解説

ファジィクラスタリングにおける ファジィネスとロバストネス†

— 正則化, ロバスト化と IRLS —

市橋 秀友*・本多 克宏*

1. はじめに

Kohonen のクラスタリングネットワーク (KCN) [20] は, ファジィc 平均法 (FCM 法) [1,15] と同様の教師無し学習法で, よく知られたクラスタリング法である. 文献 [31] で指摘されているようにそのアルゴリズムはヒューリスティックな方法であり, 目的関数が最適化されるわけではなく, 学習の収束が保障されず, クラスタリング結果はデータの入力順に依存する. しかし, 分かりやすくプログラミングも簡単なため, LVQ 法や SOM 法 [20] などへの発展と相まって, FCM 法以上に良く用いられる手法である. FCM 法も確率統計分野の手法に比べて平易で分かりやすく, また繰り返し重み付最小 2 乗法 (IRLS) [10] などの考え方を採用してさらなる簡単化も可能と考えられ, より広範な応用が期待される.

著者らは, クラスタ分析によってデータの集合をファジィ部分集合 (ファジィクラスター) に分割することによって, 知識発見やデータマイニングに生かそうとする研究に取り組んできた [11,12,14]. そこでは, 主要な傾向を捉え情報を縮約するために主成分分析 (PCA) が有効であり, 異常データや外れ値, さらに欠測値に影響されないためにロバスト M-推定 [16] も有効な方法である. ファジィクラスター分析法のなかでも FCM 法はよく知られた方法で, その解法アルゴリズムの不動点反復法は統計分野でのロバスト M-推定法で用いられる IRLS に類似している. ロバストなクラスタリング結果を得ようとするノイズクラスタリング法 [3,5] は FCM 法の修正アルゴリズムであるが M-推定との密

接な関係が指摘されている [4]. また, FCM 法は確率統計分野でのクラスタリング法であるガウス混合モデル [6,9] や確率的 PCA によるクラスター分析法である MPCA [28] と関連し, 特に主成分分析とクラスタリングの同時分析法であるファジィc 多様体法 (FCV 法) もこれらと関連が深い [13]. さらに決定論的アニーリング法 (DA) [26] と密接な関係にある. このようにファジィクラスタリング法は統計的な手法と多くの類似性がある反面, 確率論に則らないために, 分かりやすさやパラメータによる調整の自由度を高めている. 本稿ではこのような観点からファジィクラスタリングと統計手法との関連についてまとめ, 相違点について考察する.

2. ロバスト M-推定と FCM クラスタリング

M-推定は, 最も良く利用されるロバスト推定手法のひとつであり, 最小 2 乗法で用いられる最小 2 乗 (LMS) 基準の代わりに, それを少し変形し 2 乗よりも増大の仕方が小さい ρ -関数を用いる. ここで, 多次元データ集合の平均値 (中心, 重心) を外れ値に影響されずに推定することを考える. 目的関数を

$$L_{\rho} = \sum_{k=1}^n \sum_{j=1}^m \rho(r_{kj}) \quad (1)$$

とする. $r_{kj} = x_{kj} - v_j$ は m 次元データ $\mathbf{x}_k$ と平均の推定値 $\mathbf{v}$ との誤差の第 j 要素でスカラーである. そして, m 次元パラメータベクトル $\mathbf{v}$ を推定することが目的であるとする. 最小 2 乗基準はすべての誤差を均等な重みで扱っているため, ひとつの外れ値によって大きな影響を受けてしまう. そこで, M-推定では, 外れ値には小さな重みを与えるような評価基準としている. ここで, $\rho(r_{kj})$ は, $r_{kj} = 0$ で唯一の最小値を持つ正定値偶関数である. $\rho = (r_{kj})^2$ ならば, 最小 2 乗基準と同じになる. $\rho(r_{kj})$ による M-推定は次の連立方程式の解 $\mathbf{v}$ である.

$$\sum_{k=1}^n \psi(r_{kj}) \frac{\partial r_{kj}}{\partial v_j} = 0, \quad j = 1, \dots, m \quad (2)$$

† Fuzziness and Robustness in Fuzzy Clustering

—Regularization, Robustification and IRLS—
Hidetomo ICHIHASHI and Katsuhiro HONDA

* 大阪府立大学大学院 工学研究科 電気・情報系専攻
知能情報工学分野

Graduate School of Engineering, Electrical Engineering
and Information Science, Computer Sciences and Intelligent
Systems, Osaka Prefecture University

ただし, $\psi(z) = d\rho/dz$ である.

