Title	FCM 融合型多目的クラスタリング(モデリングと最適化の理論)
Author(s)	春名, 亮; 石井, 博昭
Citation	数理解析研究所講究録 (2006), 1526: 56-60
Issue Date	2006-12
URL	http://hdl.handle.net/2433/58877
Right	
Type	Departmental Bulletin Paper
Textversion	publisher

数理解析研究所講究録
1526巻 2006年 56-60

FCM融合型多目的クラスタリング

大阪大学大学院情報科学研究科情報数理学専攻　春名　亮*　Ryo HARUNA
大阪大学大学院情報科学研究科情報数理学専攻　石井　博昭　Hiroaki ISHII
Depertment of Information and Physical Science,
Graduate School of Information Science and Technology, Osaka University

1　はじめに

一般的に数理的意思決定問題の大半は単一の評価基準を利用することが非常に多く，それはクラスタリングについても同様の状況であるといえる．特に，クラスタリングにおいては，単一の評価基準としてデータと中心との距離の最小化を採用することは非常に典型的な方法である．従来のクラスタリングアルゴリズムは，基本的なクラスターの様々な形に従わないかもしれない唯一の評価基準が利用されている [1]．しかし，実際には本質的に複数の評価基準を持つことが多いと考えられ，一元的な価値基準で評価することができる場合のほうがむしろ稀である．そこで，クラスタリングにおいても同時に多数のクラスタリング評価関数を用いることが可能な新しいクラスタリングの方法を考えることが必要とされる．そのアプローチの一要素として，クラスターを再抽出する技法を用いて多数のクラスターの効用を評価するクラスターの適合度関数を用いている．

複数の評価基準をもつ場合として，クラスター中心と各点との偏差を利用するクラスターのコンパクト性に関する評価基準，および近傍データ同士がどの程度同じクラスターグループに属しているかを測定するデータ点の連結性に関する評価基準が用いられた．また，多目的クラスタリングでもクラスター内のデータの所属度について検討する必要があるといえる．Wang and Wu[2] は，ファジィc-平均法 (FCM：Fuzzy c-means) を2目的最適化クラスタリングに変更する方法を提案したり，M.Sato and Y.Sato[3] らは状況に依存したデータのファジィクラスタリングを多目的最適化として定式化を行い，そのパレート最適解を考察する方法を提案している．

我々は，新しい方法としてクラスターの選択方法を導入するが，0-1整数計画問題として定式化されるので整数緩和も兼ねてクラスター選択をファジィ化する．さらにクラスター内のデータの所属にあいまいさを導入した多目的性を伴うファジィクラスタリングモデルを定式化し，パレート最適解法に依らない非線形最小化を行う方法を提案する．

2　多目的クラスタリングとファジィc-平均法

多目的クラスタリングの目標は，幾つかの異なる目的関数に対応するクラスタリングアルゴリズムを適用することによってデータ集合内でのクラスターを見つけることである．我々は1つの分割へと異なるクラスタリングアルゴリズムの出力を統合するファジィc-平均クラスタリングアルゴリズムを提案する．正確に言えば，異なるクラスタリング目的関数が与えられ，我々はデータ空間の異なる部分に対して適切な目的関数を利用して分割を

求める.

適合度関数に対する1つの可能な候補は，データ集合の再抽出のもとでクラスターの安定性があるものであり，多目的クラスタリングではしばしば発見されたクラスターが互換性のない感覚の中で矛盾する基準に遭遇する.

2.1 問題の状態

我々はデータ集合 $\mathcal{D} = \{\boldsymbol{x}_1, \cdots, \boldsymbol{x}_m\}$ と L 個のクラスタリングアルゴリズムが与えられたと仮定して，各アルゴリズム $\mathcal{A}_i$ は対応する目的関数 f_i を最大にする $\mathcal{D}$ の分割 $P^{(i)}$ へ戻すものである．正式には，

$$P^{(i)} = \{S_1^{(i)}, S_2^{(i)}, \cdots, S_{C_i}^{(i)}\} = \arg\max_{P(\mathcal{D})} f_i(P(\mathcal{D})) \tag{1}$$

