

K-L 情報量正則化 FCM クラスタリング法

Fuzzy c -Means Clustering with Regularization by K-L Information

日本ファジィ学会誌, 13, 4, 406-417 (2001)

宮岸 聖高
市橋 秀友
本多 克宏

大阪府立大学大学院 工学研究科 電気・情報系専攻 経営工学分野
〒 599-8531 大阪府堺市学園町 1 - 1

Kiyotaka MIYAGISHI
Hidetomo ICHIHASHI
Katsuhiro HONDA

Graduate School of Engineering, Electrical Engineering and Information
Science, Industrial Engineering, Osaka Prefecture University
1-1 Gakuen-cho, Sakai, Osaka 599-8531, Japan

要 約

Fuzzy c -Means (FCM 法) は目的関数の最適化によるクラスタリング法である．一方，ガウス混合モデルは複数の正規分布を足し合わせて密度関数を表現し，パラメータ推定に EM アルゴリズムを適用する方法で一種のクラスタリング法とも考えられている．本研究では K-L 情報量正則化による FCM 法 (KFCM 法) を提案し，ガウス混合モデルとの類似点と相違点を議論する．ガウス混合モデルは提案法におけるファジィ度を制御するパラメータ λ が特定の値を持つときのみ同等のアルゴリズムが導かれるが特定の値以外の場合には対応する混合モデルが存在しないことを示す．そのことから，KFCM 法ではパラメータ λ の値を温度とみなし，アニーリングを行なうことが可能で，クラスタリング結果を改善できることを示す．また，Gustafson-Kessel の制約項を導入した GKFCM 法を提案し，ノイズのあるデータに対して可能性的クラスタリングと同様の結果を得ることができることを示す．

キーワード

ファジィクラスタリング，ガウス混合モデル，K-L 情報量正則化，可能性的クラスタリング

Abstract

Gaussian mixture model with EM algorithm is a popular density estimation method that uses the likelihood function as a measure of fit. It can be used as a tool for clustering. The thesis of the paper is that although the iterative algorithm of Fuzzy c -Means (FCM) clustering with entropy regularization is similar to that of the Gaussian mixture model, the FCM clustering has more flexible structure since the algorithm is based on the objective function method. We show that just the same algorithm as the Gaussian mixture model can be derived from a modified objective function with regularization by K-L information, and in a slightly different manner such as installing an annealing parameter and addition of Gustafson and Kessel's constraints, the proposed algorithm provides more valid or useful clustering results.

Keywords

Fuzzy clustering, Gaussian mixture model, Regularization by K-L Information, Possibilistic clustering

1 はじめに

クラスタリングの手法は大別して、階層的クラスタリングと非階層的クラスタリングに分かれる．非階層的方法に、クラスタリング結果の良さを評価するための目的関数を設定する、最適化クラスタリングの一つである K-平均法 (K-means) がある．それから発展した FCM(fuzzy c-means) 法 [1] でも目的関数の局所最適解が最適なクラスタリングであるとされる．

類似性の測度としてユークリッド距離を用いる場合は、目的関数にはクラスタ中心からの 2 乗距離の総和が用いられ、距離の定義や目的関数の変更によって多くのファジィクラスタリング法が研究され、開発されてきた．一方、確率統計やニューラルネットの分野でも類似のアルゴリズムが開発されている．中でも、ガウス混合モデル [2, 3] は複数の正規分布を足しあわせて密度関数を表現し、パラメータ推定に EM アルゴリズムを適用する方法であり、確率ニューラルネットワークの一種とも見なされている．ガウス混合モデルは、最近活発に研究されているエントロピー正則化による FCM 法 [4, 5] とそのアルゴリズムに類似性がある [6]．

本論文では、K-L 情報量 (相互情報量) 正則化による FCM 法 (KFCM 法) を提案し、ガウス混合モデルにおける EM アルゴリズムとの相違点を議論し FCM 法の有効性を明らかにする．まず、提案の KFCM 法でのファジィ度を制御するパラメータ λ を 2 とするとガウス混合モデルと同じアルゴリズムが導かれるが、2 以外の場合には対応する混合モデルが存在しないことを示す．そのことから、KFCM 法ではパラメータ λ の値を温度とみなし、反復アルゴリズムの中で逐次減少させる、いわゆるアニーリングが可能で、クラスタリング結果を改善できることを示す．また、Gustafson-Kessel [7] の制約項を導入した KFCM 法を提案し、可能性的クラスタリング [8, 9] と類似したクラスタリング結果を得ることができることを示す．数値実験ではガウス混合モデルとの分類結果や局所解の頻度、Gustafson-Kessel の制約を付加する場合との比較を行い、それぞれの特徴を明らかにする．

2 ガウス混合モデルと KFCM 法

正規分布や χ^2 分布などの単一の分布型を仮定しない確率密度関数の推定法にガウス混合モデルがある．そこでは、判別分析のためのクラスも考慮される場合があるが、ここではクラスは一つであるとする．密度関数は複数の正規分布の凸結合で表されるので、各正規分布の平均の周りにデータが分布するように、平均 (中心) が定められる．したがって、クラスタ中心の周りにデータが分布するように分割する、ファジィクラスタリングによく似た方法であると考えられる．ファジィクラスタリングではファジィ部分集合へのメンバシップによって分割するのに対して、ガウス混合モデルでは、正規分布の確率密度関数に従って個体が生起すると考える．データの集合をクラスタの集まりと見るか、確率分布の集まりと見るかの違いはあるが、両者には類似性がある．

n 個の個体 (サンプル) の特徴量を s 次元データベクトル \mathbf{x}_k 、 c 個の正規分布の平均を s 次元ベクトル \mathbf{v}_i とする．また、 ϕ^* をパラメータの推定値、 ϕ を更新後のパラメータ推定値とする．ガウス混合モデルは密度関数を

$$g(\mathbf{x}|\phi) = \sum_{i=1}^c \pi_i p_i(\mathbf{x}|\phi_i) \quad (1)$$

$$p_i(\mathbf{x}|\phi_i) = \frac{1}{(2\pi)^{\frac{s}{2}} |A_i|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{v}_i)^T A_i^{-1}(\mathbf{x} - \mathbf{v}_i)\right) \quad (2)$$