ここで, 重み関数を $w(z) = \psi(z)/z$ と定めると, 式 (2) は

$$\sum_{k=1}^n w(r_{kj}) r_{kj} \frac{\partial r_{kj}}{\partial v_j} = 0, \quad j = 1, \dots, m \quad (3)$$

となる. これは, 次の関数 L_{irls} の最小値を IRLS で求めることに等しい. 式 (4) を最小化するための必要条件は式 (3) の連立方程式と等しくなり, IRLS では重み w を更新しながら式 (3) を繰り返し解くことになる.

$$L_{irls} = \sum_{k=1}^n \sum_{j=1}^m w(r_{kj}) r_{kj}^2 \quad (4)$$

すなわち, 式 (4) を最小化する解は $\tilde{x}_k$ の重み付平均

$$v_j = \frac{\sum_{k=1}^n w(r_{kj}) x_{kj}}{\sum_{k=1}^n w(r_{kj})}, \quad j = 1, \dots, m \quad (5)$$

であるので, $r_{kj} = x_{kj} - v_j$, $w(r_{kj}) = \psi(r_{kj})/r_{kj}$ と重みを更新しながら式 (5) を繰り返すのが IRLS である.

関数 $\rho = (r_{kj})^2$ の場合はデータが v から離れるにつれてその影響が線形に増大する. 一方, たとえば, Geman-McClure の ρ -関数 [7], $\rho(z) = z^2/(z^2 + \sigma_j^2)$ では, データが v からある程度離れるとその影響はほとんどなくなり 0 に近づく. σ_j はスケールパラメータと呼ばれ ρ -関数の凸性の度合いを指定するものである.

クラスター分析の一つである FCM 法 [1, 15] はメンバシップを導入してデータの集合をいくつかのファジィクラスター (ファジィ部分集合) に分割する方法である. クラスターごとの中心 (重心) はプロトタイプと呼ばれ m 次元ベクトル v_i で表す. u_{ik} は k 番目のデータの i 番目のクラスターへのメンバシップとする. 良いクラスタリング結果を得るための目的関数は

$$\begin{aligned} L_{fcm} &= \sum_{i=1}^c \sum_{k=1}^n \sum_{j=1}^m (u_{ik})^\theta r_{kj}^2 \\ &= \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^\theta d_{ik}^2 \end{aligned} \quad (6)$$

と表される. c はクラスター数を表す.

$$d_{ik}^2 = (\tilde{x}_k - v_i)^\top (\tilde{x}_k - v_i) \quad (7)$$

は $\tilde{x}_k$ から i 番目のクラスター中心までのユークリッド距離の 2 乗である.

θ は fuzzifier とよばれ, 大きく設定するとよりファジィなクラスターが得られる. ただし, u_{ik} は非負でデータごとに和が 1 となるように基準化される. また解法

アルゴリズムは, 簡便さが優先されていて, L_{fcm} の未知パラメータについての 1 階微分を 0 とするように反復するもので, 不動点反復法や Picard 反復法と呼ばれる.