である．ここで，$P(\mathcal{D})$ は $\mathcal{D}$ の任意分割を表し，$S_c^{(i)}$ は $P^{(i)}$ における c 番目のクラスターである．$\mathcal{S} \equiv \cup_i P^{(i)}$ を候補のアルゴリズムによって，発生させた全てのクラスターの集まり

$$\mathcal{S} = \{S_1^{(1)}, \cdots, S_{C_1}^{(1)}, \cdots, S_1^{(L)}, \cdots, S_{C_L}^{(L)}\} \tag{2}$$

である．目標は，分割 $\{P^{(i)}\}$ に基づいて「合意形成な」分割 T を見つけることである．換言すると，クラスターの集まりの目標は，T は S におけるクラスターから得られる.

実際に，我々は全ての特徴空間に適用したクラスタリングアルゴリズムによって得られたクラスターの比較および選択をしなければならない．このため，我々はクラスタリングアルゴリズムによって使用される評価基準に外部の付加的な目的関数を導入する必要がある．目的関数を調和する方法においてクラスターの特性を測定する適合度関数 $g_j(\mathcal{C}_i, \mathcal{D})$ とする．適合度関数 $g_j(\mathcal{C}_i, \mathcal{D})$ を比較することによって，我々は異なるデータの部分集合 (クラスター) に対する最も近似的なクラスタリング評価基準 f_i を間接的に採用することができる.

2.2 適合度関数

適合度関数 $g_j(\mathcal{C}_i, \mathcal{D})$ はクラスター $\mathcal{C}_i$ と単に $\mathcal{C}_i$ の代わりに，全体のデータ集合 $\mathcal{D}$ の両方に依存する．なぜなら一般的にクラスターの適合度は全てのデータ点との関連でその重要性を必要とするからである．正当な適合度関数は次の性質を持つべきである.

性質1： クラスターの評価基準 f_i または f_j を最適にするクラスタリングアルゴリズム $\mathcal{A}_j$ と関係を持つべきである．適合度関数 $g_j(\mathcal{C}_i, \mathcal{D})$ の値が大きければ，さらに f_j または同等に $\mathcal{A}_j$ と関係がある.

性質2： 異なるクラスタリング目的関数に対して比較するべきである．換言すると，$g_j(\mathcal{C}_i, \mathcal{D}) > g_l(\mathcal{C}_i, \mathcal{D})$ ならば，クラスター $\mathcal{C}_i$ の特性は f_l よりも f_j と関係がある．例えば，$g_j(\mathcal{C}_i, \mathcal{D}) = g_j(\mathcal{C}_l, \mathcal{D})$ はクラスター $\mathcal{C}_i$ と $\mathcal{C}_l$ は同等に良く，評価基準 f_j と関係があることを意味する.

マーチンらによって，適合度関数はクラスターの安定性を基にするべきだと提案されている [1]．クラスターの安定性はデータの摂動の下でクラスタリング解法における変化を反映し，異なるクラスタリングアルゴリズムが使える．摂動は，置換の有無にかかわらず，再抽出しているデータによって生じる．安定したクラスターはたいてい選ぶに値するものである．なぜなら同じクラスターがデータ集合における軽微な変化に関係なく形成されるからである．安定したクラスターは良い分離の結果またはクラスターの緊密さをもつ．$g_j(\mathcal{C}_i, \mathcal{D})$ を計算するための擬似コードはアルゴリズム 1 によって与えられる．
[アルゴリズム 1]

I) 以下の操作を M 回行う．

- 摂動を与えるデータ集合 $\mathcal{D}'$ を得るための交換の有無に関わらず $\mathcal{D}$ を再抽出
- 入力として $\mathcal{A}'$ を用いて $\mathcal{A}_j$ を実行して $P(\mathcal{D}')$ を求める．
- $P(\mathcal{D}')$ はクラスターの意味に従って $\mathcal{D}/\mathcal{D}'$ におけるデータを分類することによって $P(\mathcal{D}')$ に変換
- $\mathrm{sim}(\mathcal{C}_i, \mathcal{D})$ を計算