と表して、未知パラメータである行列 A_i 、平均 \mathbf{v}_i 、足しあわせの比率 π_i を最尤法によって求める． \mathbf{x}_k が与えられた場合のモデル l の生起する事後確率を

$$u_{lk} = \frac{\pi_l^* p_l(\mathbf{x}_k|\phi_l^*)}{\sum_{j=1}^c \pi_j^* p_j(\mathbf{x}_k|\phi_j^*)} \quad (3)$$

とすると、EM アルゴリズムは、対数尤度

$$\begin{aligned} Q(\phi|\phi^*) &= \sum_{i=1}^c \sum_{k=1}^n \log[\pi_i p_i(\mathbf{x}_k|\phi_i)] u_{ik} \\ &= \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log \pi_i \\ &\quad + \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log p_i(\mathbf{x}_k|\phi_i) \end{aligned} \quad (4)$$

の最大化として求められる．

まず，E ステップとして

$$u_{lk} = \frac{\pi_l \exp\left(-\frac{1}{2}(\mathbf{x}_k - \mathbf{v}_l)^T A_l^{-1}(\mathbf{x}_k - \mathbf{v}_l)\right) |A_l|^{-\frac{1}{2}}}{\sum_{i=1}^c \pi_i \exp\left(-\frac{1}{2}(\mathbf{x}_k - \mathbf{v}_i)^T A_i^{-1}(\mathbf{x}_k - \mathbf{v}_i)\right) |A_i|^{-\frac{1}{2}}} \quad (5)$$

M ステップとして

$$\pi_i = \frac{1}{n} \sum_{k=1}^n u_{ik} \quad (6)$$

また， A_i と \mathbf{v}_i は，式 (4) の第 1 項目には含まれていないので，第 2 項目から，

$$A_i = \frac{\sum_{k=1}^n u_{ik} (\mathbf{x}_k - \mathbf{v}_i)(\mathbf{x}_k - \mathbf{v}_i)^T}{\sum_{k=1}^n u_{ik}} \quad (7)$$

$$\mathbf{v}_i = \frac{\sum_{k=1}^n u_{ik} \mathbf{x}_k}{\sum_{k=1}^n u_{ik}} \quad (8)$$

となる．上記の E ステップと M ステップを交互に繰り返す，EM アルゴリズムが知られている [3, 10]．EM アルゴリズムでは，行列 A_i も決定変数としていて， A_i が正定値行列とならない場合が起こりうる．また局所解に陥りやすいという問題もしばしば指摘されている．

J.C.Bezdek による文献 [1] の FCM 法は，メンバシップを求めることで個体の集合をファジィ部分集合に分割するクラスタリング法である．前述の c 個の正規分布の平均に代えて，クラスタの中心ベクトルを s 次元ベクトル \mathbf{v}_i とし，個体 k がクラスタ i に属する度合いをメンバシップ u_{ik} とする．目的関数は次のように定められている．

$$J_m = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m d_{ik} \quad (9)$$

ただし，メンバシップ u_{ik} は非負で， c 個のクラスタについての和は各個体毎に 1 であるという制約条件の下での最小化を考える．

$$d_{ik} = (\mathbf{x}_k - \mathbf{v}_i)^T A_i^{-1}(\mathbf{x}_k - \mathbf{v}_i) \quad (10)$$

は，ユークリッド距離やマハラノビス距離などの，事前に選ばれた重み付き距離である． A_i を単位行列としてユークリッド距離が用いられることが多いが，D.E.Gustafson と W.C.Kessel による修正 FCM 法 [7]

では $|A_i|$ の大きさに制約を付加していて， A_i の要素も決定変数である．指数の m は Bezdek の論文 [11] で初めて導入されたもので， J_m はメンバシップの m 乗で重みづけられた各個体からクラスタ中心までの距離の和である．したがって， J_m を最小化する解はクラスタ中心からの距離を最小化するファジィクラスタである．添え字の集合を次のように定めると，

$$I_k = \{i | 1 \leq i \leq c; d_{ik} = \|\mathbf{x}_k - \mathbf{v}_i\| = 0\}$$

$$\tilde{I}_k = \{1, 2, \dots, c\} - I_k$$

J_m の最適性の必要条件はメンバシップ u_{ik} について， $I_k = \phi$ ならば，

$$u_{ik} = \frac{1}{\sum_{j=1}^c \left(\frac{d_{jk}}{d_{jk}}\right)^{\frac{1}{m-1}}} \quad (11)$$

$I_k \neq \phi$ ならば，

$$u_{ik} = 0 \quad \forall i \in \tilde{I}_k \text{ かつ } \sum_{i \in I_k} u_{ik} = 1 \quad (12)$$

また，クラスタ中心 \mathbf{v}_i について

$$\mathbf{v}_i = \frac{\sum_{k=1}^n (u_{ik})^m \mathbf{x}_k}{\sum_{k=1}^n (u_{ik})^m} \quad (13)$$

となる．

すべての u_{ik} を固定すると，すべての \mathbf{v}_i の最適値が求まり，逆にすべての \mathbf{v}_i を固定すると，すべての u_{ik} の最適値を求めることができる．そこで，FCM 法のアルゴリズムは， J_m の最適性の必要条件である式 (11)～式 (13) により，決定変数のグループ毎に最適解を求めることを繰り返すもので，Picard 反復と呼ばれている．

メンバシップのべき乗を用いるオリジナルな FCM 法には，式 (12) のような技術的トリックが存在する．FCM 法では x_k とクラスタ中心 \mathbf{v}_i が同じ値で，その間の距離が $d_{ik} = 0$ でメンバシップの算式 (11) の分母が 0 となることを特異であると呼ぶ．一方，文献 [4, 5] で，宮本らはラグランジュ乗数法で解が求まらないという意味でクリスプなクラスタを特異と呼び，ファジィクラスタが求まるようにエントロピー項を導入する方法を提案している．エントロピー正則化による場合は上記のような技術的トリックは必要でない．一般に正則化は滑らかな解を得るための方法という意味合いで使

われることがあり，目的関数 J_1 にある種の正則化項 K を正数パラメータ λ とともに付加し， $J = J_1 + \lambda K$ を最適化することはしばしば行われている．最適化される目的関数は， J_m の代わりに，

$$J_\lambda = \sum_{i=1}^c \sum_{k=1}^n u_{ik} d_{ik} + \lambda \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log u_{ik} \quad (14)$$

を採用する．ラグランジュ乗数法を用いることによって u_{ik} の最適解が次式で与えられる [5]．

$$u_{ik} = \frac{\exp(-\frac{1}{\lambda} d_{ik})}{\sum_{j=1}^c \exp(-\frac{1}{\lambda} d_{jk})} \quad (15)$$

