

## 総合距離尺度を用いたクラスタリングの 従来手法との比較実験

5Q-9

藤澤 肇 加藤 常員 小沢 一雅

大阪電気通信大学

### 1. はじめに

クラスタリングとは、ものを分類、分割する操作で、入れ子状に分類する階層的クラスタリングと単一の分類を行う非階層的クラスタリングに分けられる。クラスタリングでは、分割状態を評価する何らかの尺度の設定をまず行う。目的に合わせて尺度を変えることで様々な分割結果が得られる。

本稿では、視覚的クラスタリングを想定して定義された総合距離尺度<sup>(1)</sup>を用いた分割実験を行う。実験は制御パラメータをひとつもつクラスタリング手法である最大距離法との比較実験である。比較実験で得られた知見をもとに、総合距離尺度により分割されるクラスタの属性について考察を行う。以下、総合距離尺度と実験の概要を示し、実験より明らかになった事項について述べる。

### 2. 総合距離尺度

人間の視覚的あるいは主観的な分割をモデルとしたクラスタ評価尺度を定義する。想定するモデルは、人間の主観的認識が対象を分ける方向で認識しようとする分散作用と固める方向で認識しようとする凝集作用の2つの作用があり、各々の対象について両者のバランスをとることで最終的認識がなされるとするものである。このモデルに対応させて、凝集と分散のそれぞれを表わす評価式を制御パラメータ $\alpha$ でバランスを取った評価式：総合距離尺度 $D$ を式(1)のように定義する。

$$D = \alpha D_i + (1.0 - \alpha) D_o \quad (0.0 \leq \alpha \leq 1.0) \quad (1)$$

$$D_i = \sum_{k=1}^g \sum_{i \in C_k} \sum_{j \in C_k} d_{ij} \quad (2)$$

$$D_o = \sum_{k=1}^g \sum_{i \in C_k} \sum_{j \in C_k} (d_{max} - d_{ij}) \quad (3)$$

ここで、 $d_{ij}$ は個体 $i$ - $j$ 間の距離、 $d_{max}$ は個体間の最長距離、 $g$ はクラスタ数、 $C_k$ は $k$ 番目に属する個体の集合である。 $D_i$ はクラスタ内距離和、同じクラスタに属する点間の距離を総合計したものである。 $D_o$ はクラスタ外距離和、相異なるクラスタに属する点間の距離を点間の最大距離から引いたものを総合計したものである。評価値 $D$ は小さいものほど適した分割あるとする。 $D$ は $\alpha$ を変化させることにより一群を最適とする評価からシングルTONを最適とする評価まで制御することができる(図1参照)。

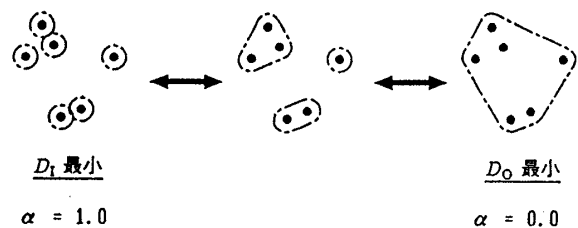


図1 パラメータ $\alpha$ と分割状態

### 3. 分割最適化型手法と最大距離法

4章で述べる実験で用いる手法を紹介する。分割最適化型手法の評価尺度に総合距離尺度を適用する。

- (1) 分割最適化型手法: 評価基準と初期分割を定め、評価値を最大化(あるいは最小化)するように、初期分割を改良して行く手法である。具体的に各個体の所属するクラスタを換え、評価値の改善が見られれば改良を行う。代表的な手法として、*k-means*法が挙げられる。
- (2) 最大距離法: 各クラスタ内の個体間の最大距離の最大がクラスタ代表個体間の最大距離にパラメータ $\theta$ をかけたものより大きいものが存在しないように分割する手法である。クラスタ代表個体は、もっとも離れた2個体から始め、逐次的に代表個体を決定する。