II) スコアー l の平均を $g_j(\mathcal{C}_i, \mathcal{D})$ とする．

ここで，$\mathrm{sim}(\mathcal{C}_i, \mathcal{D})$ は任意のデータ分割 $P(\mathcal{D})$ を伴うクラスター $\mathcal{C}_i$ を比較する類似度である．

2.3 クラスターの選択

統合されたクラスター $\mathcal{S} = \{\mathcal{C}_1, \cdots, \mathcal{C}_M\}$ のリスト与えられ，我々は適合度関数を用いて，目標のクラスター集合 $T = \{\mathcal{C}_1^*, \cdots, \mathcal{C}_M^*\}$ を見つける．$\boldsymbol{u}$ はクラスターの選択に関するメンバシップで構成されるベクトルである．ここでもしメンバシップが 1 ならば選択において 1 つのクラスターが T の中で選ばれ，そうでなければ 0 である．集合 T は最適なメンバシップを求めることによって構築される．

w_{ij} は $\mathcal{C}_i$ および $\mathcal{C}_j$ が集合 T 内で選ればれたならば，非対立な性質に反していることによるペナルティーを示しており，行列 W を定義する．従って $\boldsymbol{u}$ に対する非対立な性質に反している全体のペナルティーは 2 次項 $\frac{1}{2}\boldsymbol{u}^\top W\boldsymbol{u}$ として考えることができる．n_{ij} を $\mathcal{C}_i$ および $\mathcal{C}_j$ の両方に入るデータ点のメンバシップ

$$n_{ij} = \left(\sum_k q_{ki}\right) \wedge \left(\sum_k q_{kj}\right) \tag{3}$$

とする．ここで，q_{kc} はクラスター c のデータ集合 $\mathcal{D}$ における $\boldsymbol{x}_k$ のメンバシップである．w_{ij} の 1 つの妥当な定義は，

$$w_{ij} = \frac{n_{ij}}{\max(|\mathcal{C}_i|, |\mathcal{C}_j|)} = \frac{n_{ij}}{\mu} \tag{4}$$

である．直観的に言えば，w_{ij} はまたより大きなクラスターへ割り当てられ，より小さなクラスターからのデータ点の比率を表す．$\xi(\boldsymbol{u})$ は $\boldsymbol{u}$ において現在のクラスターによって

割り当てられずに残っているデータ点と全体の比とする．もし，完全な分布範囲のペナルティーが満たされれば，$\xi(\boldsymbol{u}) = 0$ である．さらに，ファジィc-平均クラスタリング法 [?] の目的関数

$$J_{\mathrm{FCM}}^{\theta} = \sum_{k}\sum_{c}(q_{kc})^{\theta}\|\boldsymbol{x}_k - \boldsymbol{v}_c\|^2 \tag{5}$$

も付加する．$\xi(\boldsymbol{u})$ および $\boldsymbol{u}^{\top}W\boldsymbol{u}$ を最小化したいのであるが，これはクラスターの適合度関数の和を最大化することと同様であるので，ファジィc-平均クラスタリング法の目的関数 [4] も追加し，3個の正の変数 ω_1, ω_2, ω_3 を導入して，最小化目的関数を次のように定義する．

$$J = -\boldsymbol{s}^{\top}\boldsymbol{u} + \omega_1\boldsymbol{u}^{\top}W\boldsymbol{u} + \omega_2\xi(\boldsymbol{u}) + \omega_3 J_{\mathrm{FCM}}^2 \tag{6}$$

ここで，$\boldsymbol{s}$ は適合度関数値で構成されるベクトルであり，$\boldsymbol{s}_i = g_j(C_i, \mathcal{D})$，$j$ は C_i を創生するアルゴリズム $\mathcal{A}_j$ を記す．関数 $\xi(\boldsymbol{u})$ は

$$\xi(\boldsymbol{u}) \approx \frac{1}{m(\boldsymbol{d}^{\top}\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^{\top}N\boldsymbol{u})} \tag{7}$$