また， v_i の最適解は次のようになる．

$$v_i = \frac{\sum_{k=1}^n u_{ik} \mathbf{x}_k}{\sum_{k=1}^n u_{ik}} \quad (16)$$

式 (13) では平均 (中心) の計算がメンバシップの m 乗の重み付きとして求められるが，式 (16) では m 乗が不要である．メンバシップのべき乗による FCM 法では，十分遠くにある個体はどのクラスタにもほぼ等しいメンバシップを持つようになるが，エントロピー正則化では，0 あるいは 1 に近づく [5]．

本論文では，エントロピー正則化 FCM 法の目的関数である式 (14) のエントロピー項を K-L 情報量に変更し，制約条件をラグランジュ関数として含めて

$$\begin{aligned} J_{\lambda\tau} = & \sum_{i=1}^c \sum_{k=1}^n u_{ik} d_{ik} + \lambda \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log \frac{u_{ik}}{\pi_i} \\ & + \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log |A_i| - \sum_{k=1}^n \eta_k \left(\sum_{i=1}^c u_{ik} - 1 \right) \\ & - \tau \left(\sum_{i=1}^c \pi_i - 1 \right) \end{aligned} \quad (17)$$

とする． d_{ik} は式 (10) とし， η_k と τ はラグランジュ乗数で，対応する項はそれぞれ，メンバシップおよび新たに導入した変数 π_i の i についての c 個の和が 1 であることを表している．式 (14) のエントロピー項はメンバシップができるだけ均等になるように設けられたものである．一方，式 (17) の第 2 項目は u_{ik} と π_i がすべての i について等しければ，0 となり， u_{ik} と π_i の分布の近さを測るための K-L 情報量を表している．すなわち，クラスタ i 内の個体のメンバシップ u_{ik} がすべて π_i に，できるだけ等しくなるように設けられ

た項で，パラメータ λ でその程度を指定する．K-L 情報量正則化項を導入することで，クラスタの容量を考慮しながらクラスタリングを行なうことになり，大きさに差があるクラスタを適当に分割することができる [12]．提案のファジィクラスタリング法は確率統計理論におけるエントロピーや K-L 情報量の形式を採用しているが，ファジィなクラスタを得ようとするもので，EM アルゴリズムのような確率的アプローチではない．

式 (17) の最適性の必要条件

$$\frac{\partial J_{\lambda\tau}}{\partial u_{ik}} = 0 \quad (18)$$

より，

$$u_{ik} = \pi_i |A_i|^{-\frac{1}{\lambda}} \exp\left(-\frac{1}{\lambda} d_{ik}\right) \exp\left(\frac{\eta_k}{\lambda} - 1\right) \quad (19)$$

また，

$$\sum_{i=1}^c u_{ik} = 1 \quad (20)$$

であるので，

$$u_{ik} = \frac{\pi_i \exp(-\frac{1}{\lambda} d_{ik}) |A_i|^{-\frac{1}{\lambda}}}{\sum_{j=1}^c \pi_j \exp(-\frac{1}{\lambda} d_{jk}) |A_j|^{-\frac{1}{\lambda}}} \quad (21)$$

さらに，

$$\frac{\partial J_{\lambda\tau}}{\partial \pi_i} = 0 \quad (22)$$

より，

$$\pi_i = \frac{-\lambda}{\tau} \sum_{k=1}^n u_{ik} \quad (23)$$

また，

$$\sum_{i=1}^c \pi_i = 1 \quad (24)$$

より，

$$\pi_i = \frac{\sum_{k=1}^n u_{ik}}{\sum_{j=1}^c \sum_{k=1}^n u_{jk}} = \frac{1}{n} \sum_{k=1}^n u_{ik} \quad (25)$$

となる．上式より π_i はファジィクラスタ i に含まれるデータの容量 (割合) であるといえる．同様に行列 A_i の jl 成分を a_i^{jl} とすると，

$$\frac{\partial J_{\lambda\tau}}{\partial a_i^{jl}} = 0 \quad (26)$$

より,

$$A_i = \frac{\sum_{k=1}^n u_{ik}(\mathbf{x}_k - \mathbf{v}_i)(\mathbf{x}_k - \mathbf{v}_i)^T}{\sum_{k=1}^n u_{ik}} \quad (27)$$

$$\frac{\partial J_{\lambda\tau}}{\partial \mathbf{v}_i} = \mathbf{0} \quad (28)$$

から,

$$\mathbf{v}_i = \frac{\sum_{k=1}^n u_{ik} \mathbf{x}_k}{\sum_{k=1}^n u_{ik}} \quad (29)$$

となる. したがって, 計算アルゴリズムは次のような, 式 (21), 式 (25), 式 (27), 式 (29) の繰り返しとなる. 本提案アルゴリズムを KFCM 法と呼ぶ.

KFCM アルゴリズム

Step 1: メンバシップの初期値 u_{ik} を乱数により与える.

Step 2: 式 (29) よりクラスタ中心 \mathbf{v}_i を求める.

Step 3: 式 (27) より行列 A_i を求める.

Step 4: 式 (25) よりクラスタ i の容量 π_i を求める.

Step 5: 式 (21) よりメンバシップ u_{ik} を求める.

Step 6: 収束判定条件

$$\max_{c, j} |u_{cj}^{NEW} - u_{cj}^{OLD}| < \epsilon$$

を満たせば終了. その他は *Step 2* へ戻る.

目的関数に K-L 情報量による正則化を導入し, ファジクラスタを得るための重みパラメータ λ を 2 とすることによって, ガウス混合モデルでの式 (5)~式 (8) のアルゴリズムが得られる. したがって, ガウス混合モデルと KFCM 法は分布 (確率密度) を求めるのか集合 (メンバシップ) を求めるのかの相違はあるが密接な関係にあると言える. FCM 法では行列 A_i は単位行列とされることが多いが, D.E.Gustafson と W.C.Kessel による修正 FCM 法 [7] では A_i の要素も繰り返しアルゴリズムの中で求められる.

