均法では，それぞれの場合で異なるクラスター中心の配置が得られた．エントロピー法を用いた場合は，クラスター中心の位置はそれぞれの場合でほぼ同じ位置に収束した．本数値例では，不確定性を含むデータについて，不確定性を含まないと仮定した場合の通常のクラスタリングとは異なる結果が得られた．不確定性を含むデータに対して，複数の結果が得られることで，データに対してより多面的な視点が与えられる．

本章で述べたクラスター中心計算アルゴリズムでは，目的関数はクラスター中心に関して厳密に最適化 (最小化) が行われている．これは，ファジィc-平均法のアルゴリズムのクラスター中心の最適化のステップ **FCM3** において，目的関数が変化

しないかあるいは減少していることを意味する．また，ステップFCM2の最適化も厳密に行われているので，目的関数は同様に変化しないかもしくは減少する．このステップFCM2，ステップFCM3を繰り返すのがファジィc-平均法のアルゴリズムなので，目的関数はファジィc-平均法によって明らかに単調減少する．

このことを例示するために本数値実験における，実際に調査されたデータに対する実験を行ったときの，アルゴリズムFCMの各ループにおける目的関数の値を図4.15に示す．ここでは，標準的なファジィc-平均法(パラメータ$m = 2.0$)において最長距離法を用いて，3分類を行った場合を示している．クラスター中心に関する最適化(最小化)を厳密に行っているので，目的関数の値は単調に減少している．

4.7　まとめ

本章では，L_1距離に基づくファジィc-平均法でファジィデータをクラスタリングする手法を提案した．ファジィデータとして，三角ファジィ数の直積で表されるようなファジィデータを扱ったが，三角ファジィ数のメンバシップ関数は，線形であるために，一般のファジィ数を扱う場合に比べて，より単純な計算によって，クラスタリングアルゴリズムを実行できる．ここでは，不確定性をもつ個体データの場合，L_1空間がユークリッド空間に劣らず適切であることから，L_1空間にもとづくファジィクラスタリングアルゴリズムを開発した．

不確定性を扱う方法は唯一ではないが，ここでの手法は距離の最小値と最大値の両方を考察するアプローチを用いた．最長距離法，最短距離法の2種類の集合間の距離を用いて，ファジィデータとクラスター中心との間の距離を定義し，目的関数を定義した．他の集合間の距離を用いることも可能だが，それらの距離は最長距離法と最短距離法との間の性質を持つと考えられる．

ファジィc-平均法のアルゴリズムは，クラスター中心とメンバシップの2種類の変数を含む目的関数を，それぞれの変数について交互に最適化を行うことにより最適化するアルゴリズムである．ファジィデータに対するL_1距離に基づく目的関数を，ファジィc-平均法のアルゴリズムで厳密に最小化しようとすると，クラスター中心に関する最適解を既存のファジィc-平均法と同様に求めることはできない．そこでクラスター中心に関する最適化のアルゴリズムを新たに開発した．このアルゴリズムを用いることにより，偏導関数が負から正に転じる場所が求められる．これは目

的関数のクラスター中心に関する厳密な最適解が得られることを意味し，目的関数をファジィc-平均法のアルゴリズムで厳密に交互最適化することができ，区間データに対するファジィc-平均法が確立した．

本節では，ファジィデータを扱ったが，区間の直積で表現されるデータについてもL_1空間にもとづくファジィクラスタリングを考えることができる．本手法では直接扱うことはできないが，本手法と同様の考え方で，目的関数のクラスター中心に関する偏導関数を探索するアルゴリズムを用いれば，クラスタリングを行うことができる．このアルゴリズムは本章のアルゴリズムを簡単化した形になるが，これは，L_1距離に基づく不確定性を含まないデータに対するファジィc-平均法 [43] のアルゴリズムと同様となるので，ここでは省略している．

ある人物の印象についての不確定性を含むデータが三角ファジィ数の直積で表されると仮定し，本手法でクラスタリングした結果を求めた．また，同じデータの代表値を用いて，不確定性を含まないと仮定した場合の既存のファジィc-平均法による結果を求めた．これら2つの結果を比較すると，標準的なファジィc-平均法においては，手法によってクラスター中心が異なる位置に収束したが，エントロピー法においては，クラスター中心はほぼ同じ位置に収束した．

また，本論文ではL_1距離に基づくファジィc-平均法を扱ったが，ユークリッド距離の二乗を用いた場合には，目的関数のクラスター中心に関する偏導関数は，区分的で，非線形な関数となるため，クラスター中心に関する最適化にはより複雑な解法が必要となる．この手法については，将来の課題である．