として近似され，$\boldsymbol{d}$ は C_i の大きさ d_i を要素とし，N は n_{ij} を要素とする行列である．制約条件

$$\sum_{i} u_i = 1, \quad u_i \in [0,\ 1] \tag{8}$$

$$\sum_{c} q_{kc} = 1, \quad q_{kc} \in [0,\ 1] \tag{9}$$

の下で，(6) 式をラグランジュ未定乗数法により以下の目的関数

$$\begin{aligned} L \ = \ & -\boldsymbol{s}^{\top}\boldsymbol{u} + \omega_1\boldsymbol{u}^{\top}W\boldsymbol{u} + \omega_2\left\{-m\left(\boldsymbol{d}^{\top}\boldsymbol{u} - \frac{1}{2}\boldsymbol{u}^{\top}N\boldsymbol{u}\right)\right\} + \tau(\mathbf{1}^{\top}\boldsymbol{u} - 1) \\ & + \omega_3\sum_{k}\sum_{c}(q_{kc})^2\|\boldsymbol{x}_k - \boldsymbol{v}_c\|^2 + \sum_{k}\eta_k\left(\sum_{c} q_{kc} - 1\right) \end{aligned} \tag{10}$$

を最小化する．ここで，$\xi(\boldsymbol{u})$ を2種類に場合分けして用いることによって，メンバシップの更新式も2種類考えなければならない．ファジィc-平均法を融合した不動点反復アルゴリズム [6, 7] を示す．

Step 1: 以下に示すパラメータの初期値を与える

- C：クラスター数
- ω_1, ω_2, ω_3
- (7) 式における m
- 2つの小さな正数 ε_1，ε_2

2つのメンバシップ $(u_i,\ q_{kc})$，クラスター中心 $\boldsymbol{v}_c$ の初期値は乱数を用いる

Step 2: クラスター中心 $\boldsymbol{v}_c$ を更新する

Step 3: 2つのメンバシップ (u_i, q_{kc}) を更新する

Step 4: もし以下の条件

$$\max |\boldsymbol{u}^{\mathrm{NEW}} - \boldsymbol{u}^{\mathrm{OLD}}| < \varepsilon_1 \tag{11}$$

$$\max |q_{kc}^{\mathrm{NEW}} - q_{kc}^{\mathrm{OLD}}| < \varepsilon_2 \tag{12}$$

を満足すれば終了，そうでなければ **Step 2** へ戻る

3 むすび

我々は，多目的クラスタリングおよびファジィc-平均法を融合したモデルの定式化を示した．従来のクラスタリングの評価基準にクラスターの選択に関する評価基準を付加して，さらにクラスターの選択指標をファジィ化するとともに 0-1 整数条件を緩和したファジィc-平均法を融合した不動点反復アルゴリズムを提案した．

今後の課題として，関数 $\xi(\boldsymbol{u})$ の近似方法の改善や新たな $\mathrm{sim}(\mathcal{C}_i, \mathcal{D})$ の定義などが残される．

参考文献

[1] Martin H.C.Law, Alexander P.Topchy and Anil K.Jain : Muliobjective Data Clustering, To appear in *IEEE Computer Society on Conference on Computer Vision and Pattern Recognition*, 2004.

[2] H.-F. Wang and G.-Y. Wu, Bi-Criteria fuzzy clustering systems, VI IFSA World Congress, Sao Paulo, Brazil, 1995, Vol.1, pp.633-636, 1995

[3] M.Sato and Y.Sato, On a multicriteria fuzzy clustcing method, Fifth IFSA World Congress, Seoul, Korea, 1993, pp.473-776, 1993

[4] J.C.Bezdek : Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press (1981)

[5] Frank Hoppner, Frank Klawonn, Rudolf Kruse, Tomas Runkler, -FUZZY CLUSTER ANALYSIS AND IMAGE RECOGNITION-, WILEY (2000)

[6] A .K. Jain and R. C. Dubes, "Algorithms for clustering data", Prentice Hall. Englewood Cliffs, New Jersey. (1988)

[7] F. Klawonn, E.-P.Klement, Mathematical Analysis of Fuzzy Classifiers. In : X. Liu, P. Cohen, M. Berthold (eds.). Advances in Inteligent Data Analysis. Springer, Berlin (1997), pp.359-370