式 (6) は式 (4) でのデータの平均値をクラスごとに推定するようにしたものと考えられる. FCM 法でのメンバシップ関数は重み関数に対応する. 両者の相違点は, 式 (4) では重みを変数ごと (次元ごと) に付けられているのに対して, 式 (6) ではデータごとである. しかし, M-推定で重みをデータごとにしたり, FCM 法においてもメンバシップを各変数ごとにすることは可能である. 重み関数またはメンバシップ関数が誤差 r_{kj} のどのような関数として表されるかという点と, FCM 法ではクラスターへのメンバシップの和が, 各データごとに 1 と基準化される点も相違点である. この基準化はクラスター中心が競合的になるために必要で, これを省略してアルゴリズムを実行すれば, すべてのクラスターの中心が全データの中心に集まってしまう. ρ -関数から導かれる M-推定の重み関数は誤差の減少関数であり, さまざまな ρ -関数に対応する重み関数が提案されている. FCM 法では, 上記の L_{fcm} の最適性の必要条件から導かれる標準 FCM 法でのべき乗型のメンバシップ関数とエントロピー正則化 FCM [24] から導かれる指数関数型のメンバシップ関数が代表的である. 場合分けや基準化のための操作が必要であるが, FCM 法の変形版として, 他に 2 次正則化による方法 [32] や統計力学に基づいて De Luka と Termini のファジィエントロピーで正則化する方法 [33] などが提案されている.

エントロピー正則化 FCM 法は宮本と向殿により提案されたもので Shannon のエントロピー K と正のパラメータ λ を導入して L_{fcm} の代わりに $L_{\lambda fcm} = \sum_{i=1}^c \sum_{k=1}^n u_{ik} d_{ik}^2 + \lambda K$ を最小化する. 標準的な FCM 法ではどのクラスター中心からも遠く離れたデータ点のメンバシップは $1/c$ に近づくが, エントロピー正則化では遠ざかるほどクリスプになることが証明されている [22, 25]. この方法は目的関数最小化として定式化され解法アルゴリズムが導かれているが, 確率統計分野の決定論的アニーリング法 [26] での EM アルゴリズムによる解法とも密接に関係している. メンバシップ関数は

$$u_{ik}^{entro} = \exp(-d_{ik}^2/\lambda) \quad (8)$$

を和が 1 となるように基準化したものとなる. Welsch による M-推定での ρ -関数を用いて

$$\rho(d_{ik}) = \frac{\lambda}{2} (1 - \exp(-d_{ik}^2/\lambda)) \quad (9)$$

とすれば, その対応する重み関数 $w(d_{ik}) = \psi(d_{ik})/d_{ik}$ は式(8)に等しい. すなわち, エントロピー正則化 FCM 法のアルゴリズムは, Welsch の ρ -関数を用いる M-推定において, データの平均値を複数 (クラスター数) 求めるようにして, かつそれらが競合するように, 重み関数の和が1となるように基準化すれば同じものが導かれる. IRLS による M-推定では, ρ -関数から重み関数が導かれ, 重み付最小 2 乗法が用いられるが, 最初に重み関数を定義し, ρ -関数は陰に定めるだけとすれば, さまざまな M-推定法が考えられる. 同様に, FCM 法においても重み関数と類似の働きをしているメンバシップ関数を最初に定め IRLS で求解すれば, さまざまなメンバシップ関数を用いることができる. IRLS との関連は述べられていないが Runkler と Bezdek [27] は ACE (alternating cluster estimation) と呼び, 区分線形などのさまざまなメンバシップ関数を用いてクラスタリング結果の比較を行っている.

以上述べた M-推定と FCM との密接な関係から FCM 法はロバストなクラスタリング法であると言える. しかし, 次章で述べるノイズクラスタリングと比較すると, むしろスケールパラメータを一定値とする M-推定がファジィな推定法であると言ったべきかもしれない.