式 (21) のメンバシップ関数は, エントロピー正則化の FCM 法に比べて π_i が分母分子にかかっている. 式 (25) より, π_i はクラスタ i の容量を表していると考え

られ, 式 (17) の目的関数は未知数 π_i についても最小化されるので, K-L 情報量による正則化ではクラスタ容量についても最適化を図っているといえる. ガウス混合モデルと同様, KFCM 法でもしばしば局所解に陥り, A_i も決定変数とすると, A_i が正定値行列とならない場合が起こりうる. この点のアニールングによる改善については 5.2 節の数値実験で述べる.

3 ガウス混合モデルと KFCM 法との比較

K-L 情報量による正則化を付加した目的関数式 (17) から求まるメンバシップ u_{ik} には, 式 (21) のように変数 λ が含まれている. $\lambda = 2$ の場合は u_{ik} , π_i , A_i , \mathbf{v}_i のすべての未知パラメータを求めるアルゴリズムが, ガウス混合モデルの場合と一致している. また, λ は式 (21) の u_{ik} にのみ含まれている. そこで, $\lambda \neq 2$ の場合に相当するガウス分布以外の何らかの混合モデルが存在するかを考察する.

式 (21) は事前確率から事後確率を求めるベイズの定理に相当する. そこで, E ステップによって事後確率を表す式 (21) が得られる次の関数を考える.

$$p_i(\mathbf{x}) = \frac{|A_i^{-1}|^{1/\lambda}}{K} \exp\left(-\frac{1}{\lambda} d_{ik}\right) \quad (30)$$

この関数が確率密度関数であるかを以下に議論する. ただし, K は定数である. K を適当に定めることによって, 式 (30) の積分が 1 となれば, 何らかの確率密度関数であり, E ステップの式 (21) をベイズの定理から定めることができる. ここでは説明を分かりやすくするために, 2 変数の密度関数の場合について考える. 式 (30) の関数の全領域での積分値は,

$$\begin{aligned} & \frac{|A_i^{-1}|^{1/\lambda}}{K} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{\lambda}(\mathbf{x} - \mathbf{v}_i)^T A_i^{-1}(\mathbf{x} - \mathbf{v}_i)\right) d\mathbf{x} \\ &= \frac{\lambda\pi}{K} |A_i|^{\frac{\lambda-2}{2\lambda}} \end{aligned} \quad (31)$$

となる. $K = \lambda\pi$ で $\lambda = 2$ のときは正規分布に等しく積分は 1 となる. しかし, $\lambda \neq 2$ のとき積分値が 1 となるためには, $K = \lambda\pi |A_i|^{\frac{\lambda-2}{2\lambda}}$ でなければならず,

K は $|A_i|$ に依存するので、すべての i について定数とすることができるのは、すべての i について $|A_i|$ が一定値のときのみである。ベイズの定理を用いて、式 (21) を E ステップから求めることができる密度関数は式 (30) の形に限られる。すなわち K は定数でなければならないので、 $|A_i|$ が定数でなければその積分値が常に 1 になるような確率密度関数は存在しない。したがって、 $\lambda \neq 2$ の場合は対応する混合分布モデルは存在せず、提案法はファジィクラスタリング固有のものであるといえる。2 変量の場合の分布について述べたが、一般に p 変量に拡張することができる。

KFCM 法の目的関数式 (17) の第 2 項目は各 k について u_{ik} と π_i の離散分布 ($i = 1, \dots, c$) 間の離れ具合を表す K-L 情報量である。したがって、第 2 項目の値は非負であり、最小値は 0 である。また、最小値をとるのは $u_{ik} = \pi_i$ ($i = 1, \dots, c$) の場合のみである。このことから、 λ を大きくすると、すべての k について u_{ik} は π_i に近い値として求まると考えられる。すなわち、すべてのデータの第 i クラスタに対するメンバシップ値が $u_{ik} \simeq \pi_i$ となり、あいまいさの大きなクラスタリング結果が得られる。 λ が 0 の場合は、 u_{ik} について線形制約の下での線形最適化問題になり、 u_{ik} は端点、すなわち 0 または 1 として求まる。このように、式 (9) の m と同様に、式 (17) の λ を変化させることでハードやファジィなクラスタリングを行なうことができる。

ガウス混合モデルと異なり、KFCM 法では係数 λ を自由に定めることができるために、 λ の値を反復アルゴリズムの中で小さくしていく、いわゆる、アニーリングを行なうことが可能である。この点においては 5.2 節の数値実験で示す。

$\sum_{k=1}^n u_{ik}$ は、ファジィクラスタ i に含まれるデータ件数に相当している。提案アルゴリズムでは、 $J_{\lambda\tau}$ の最適性の必要条件から、 A_i は式 (27) のようにサンプルデータのファジィ分散共分散行列として求まる。今、行列 A_i は m 個の 1 次独立な固有ベクトル、すなわち主成分ベクトル (p_1, p_2, \dots, p_m) を持つとすると、

$$\log |A_i| = \log \prod_{j=1}^m \delta_j^2 = \sum_{j=1}^m \log \delta_j^2 \quad (32)$$

である。ただし、 δ_j^2 は A_i の固有値である。したがっ

て、 $\log |A_i|$ はクラスタ内のデータの各主成分の分散の対数の和であり、クラスタの広がりを表している。また、 $J_{\lambda\tau}$ の第 3 項目 $\sum_{k=1}^n u_{ik} \log |A_i|$ は、クラスタ毎のデータ件数で重み付けたクラスタの広がりを最小化するように A_i を求めることを意味していて、式 (10) のマハラノビス距離 d_{ik} は A_i^{-1} で定義されている。ここで、 $J_{\lambda\tau}$ の第 3 項目に重み係数 φ を付加すると、最適性の必要条件より導出されるメンバシップ値 u_{ik} に φ が含まれるが、 λ/φ を新たな λ とおくことにより、 φ を 1 とした場合に等しくなる。したがって、重み係数の一方を固定してもう一方をパラメータとすればよい。そこで今回提案する KFCM 法では係数 λ を第 2 項目に付加し、第 3 項目は重み係数を 1 としている。