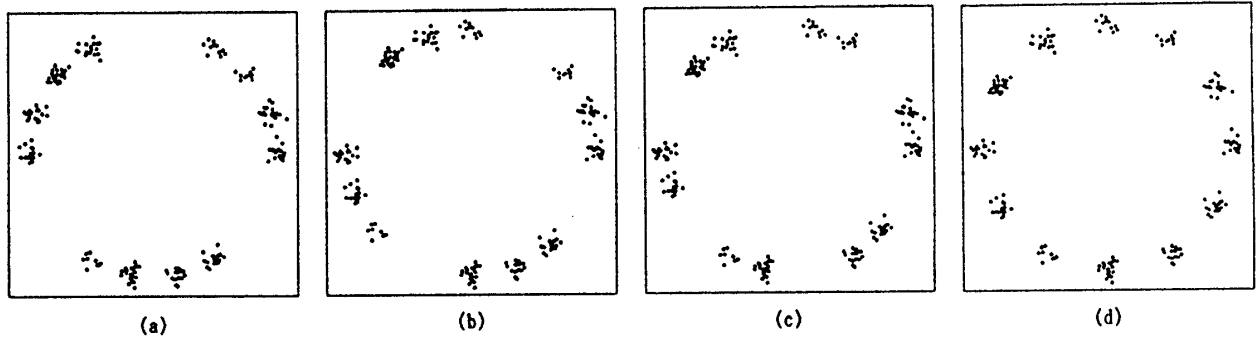


図2 点配置パターン

4. 比較実験

総合距離尺度を用いた分割最適化型手法と最大距離法を用いた分割実験を次のように行った。分割対象は、図2(a)~(d)に示す(0, 1]の正方領域の点配置パターンである。点配置パターンはそれぞれ、1塊が平均16点からなる12個の塊を3、4、6、12分割状になるよう円周上に配置したものである。2つの手法それぞれのパラメータ $\alpha$ 、 $\theta$ を0~1の範囲で0.01刻みで変化させ、分割状態の変化を観測する。

5. 実験結果および考察

図3に図2(c)の6分割パターンの場合の各パラメータの値と分割数の関係のグラフを示す。グラフの横軸はパラメータ値、縦軸は分割数を表わす。

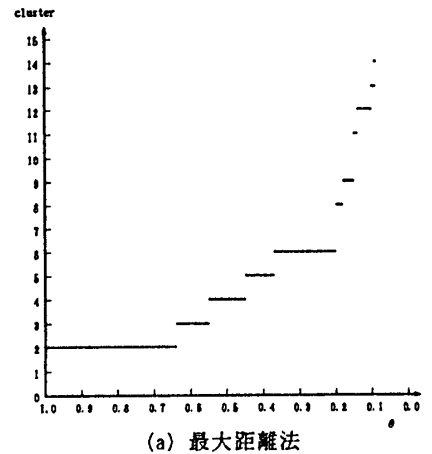
2つのグラフを比較すると、ともに6分割および12分割に対応するパラメータの範囲(幅)は、その前後の分割に比べ広い。この場合の分割結果は、意図した6分割および12分割になっている。図2(c)を6、12分割以外の分割を視覚により行うおうとすると3分割であろう。このときのパラメータの幅は、最大距離法では、3、4、5分割とあまり変わらないが、総合距離尺度では、ピンポイント的なローカルな分割を含むが、2、4、5分割と比べ明らかに幅が広い。また、実際の分割状態は、図4に示すように最大距離法では、もとの12塊を無視した分割がなされているのに対し、総合距離尺度による分割は、塊単位で分割がなされている。このことは、総合距離尺度による分割が、視覚的分割をうまくシミュレートしているものと考えられる。

6. おわりに

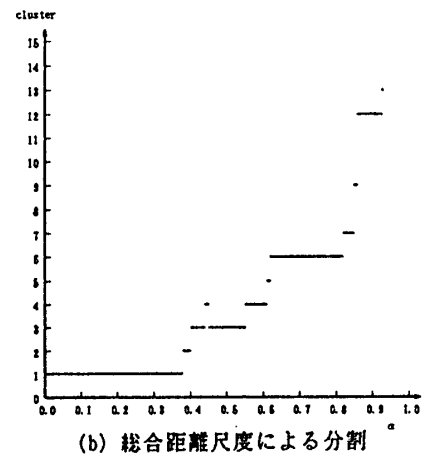
本稿では、総合距離尺度を評価基準としたクラスタ分割の属性を知るための比較実験について述べた。実験結果より、最大距離法よりも総合距離尺度による分割の方が視覚的分割に近い結果を得ることを示した。また、数通りの視覚的分割が可能な場合、総合距離尺度ではパラメータの幅に対応することを示唆した。

参考文献

- (1) 岡田, 加藤, 小沢: 総合距離尺度によるクラスタ分割(CLAIM法), 情報処理学会論文誌, Vol. 35, No. 11, pp. 2291-2299(1994).

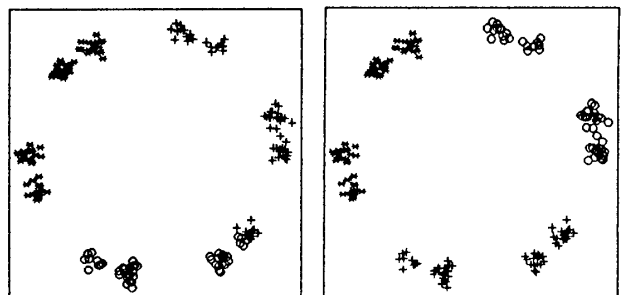


(a) 最大距離法



(b) 総合距離尺度による分割

図3 パラメータ値-分割数のグラフ



(a) 最大距離法

(b) 総合距離尺度による分割

図4 図2(c)の3分割状態