3. M-推定とノイズクラスタリング

クラスター分析も外れ値やノイズに影響されずにデータが密に集積した部分をクラスターとして抽出できれば有効である. 標準的な FCM 法はメンバシップの和が 1 に基準化されるのでその意味で確率的ファジィクラスタリングと呼ばれることがある. それに対して, メンバシップは単位区間 $[0, 1]$ の実数値であれば良いとするのが可能性的な制約を用いるクラスタリングで, この方法もデータが密に集積した部分をクラスターとして抽出する方法であると考えられる [21]. Davé によるノイズクラスタリング [3-5] も類似のものであるが, シミュレーションでは可能性的な制約を用いるクラスタリングに比べてうまく収束する場合が多い. これはノイズクラスタリングでは前述のメンバシップの和を 1 と基準化してクラスター中心同士を競合させるためと考えられる. ノイズクラスタリングでは, 外れ値やノイズデータは, それ以外とは異なるクラスター (またはクラス) として扱われる. 通常の (good) クラスター数を $c-1$ とし第 c クラスターをノイズクラスターとすると, ノイズクラスタリングの目的関数は

$$L_{nfc} = \sum_{i=1}^{c-1} \sum_{k=1}^n u_{ik} d_{ik}^2 + \delta \sum_{k=1}^n u_{ck}$$

$$+ \lambda \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log u_{ik} \quad (10)$$

ただし, $\sum_{i=1}^c u_{ik} = 1$, $k=1, \dots, n$ であり, 式(10)はエントロピー正則化の場合を示している. この目的関数を小さくするために, $d_{ik}^2 > \delta$ であれば, $\tilde{x}_k$ はノイズクラスターに含まれるようになる. δ は正の小さな値とすべきで, δ が大きければノイズクラスターに含まれる割合が小さくなるし, 負の値にすればすべてのデータがノイズと見なされる. したがって, δ はロバストにするための誤差の閾値であると考えられる.

最適性の必要条件よりメンバシップ関数は式(8)を用いて,

$$u_{ik} = \frac{u_{ik}^{entro}}{\sum_{l=1}^c u_{il}^{entro}}, \quad i=1, \dots, c, \quad k=1, \dots, n \quad (11)$$

ただし, $i=c$ または $l=c$ のとき $d_{ck}^2 = \delta$ である. また, $i=1, \dots, c-1$ について,

$$v_{ij} = \frac{\sum_{k=1}^n u_{ik} x_{kj}}{\sum_{k=1}^n u_{ik}}, \quad j=1, \dots, m \quad (12)$$

である.

文献 [4] ではノイズクラスタリングと M-推定の密接な関係が明らかにされているが, 詳細に検討すれば相違点も存在する. 前章の M-推定や FCM 法ではロバストにするための誤差の閾値に相当するパラメータ δ が含まれていない, ノイズクラスタリングでは重み関数に相当するメンバシップのパラメータ λ も閾値 δ も指定できるので真にロバストな方法といえる. λ はファジィネスを指定するパラメータで δ はロバストネスを指定するためのパラメータである. このことはクラスター数を $c=2$ にして通常のクラスターとノイズクラスターをそれぞれ 1 個ずつとし全データの平均値を推定する問題とすればより明確になる.

4. M-推定とロバスト主成分分析

M-推定を用いて主成分分析をロバストにする方法がいくつか存在する. 一つは分散共分散行列を重み付 (M-推定) にしてマハラノビス距離をロバストなものにする方法 [2] で Hampel の bounded influence function [8] などが用いられる. ロバストな分散共分散行列と平均は次のように表される.

$$S^* = \frac{\sum_{k=1}^n w_2(M_k^2) (\tilde{x}_k - v^*) (\tilde{x}_k - v^*)^T}{\sum_{k=1}^n w_2(M_k^2) - 1} \quad (13)$$

$$v^* = \frac{\sum_{k=1}^n w_1(M_k^2) \tilde{x}_k}{\sum_{k=1}^n w_1(M_k^2)} \quad (14)$$

ただし, $w_1(M_k^2)$ と $w_2(M_k^2)$ はスカラーの重みで, マハラノビス距離 M_k^2 の関数である. S^* は繰り返し推定さ

れる. Huber の重み関数 [16] を用いて $w_2(M_k^2) = (w_1(M_k^2))^2$ などとされる. この場合は重みは変数ごとでなくデータごとである.