4 擬似的な可能性的クラスタリング

前章では、KFCM 法がガウス混合モデルに対する EM アルゴリズムを包含していることを示した。本章では、目的関数の第 3 項目に D.E.Gustafson と W.C.Kessel による $|A_i|$ の大きさを制約する方法 [1, 7] を採用した KFCM 法 (GKFCM 法) を提案し、可能性的クラスタリング [8, 9] に類似したファジィクラスタリングが実現できることを示す。D.E.Gustafson と W.C.Kessel は最適なクラスタの形状を得るために、マハラノビス距離を定める行列 A_i を導入し、その最適値を導出するために、行列 A_i に制約を設けている [7]。Gustafson-Kessel の手法では、初めに $|A_i|$ の大きさ ρ_i を指定する必要がある。本章では、KFCM 法の目的関数に $c+1$ 番目のクラスタに対して Gustafson-Kessel の制約項を入れた目的関数、

$$\begin{aligned} J_{\lambda\tau} = & \sum_{i=1}^{c+1} \sum_{k=1}^n u_{ik} d_{ik} + \lambda \sum_{i=1}^{c+1} \sum_{k=1}^n u_{ik} \log \frac{u_{ik}}{\pi_i} \\ & + \sum_{i=1}^c \sum_{k=1}^n u_{ik} \log |A_i| \\ & + \gamma \left(\sum_{k=1}^n u_{c+1k} \log |A_{c+1}| - \rho \right) \\ & - \sum_{k=1}^n \eta_k \left(\sum_{i=1}^{c+1} u_{ik} - 1 \right) - \tau \left(\sum_{i=1}^{c+1} \pi_i - 1 \right) \end{aligned} \quad (33)$$

を提案する．D.E.Gustafson と W.C.Kessel による制約条件 $|A_i| = \rho, i = 1, \dots, c+1$ を付加する方法は宮本ら [8] によって提案されているが，式 (33) では制約条件を $\sum_{k=1}^n u_{c+1k} \log |A_{c+1}| = \rho$ と変更している． γ はラグランジュ乗数を表しており， ρ は正の定数である．

未知パラメータは最適性の必要条件より，

$$\mathbf{v}_i = \frac{\sum_{k=1}^n u_{ik} \mathbf{x}_k}{\sum_{k=1}^n u_{ik}} \quad (34)$$

と求まる．また，

$$W = \sum_{j=1}^c \pi_j \exp\left(-\frac{1}{\lambda} d_{jk}\right) |A_j|^{-1/\lambda} + \pi_{c+1} \exp\left(-\frac{1}{\lambda} d_{c+1k}\right) |A_{c+1}|^{-\gamma/\lambda} \quad (35)$$

とおくと， $i \leq c$ のとき，

$$u_{ik} = \pi_i \exp\left(-\frac{1}{\lambda} d_{ik}\right) |A_i|^{-1/\lambda} / W \quad (36)$$

$i = c+1$ のとき，

$$u_{ik} = \pi_{c+1} \exp\left(-\frac{1}{\lambda} d_{c+1k}\right) |A_{c+1}|^{-\gamma/\lambda} / W \quad (37)$$

$$\pi_i = \frac{\sum_{k=1}^n u_{ik}}{\sum_{j=1}^c \sum_{k=1}^n u_{jk}} = \frac{1}{n} \sum_{k=1}^n u_{ik} \quad (38)$$

とそれぞれ求まり，行列 A_i は， $i \leq c$ のとき，

$$A_i = \frac{\sum_{k=1}^n u_{ik} (\mathbf{x}_k - \mathbf{v}_i)(\mathbf{x}_k - \mathbf{v}_i)^T}{\sum_{k=1}^n u_{ik}} \quad (39)$$

$i = c+1$ のとき，

$S_i = \sum_{k=1}^n u_{ik} (\mathbf{x}_k - \mathbf{v}_i)(\mathbf{x}_k - \mathbf{v}_i)^T$ とおくと，

$$A_{c+1} = \frac{S_{c+1}}{|S_{c+1}|^{1/p}} \{\exp(\rho/n\pi_{c+1})\}^{1/p} \quad (40)$$

と求まる．式 (40) の $\{\exp(\rho/n\pi_{c+1})\}^{1/p}$ は，変数グループ A_{c+1} を最適化する繰り返しアルゴリズムの中では定数である．また，目的関数式 (33) の第 2 項目のパラメータ λ は，前述の議論からクラスタのファジィ度を制御するパラメータとみなすことができる．

クラスタリングアルゴリズムは KFCM 法のアルゴリズムにおいて，メンバシップを式 (35)～式 (37) で求め， A_{c+1} を式 (40) で求める．本提案アルゴリズムを GKFCM 法と呼ぶ．

ファジィクラスタリングのための制約条件には，確率的制約と可能性的制約 [8, 9] がある．これらはそれぞれ

$$M_{prb} = \{u_{ik} | \sum_{i=1}^c u_{ik} = 1, u_{ik} \geq 0, \forall i, k\} \quad (41)$$

$$M_{pos} = \{u_{ik} | \sum_{i=1}^c u_{ik} > 0, u_{ik} \geq 0, \forall i, k\} \quad (42)$$

で与えられている． M_{pos} では M_{prb} の u_{ik} の和に関する制約が外されている．そのことから，ノイズが広範囲に，一様に含まれたデータからクラスタを抽出することができる [8]．5.3 節の数値実験に示すように，式 (33) の目的関数から導かれた GKFCM 法では M_{pos} に類似したクラスタリング結果を得ることができる．

5 数値実験

数値実験では，本研究で作成した KFCM Simulator (図 1) を用いた．本章では，KFCM 法とガウス混合モデルにおける EM アルゴリズムとの比較，および，KFCM 法と目的関数の第 3 項目に Gustafson-Kessel の制約項を付加した GKFCM 法でのクラスタリング結果を比較する．

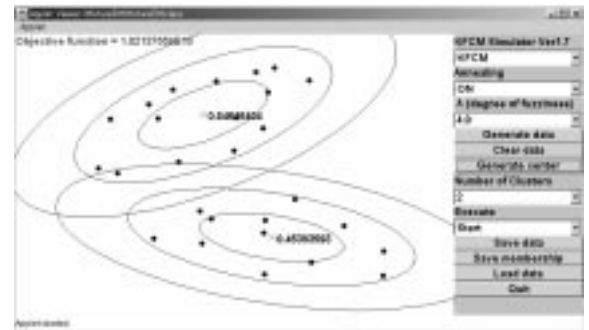


図 1: KFCM シミュレータ

5.1 ガウス混合モデルと KFCM 法の比較

KFCM 法では λ が大きい場合、広域のデータを含んだクラスタを形成する傾向がある。図 2～図 5 に分類結果を示す。ガウス混合モデル (KFCM 法では $\lambda = 2$ の場合に相当する) による結果 (図 2, 図 3) と KFCM 法の $\lambda = 4$ での結果 (図 4, 図 5) では、図中右上のデータが密集しているところを抽出するには、KFCM 法が有効であることが分かる。図 2, 図 4 の等値線はマハラノビス距離を、図 3 と図 5 は分類関数 [5] を表示したものである。