一方, ファジィクラスタリングにおいてもクラスター容量を考慮した重み付きマハラノビス距離を基準とする K-L 情報量正則化 FCM 法が提案されている [19, 23]. クラスタ容量の最適化も行うこの手法では, 各クラスターが含むデータの比率も考慮してメンバシップが与えられる. Davé のノイズクラスタリングを K-L 情報量正則化にして, 式 (10) のエントロピー項を K-L 情報量に, 距離 d_{1k}^2 をマハラノビス距離にすればガウス混合モデルと類似のアルゴリズムが導かれる [17]. マハラノビス距離を定める分散共分散行列もアルゴリズムの中で反復的に決定される. 簡単のために $c=2$ とし第 2 クラスタをノイズクラスタとする. 目的関数を

$$\begin{aligned} L_{nklfcm} = & \sum_{k=1}^n u_{1k} d_{1k}^2 + \delta \sum_{k=1}^n u_{2k} \\ & + \lambda \sum_{k=1}^n u_{1k} \log \frac{u_{1k}}{\pi} \\ & + \lambda \sum_{k=1}^n u_{2k} \log \frac{u_{2k}}{1-\pi} \\ & + \sum_{k=1}^n u_{1k} \log |S| \end{aligned} \quad (15)$$

$$d_{1k}^2 = (\tilde{\mathbf{x}}_k - \mathbf{v})^\top S^{-1} (\tilde{\mathbf{x}}_k - \mathbf{v}) \quad (16)$$

とする. ただし, $1-\pi$ はノイズクラスタの容量が全体に占める割合を表してアルゴリズムの中で決定される. u_{1k} は d_{1k} の関数として表され, クラスタ中心は

$$\mathbf{v}_1 = \frac{\sum_{k=1}^n u_{1k} \tilde{\mathbf{x}}_k}{\sum_{k=1}^n u_{1k}} \quad (17)$$

分散共分散行列は

$$S = \frac{\sum_{k=1}^n u_{1k} (\tilde{\mathbf{x}}_k - \mathbf{v}_1)(\tilde{\mathbf{x}}_k - \mathbf{v}_1)^\top}{\sum_{k=1}^n u_{1k}} \quad (18)$$

となる [17]. 式 (13)–(14) の w_1 と w_2 が式 (18) の u_{1k} に対応している. ここでも, 式 (10) と同様に δ はロバスト化の為の誤差の閾値を調節するパラメータであり, 適当な値を設定できれば, 多くのノイズに埋もれた線形クラスタをうまく抽出できる [17].

上記の方法では重み w_1 , w_2 やメンバシップ u_{ci} はデー

タサンプルごとに外れ値と見なす程度を指定する重みである. したがって, 多次元データの要素ごとの外れ値の影響を取り除きたい場合には有効ではない. たとえば, 比較的少ない数の画像データを圧縮するために主成分分析を用いるときに, 一枚の画像全体をノイズ扱いするのではなく画像のピクセルごとに外れ値を検出してその影響を取り除いた方が捨て去られる情報を少なくできる. また, 画像データのような次元数の大きなデータでは分散共分散行列の計算だけでも $O(nm^2)$ となり難しくなる. 主成分分析は形や動きの抽出などコンピュータビジョンの分野でよく用いられていて, サンプル内ノイズを含む画像データの主成分分析法が Torre と Black [29, 30] により提案されている. $X = (x_{kj})$ を $(n \times m)$ のデータ行列で, n 個の m 次元データからなるとする. X の第 j 列を要素とする n 次元列ベクトル $\mathbf{x}_j, j=1, \dots, m$ を用いて, $X = (\mathbf{x}_1, \dots, \mathbf{x}_m)$, または X の第 k 行を要素とする m 次元列ベクトル $\tilde{\mathbf{x}}_k, k=1, \dots, n$ を用いて $X = (\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_n)^\top$ と表す. $\sim$ は行列の行を要素とする列ベクトルであることを表している. 主成分はデータ行列 X の低階数近似として抽出され, それは

$$\text{tr}((X - FP^\top - \mathbf{1}_n \mathbf{v}^\top)^\top (X - FP^\top - \mathbf{1}_n \mathbf{v}^\top)) \quad (19)$$

を

$$P^\top P = I_p \quad (20)$$

のもとで最小化することにより求められる. ただし, $P = (\tilde{\mathbf{p}}_1, \dots, \tilde{\mathbf{p}}_m)^\top, p < m, F = (\tilde{\mathbf{f}}_1, \dots, \tilde{\mathbf{f}}_n)^\top$, で $\mathbf{1}_n$ はすべての要素が 1 である $n \times 1$ ベクトルである. I_p は p 次元単位行列である. Torre と Black は, これを要素ごとに表して, かつ最小 2 乗誤差基準を ρ -関数に修正して目的関数を

$$L_{rpca} = \sum_{k=1}^n \sum_{j=1}^m \rho(x_{kj} - \sum_{q=1}^p f_{iq} p_{jq} - v_j) \quad (21)$$

としている. 文献 [29, 30] では Geman-McClure [7] の ρ -関数が用いられている. 式 (21) を最小化するために重み関数 $w(r_{kj})$, $r_{kj} = x_{kj} - \sum_{q=1}^p f_{iq} p_{jq} - v_j$ を用いて IRLS の式 (3) に対応させて書き換えると, $j=1, \dots, m, k=1, \dots, n$ について,

$$(F^\top W_j F) \tilde{\mathbf{p}}_j = F^\top W_j (\mathbf{x}_j - \mathbf{1}_n v_j) \quad (22)$$

$$P^\top \tilde{W}_k P \tilde{\mathbf{f}}_k = P^\top \tilde{W}_k (\tilde{\mathbf{x}}_k - \mathbf{v}) \quad (23)$$

$$\left(\sum_{k=1}^n \tilde{W}_k \right) \mathbf{v} = \sum_{k=1}^n \tilde{W}_k (\tilde{\mathbf{x}}_k - P \tilde{\mathbf{f}}_k) \quad (24)$$

なる連立方程式が得られる. $W_j = \text{diag}(w_{1j}, \dots, w_{nj})$, $\tilde{W}_k = \text{diag}(w_{k1}, \dots, w_{km})$, $w_{kj} = w(r_{kj})$ である.

IRLS は式 (22) を $\hat{p}_j$ について, 式 (23) を $\hat{f}_k$ について, 式 (24) を v について交互に解いてその解を代入することを目的関数 (21) が収束するまで繰り返す. また, スケールパラメータ σ_j を温度と見なした決定論的アニーリングも用いられている.

この方法では, データ集合全体の主成分を外れ値に影響されないように抽出することが目的であるが, いくつかのクラスターに分割して同時にその主成分を求めるには, FCV 法に M-推定を用いて IRLS アルゴリズムを適用すればよい [12]. また, M-推定の代わりにノイズクラスタリングを用いてロバストネスを調整することもできる [18].

5. おわりに

ファジィクラスタリングにおけるファジィネスとロバストネスについて, 統計分野の M-推定との関連を中心に紹介した. IRLS ではノイズと見なされる要素に小さな重みを与えることでその影響を無視するが, 分析における重要度を調節する重みはむしろファジィクラスタリングにおけるメンバシップ (ファジィネス) に類似しているといえる. 一方, ノイズクラスタリングに代表されるロバスト手法を M-推定と関連付けた論文もあるが, ロバストネスを指定するパラメータをファジィネスのそれと区別することで, 異なる概念に基づく手法であると考えられる. したがって, M-推定の代わりにノイズクラスタリングを用いてロバストネスを調整する手法に関する研究が期待される.

また, ノイズと見なされる要素に小さな重みを与えるロバスト手法は, データに含まれる欠測値 (未観測値) を無視して観測されたデータのみからモデルを推定する欠測値処理手法に通じる点がある. IRLS では要素ごとの重みも決定変数であるが, 欠測値はモデルの推定から排除される要素があらかじめ分かっている特別な場合と考えることができ, 式 (3) の対応する重みを 0 と固定することに相当する. 両者の類似性は応用例においても顕著である. Torre と Black [29, 30] は定点カメラの映像から通り過ぎる人物をノイズとして除去することで背景画像を復元しているが, ロバストな部分空間上への射影を用いたデータの復元 (推定) と見なされることから, 主成分を用いた欠測値推定 [12, 14] と概念が似ている. その他, ノイズと欠測値の処理を併用することも可能であり, 従来は困難であった実世界の大規模データベースの解析における有用性が期待できる.