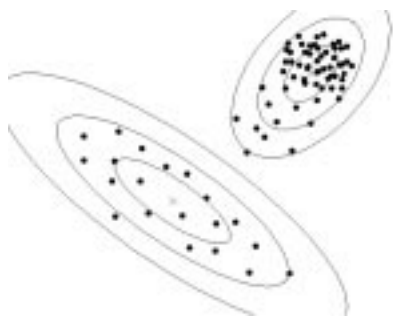


図 2: ガウス混合モデルでの結果

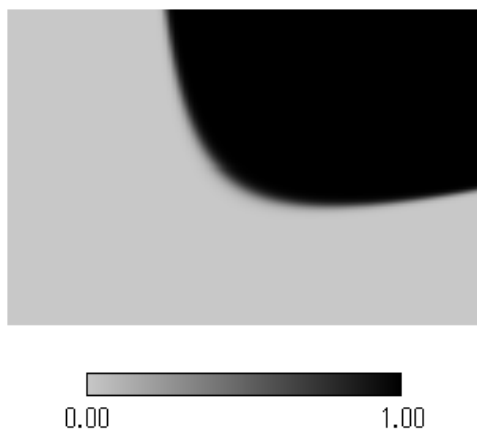


図 3: ガウス混合モデルの事後確率 (分類関数)

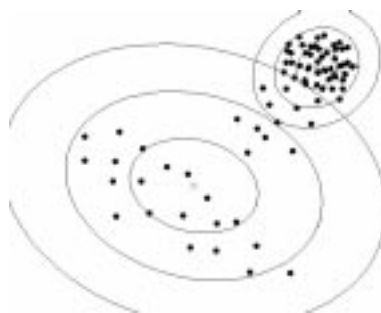


図 4: KFCM 法 ($\lambda = 4$) での結果

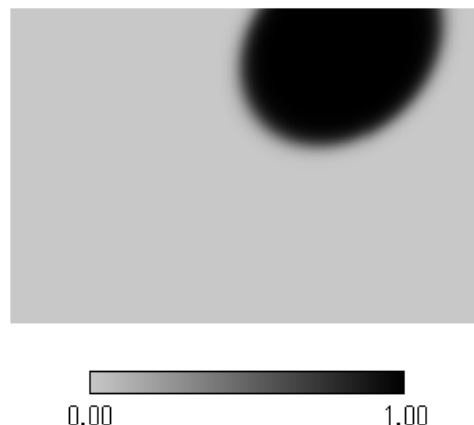


図 5: KFCM 法 ($\lambda = 4$) での第 2 クラスタのメンバシップ (分類関数)

5.2 局所解の比較

次に λ の値を温度とみなして、反復アルゴリズムの中で減少させる、アニーリングの効果を調べた。クラスタ数を 2, KFCM 法の重み係数のアニーリングスケジュールとして、 λ^* を 8 とし、S.Geman と D.Geman[13] の冷却スケジュールをまねた、

$$\lambda(t) = \lambda^* / \ln(2 + t) \quad (43)$$

を使用した。ただし t は繰り返し回数である。また、ガウス混合モデルの結果との比較のために、 λ は 2 になれば固定するように設定した。すなわち、反復回数が 53 回目で $\lambda(53) = 1.99$ になり、53 回目以降は、 $\lambda = 2$ とした。図 6 はアニーリングスケジュールによるパラメータ λ の減少変化を示している。

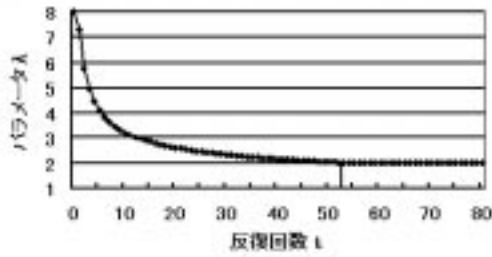


図 6: アニールスケジュール

図 7 に示すデータパターン 1 はデータ数が 10 で、得られる種々の分類結果において、目的関数値にあまり変化がないデータ配置になっている。図 7 の等値線はその一つの分類結果を示している。クラスタ中心の初期値は乱数により与えた。なお、試行回数は 20 回とした。ガウス混合モデルでの結果を表 1、アニールを行なった KFCM 法の結果を表 2 に示す。表中の度数は、同一の分類結果が得られた回数を表している。分類結果 1 から 7 はそれぞれ異なる分類結果に収束したことを示している。KFCM 法でアニールを行なった場合はガウス混合モデルに比べて、収束結果が 4 通りと少ない。しかし、収束までに要した平均反復回数はアニールを行なっているために増加している。KFCM 法でアニールを行なった場合でも局所解に陥るが、比較的良好な解に収束することが分かる。

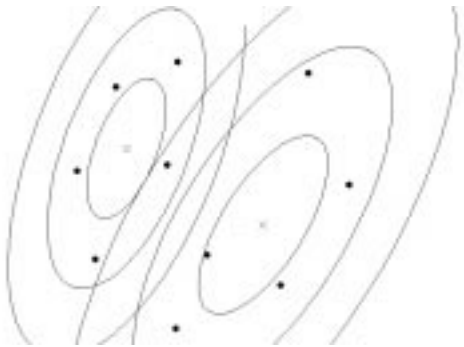


図 7: データパターン 1

表 1: ガウス混合モデルでの結果

分類結果	平均反復回数	度数	目的関数値
1	32	5	4.22×10^9
2	29	5	4.24×10^9
3	34	2	4.66×10^9
4	52	4	4.57×10^9
5	25	1	4.17×10^9
6	25	2	3.81×10^9
7	25	1	3.43×10^9

表 2: KFCM 法での結果

分類結果	平均反復回数	度数	目的関数値
1	74	14	4.22×10^9
5	59	3	4.17×10^9
6	63	2	3.81×10^9
7	60	1	3.43×10^9

同様に図 8 のデータパターンを用いて、各手法で最適解に収束する回数を調べた。ただし、100 回の試行の中で最も目的関数の値の小さいものを最適解とした。試行回数を 100 とし、その他のアニールスケジュールやパラメータの条件は前例と同じとした。Gustafson-Kessel の制約 $|A_i| = \rho_i$ をすべてのクラスタに付加する場合の FCM 法を G-K と表記する。GMM はガウス混合モデルの場合で KFCM 法で $\lambda = 2$ としたものに等しい。表 3 より最適解に収束する割合は、KFCM 法でアニールを行なう場合が他の手法と比べてかなり高い。KFCM 法ではアニールを行なうことで初期値に依存せずにほとんどの場合、最適な結果に収束することが分かる。この例ではガウス混合モデルとの結果の違いが顕著である。また、図 9 のデータパターンでの最適解に収束した度数を表 4 に示す。このデータパターンはいずれの手法でも最適解に収束しやすいが、KFCM 法ではほとんどが最適解に収束している。また、ガウス混合モデルと KFCM 法のアニールスケジュールとして、 λ を一定値ずつ減少させて 0 に近づけ、ハードクラスタリングをおこなった結果と比較した場合も違いは顕著であった。結果は省略する。

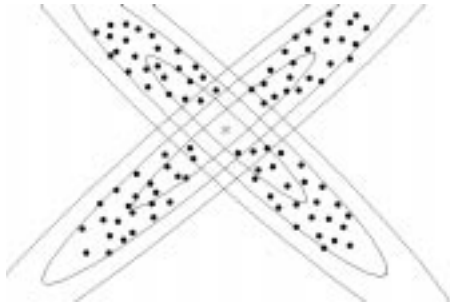


図 8: データパターン 2

表 3: データパターン 2 での結果

手法	度数	目的関数値
GMM ($\lambda = 2$)	11	3.89×10^{10}
KFCM (Annealing)	75	3.89×10^{10}
G-K (Annealing)	7	3.85×10^{10}

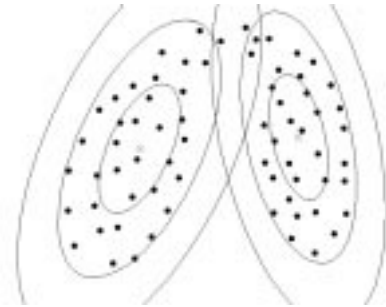


図 9: データパターン 3

表 4: データパターン 3 での結果

手法	度数	目的関数値
GMM ($\lambda = 2$)	70	2.05×10^{10}
KFCM (Annealing)	99	2.05×10^{10}
G-K (Annealing)	86	2.04×10^{10}

5.3 GKFCM 法での分類結果の比較

Gustafson-Kessel の制約項を 1 つのクラスタにのみ付加した，GKFCM 法では可能性的制約 (M_{pos}) を用いるクラスタリングに類似した分類結果を得ることができる．ほぼ一様にデータが配置されているが，2 箇

所にやや密集しているデータ (図 10) を用いた．クラスタ数は 3 とした．図 10～図 12 はガウス混合モデルでのクラスタリング結果，図 13～図 15 は GKFCM 法でのクラスタリング結果である．ガウス混合モデルの結果と比較して，GKFCM 法ではデータの密集しているところにクラスタが形成されている．このような分類結果は，メンバシップの和を 1 とする FCM 法での制約条件を除いた，可能性的クラスタリング [8] で得られているものに類似していて，広く分布しているノイズデータに影響されない分類結果が得られる．

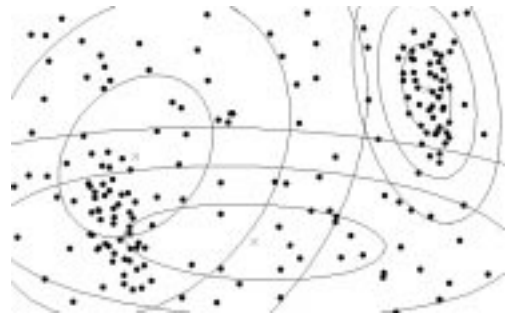


図 10: ガウス混合モデルでの結果



図 11: ガウス混合モデルでの結果 ($c=1$) (分類関数)

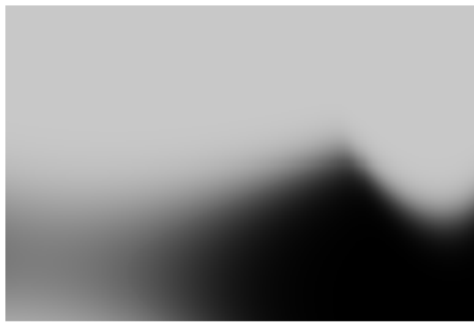


図 12: ガウス混合モデルでの結果 ($c=3$) (分類関数)

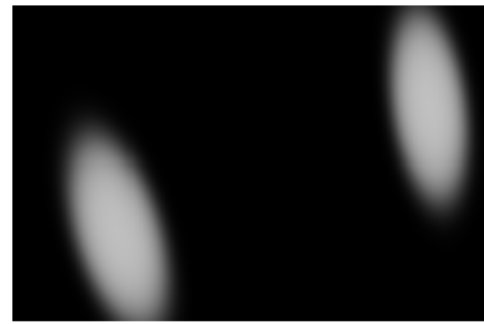


図 15: GKFCM 法での結果 ($c=3$) (分類関数)

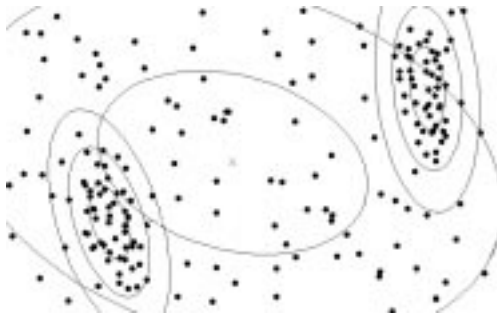


図 13: GKFCM 法での結果



図 14: GKFCM 法での結果 ($c=1$) (分類関数)

6 おわりに

本研究では，K-L 情報量正則化による FCM 法 (KFCM 法) を提案し，ガウス混合モデルにおける EM

アルゴリズムとの類似点と相違点を議論し，KFCM 法がガウス混合モデルとは異なる固有の方法であることを示した．

ガウス混合モデルと KFCM 法の相違点は，

- ガウス混合モデルでは尤度関数最大化であるのに対し，KFCM 法では目的関数最小化であるので種々の制約条件を導入しやすい．
- KFCM 法では，ファジィ度を制御する係数 λ の値を変えることで，ファジィやクリスプなクラスタが得られる．(ガウス混合モデルでは $\lambda = 2$ に固定)
- KFCM 法はファジィ度を制御するパラメータ λ のアニーリングが可能で，パラメータの初期値に依存しにくい分類結果が得られる．
- λ を大きく設定すると，ノイズのある広いデータパターンに対して，データの密集部分を抽出するようにクラスタを形成する．