参 考 文 献

- [1] J. C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, 1981.
- [2] N. A. Campbell, "Robust procedures in multivariate analysis I: Robust covariance estimation," *Applied Statistics* 29, pp. 231-237, 1980.
- [3] R. N. Davé, "Characterization and detection of noise in clustering," *Pattern Recognition Letters*, vol.12, pp.657-664, 1991.
- [4] R. N. Davé and R. Krishnapuram, "Robust clustering methods, a unified approach," *IEEE Trans. Fuzzy Syst.*, vol.5, no.2, pp.270-293, 1997.
- [5] R. N. Davé and S. Sen, "Robust fuzzy clustering of relational data," *IEEE Trans. Fuzzy Syst.*, vol. 10, no.6, pp.713-727, 2002.
- [6] R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
- [7] S. Geman and D. E. McClure, "Statistical methods for tomographic image reconstruction," *Bulletin of International Statistical Institute*, vol. LII-4, pp. 5-21, 1987.
- [8] F. R. Hampel, "The influence curve and its role in robust estimation," *J. Amer. Statist. Ass.*, vol.69, pp.383-393, 1974.
- [9] R. J. Hathaway, "Another interpretation of the EM algorithm for mixture distributions," *Statistics and Probability Letters*, vol.4, pp.53-56, 1986.
- [10] P. W. Holland, and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Communications in Statistics*, vol. A6, no. 9, pp. 813-827, 1977.
- [11] K. Honda and H. Ichihashi, "Linear fuzzy clustering techniques with missing values and their application to local principal component analysis," *IEEE Trans. Fuzzy Syst.*, vol. 12, no. 2, pp. 183-193, 2004.
- [12] K. Honda and H. Ichihashi, "Component-wise robust linear fuzzy clustering for collaborative filtering," *Int. J. Approximate Reasoning*, vol.37, no.2, pp.127-144, 2004.
- [13] K. Honda and H. Ichihashi, "Regularized linear fuzzy clustering and probabilistic PCA mixture models," *IEEE Trans. Fuzzy Syst.* (in press).
- [14] K. Honda, N. Sugiura, H. Ichihashi and S. Araki, "Collaborative filtering using principal component analysis and fuzzy clustering," *Web Intelligence: Research and Development*, Lecture Notes in Artificial Intelligence 2198, pp. 394-402, Springer, 2001.
- [15] F. Höppner, F. Klawonn, R. Kruse, T. Runkler, *Fuzzy Cluster Analysis, (Methods for Classification, Data Analysis and Image Recognition)*, John Wiley & Sons, Ltd., 1999.
- [16] P. J. Huber, *Robust Statistics*, 1st Edn., Wiley, New York, 1981.
- [17] H. Ichihashi and K. Honda, "On parameter setting in applying Davé's noise fuzzy clustering to gaussian mixture models," *Proc. 2004 IEEE Int. Conf. Fuzzy Syst.*, #0265-1090, 2004.
- [18] H. Ichihashi and K. Honda, "Robust PCA with