である．

また，ノイズの影響を受けにくい可能性的クラスタリングを実現できる GKFCM 法を提案し，ガウス混合モデルに比べデータの密集部分を抽出できることを示した．クラスタリングと多変量解析の同時分析法 [14–16] などのデータ解析やデータマイニングへの応用が

今後の課題である．

謝 辞

本研究の一部は文部科学省科学研究費補助金・基盤研究 (C)(13680375) の助成に基づいて行われたものであり，謝意を表する．

参考文献

- [1] J.C.Bezdek: *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, (1981)
- [2] R.L.Streit and T.E.Luginbuhl: Maximum likelihood training of probabilistic neural networks, *IEEE Transactions on Neural Networks*, Vol.5, No.5, pp.764-783, (1994)
- [3] 田中雅博: EM アルゴリズムの原理と応用, システム/制御/情報, Vol.42, No.2, pp.88-96, (1998)
- [4] 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)
- [5] 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.548-557, (1998)
- [6] 赤穂昭太郎: EM アルゴリズム - クラスタリングへの適用と最近の発展 -, 日本ファジィ学会誌, Vol.12, No.5, pp.594-602, (2000)
- [7] D.E.Gustafson and W.C.Kessel: Fuzzy clustering with a fuzzy covariance matrix, *Proc. IEEE CDC*, Vol.2, pp.761-766, (1979)
- [8] 渋谷和宏, 宮本定明, 高田治: ファジィ *c*-平均クラスタリングにおける制約条件と正則化; 第 10 回インテリジェントシステムシンポジウム, 東京, pp.365-368, (2000)
- [9] F.Höppner, F.Klawonn, R.Kruse, T.Runkler: *Fuzzy Cluster Analysis, (Methods for Classification, Data Analysis and Image Recognition)*, John Wiley & Sons,Ltd, (1999)
- [10] 宮本定明: クラスタ分析入門, ファジィクラスタリングの理論と応用, 森北出版, (1999)
- [11] J.C.Bezdek: *Fuzzy Mathematics in Pattern Classification, Ph.D. Thesis*, Applied Math. Center, Cornell University, (1973)
- [12] 宮岸聖高, 安富善彦, 市橋秀友, 本多克宏: KL 情報量正則化によるファジィクラスタリング; 第 16 回ファジィシステムシンポジウム, 秋田, pp.549-550, (2000)
- [13] 坂和正敏, 田中雅博: ニューロコンピューティング入門, 森北出版, (1996)
- [14] 金海好彦, 山川あす香, 市橋秀友, 三好哲也: クラスタリングと数量化分析 3 類の同時分析法, システム制御情報学会論文誌, Vol.12, No.3, pp.47-54, (1999)

- [15] 本多克宏, 山川あす香, 市橋秀友, 三好哲也, 奥山哲史: ファジィクラスタリングと回帰と主成分の同時分析法, システム制御情報学会論文誌, Vol.13, No.5, pp.236-243, (2000)
- [16] 山川あす香, 市橋秀友, 三好哲也: 正準相関係数を最大とするファジィ *c*-Means クラスタリング法, 日本経営工学会論文誌, Vol.51, pp.17-26, (2000)

【 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・情報系経営工学分野

市橋 秀友

TEL : 0722-54-9352

FAX : 0722-54-9915

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

著者略歴

宮岸 聖高 (みやぎし きよたか)

2001 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, 同年日本電気(株)入社, 現在に至る. ファジィクラスタリング, ファジィモデリング, ニューラルネットワークに興味を持つ.

市橋 秀友 (いちはし ひでとも) [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, 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のデータ解析法, その知的システムや人間機械システムへの応用研究に従事. IEEE, 日本ファジィ学会, 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.

本多 克宏 (ほんだ かつひろ) [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，現
在に至る．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．

「訂正」

式 (33) ~ 式 (40) に著者の責任による記載誤りがありましたので，以下のように訂正させていただきます．

$$\begin{aligned} J_{\lambda\tau} = & \sum_{i=1}^{c+1} \sum_{k=1}^n u_{ik} d_{ik} + \lambda \sum_{i=1}^{c+1} \sum_{k=1}^n u_{ik} \log \frac{u_{ik}}{\pi_i} + \sum_{i=1}^{c+1} \sum_{k=1}^n u_{ik} \log |A_i| \\ & + \gamma (\log |A_{c+1}| - \rho) - \sum_{k=1}^n \eta_k \left(\sum_{i=1}^{c+1} u_{ik} - 1 \right) - \tau \left(\sum_{i=1}^{c+1} \pi_i - 1 \right) \end{aligned} \quad (33)$$

を提案する．D.E.Gustafson と W.C.Kessel による制約条件 $|A_i| = \rho_i, i = 1, \dots, c+1$ を付加する方法は宮本ら [8] によって提案されているが，式 (33) では制約条件を $\log |A_{c+1}| = \rho$ のみに変更している． γ はラグランジュ乗数を表しており， ρ は正の定数である．未知パラメータ v_i, u_{ik}, π_j は最適性の必要条件より， c を $c+1$ として KFCM 法と同様に求まる．行列 A_i は， $i \leq c$ と $i = c+1$ のとき，それぞれ

$$A_i = \frac{\sum_{k=1}^n u_{ik} (\mathbf{x}_k - \mathbf{v}_i)(\mathbf{x}_k - \mathbf{v}_i)^T}{\sum_{k=1}^n u_{ik}} \quad (39)$$

$$A_{c+1} = \frac{S_{c+1}}{|S_{c+1}|^{1/p}} e^{\rho/p} \quad (40)$$

ただし， $S_{c+1} = \sum_{k=1}^n u_{c+1,k} (\mathbf{x}_k - \mathbf{v}_{c+1})(\mathbf{x}_k - \mathbf{v}_{c+1})^T$ と求まる．式 (33) の第 2 項目のパラメータ λ は，前述の議論からクラスタのファジィ度を制御するパラメータとみなすことができる．クラスタリングアルゴリズムは KFCM 法と同様の不動点反復法とする．本提案アルゴリズムを GKFCM 法と呼ぶ．