- intra-sample outlier process using noise fuzzy clustering," *Proc. of the 5th International Conference on Recent Advances in Soft Computing*, pp. 648-653, 2004.
- [19] H. Ichihashi, K. Miyagishi and K. Honda, "Fuzzy c -means clustering with regularization by K-L information," *Proc. of 10th IEEE Int. Conf. Fuzzy Syst.*, pp.924-927, 2001.
- [20] T. Kohonen, *Self-Organization and Associative Memory*, 3rd Edn. Springer, Berlin, 1989.
- [21] R. Krishnapuram and J. M. Keller, "A possibilistic approach to clustering," *IEEE Trans. Fuzzy Syst.*, vol.1, pp. 98-110, 1993.
- [22] Z.-Q. Liu, S. Miyamoto (Eds.), *Softcomputing and Human-Centered Machines*, Springer-Verlag, 2000.
- [23] 宮岸聖高, 市橋秀友, 本多克宏, K-L 情報量正則化 FCM クラスタリング法, 日本ファジィ学会誌, vol. 13, no. 4, pp. 406-417, 2001.
- [24] S. Miyamoto and M. Mukaidono, "Fuzzy c -means as a regularization and maximum entropy approach," *Proc. of the 7th International Fuzzy Systems Association World Congress (IFSA '97)*, Vol.II, pp.86-92, 1997.
- [25] 宮本定明, 馬屋原一孝, 向殿政男, ファジィ c -平均法とエントロピー正則化法におけるファジィ分類関数, 日本ファジィ学会誌, vol. 10, no. 3, pp.156-164, 1998.
- [26] K. Rose, "Deterministic annealing for clustering, compression, classification, regression, and related optimization problems," *Proc. of the IEEE*, vol.86, no.11, pp.2210-2239, 1998.
- [27] T. A. Runkler and J. C. Bezdek, "Alternating cluster estimation: a new tool for clustering and function approximation," *IEEE Trans. Fuzzy Syst.*, vol. 7, no. 4, pp. 377-393, 1999.
- [28] M. E. Tipping and C. M. Bishop, "Mixtures of probabilistic principal component analysers," *Neural Computation*, vol.11, pp. 443-482, 1999.
- [29] F. de la Torre and M. J. Black, "Robust principal component analysis for computer vision," *Proc. of International Conference on Computer Vision*, pp. 362-369, 2001.
- [30] F. de la Torre and M. J. Black, "A framework for robust subspace learning," *Int. J. of Computer Vision*, vol.54, pp.117-142, 2003.
- [31] E. C-K. Tsao, J. C. Bezdek, N. R. Pal, "Fuzzy Kohonen clustering networks," *Pattern Recognition*, vol.27, no.5, pp.757-764, 1994.
- [32] 馬屋原一孝, 宮本定明, 2 次正則化によるファジィクラスタリング, 日本ファジィ学会誌, vol.10, no.1, pp.907-914, 1998.
- [33] 安田真, 古橋武, 松崎元昭, 大熊繁, 確定的アニーリング法を用いたファジィクラスタリングとその統計力学的性質, 日本ファジィ学会誌, vol.13, no.5 pp. 522-529, 2001.

(2005年1月24日 受付)

[問い合わせ先]

〒599-8531

大阪府堺市学園町1-1

大阪府立大学大学院工学研究科

電気・情報系専攻知能情報工学分野

市橋 秀友

TEL : 072-254-9352

FAX : 072-254-9915

E-mail: ichi@cs.osakafu-u.ac.jp

著者紹介

 いちはし ひでとも
 市橋 秀友 [正会員]

1971年大阪府立大学工学部経営工学科卒業。同年松下電器産業(株)入社, 1981年大阪府立大学工学部経営工学科助手, 1987年同講師, 1989年同助教授, 1993年同教授, 2000年同大学院工学研究科教授, 現在に至る。工学博士。ファジィクラスタリングやニューラルネットワークなどでのデータ解析法, その知的システムや人間情報システムへの応用研究に従事。IEEE, 日本知能情報ファジィ学会, システム制御情報学会, 電子情報通信学会, 日本経営工学会などの会員。

 ほんだ かつひろ
 本多 克宏 [正会員]

1999年大阪府立大学大学院工学研究科博士前期課程電気・情報系専攻修了。同年日本電信電話(株)入社, 同年大阪府立大学工学部経営工学科助手, 2000年同大学院工学研究科助手, 現在に至る。博士(工学)。ニューラルネットワーク, ファジィクラスタリングの研究に従事。IEEE, 日本知能情報ファジィ学会, システム制御情報学会, 日本経営工学会の会員。