

総合距離尺度を用いた構造的クラスタ分割

2G-4

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、分割する操作である。評価基準を逐次的に適用して入れ子状の分割を得る階層的分類法と評価基準により単一に分割状態を得る非階層的分類法に分けられる。階層的分類法ではクラスタを融合(あるいは分離)するための評価基準が必要であり、非階層的分類法においては、分割状態の評価基準が必要である。これらの評価基準は、目的や対象あるいは効率を考えてさまざまなものが提案されている⁽¹⁾。

本稿では、視覚的に認知される構造を持った対象の分割を想定し、その分割の評価を行う総合距離尺度を用いた分割実験について報告する。めざす総合距離尺度を評価基準とした分割は、単一分割でなく、入れ子状の分割でもない限られた範囲で複数の状態をもつものである。たとえば、2通りの2分割状態があるといった分割である。実験は、階層的分類法、非階層的分類法による分割と、総合距離尺度を用いた分割との比較実験である。比較実験で得られた結果より、総合距離尺度による分割の特性を考察する。

2. 在来のクラスタリング手法

比較実験の対象となる在来の手法について述べる。

分割最適化型手法: 分割最適化型手法は、非階層的分類法として良く用いられる手法である。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する。実験では、分割最適化型手法の代表である *k-means* 法を対象とする。評価基準としてはクラスタ内平方和の和を用いる。

凝集型分類法: 凝集型分類法は、階層的分類法のうち、クラスタを融合していく方法である。クラスタを融合する基準を定め、全クラスタの中で基準にあてはまるクラスタを融合していき、ひとつのクラスタになるまで繰り返す手法であ

A Structure Clustering

Based on the Integrated Distance Measure
Hajime Fujisawa, Tsunekazu Kato, Kazumasa Ozawa
Osaka Electro-Communication University
Neyagawa-shi, Osaka 572, Japan

る。基準の設定によって様々なものがあるが、本実験では組み合わせ的手法の最短距離法、最長距離法、群平均法、加重平均法、重心法、メジアン法、ワード法の7つを対象とする。

3. 総合距離尺度による分割

3.1 総合距離尺度

人間の視覚により認知される構造的な分類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 α 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。

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

$(0.0 \leq \alpha \leq 1.0)$

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

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

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、 D_o はクラスタ外距離和である。評価値 D は小さいものほど適した分割あるとする。

D は α 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できる。また、 α と最適分割との関係は、傾向として α が大きな値であるほど大分割になり、安定(自然)な分割であるほど α の範囲が広がることが実験的に確かめられている⁽²⁾。

3.2 総合距離尺度法

総合距離尺度をクラスタ評価基準にし、階層的分類と非階層的分類の中間的な分割を得る手法、総合距離尺度法を提案する。

総合距離尺度は、 α を1つ決定すると1つの分割状態が決定され、分割を得る方法としては、先に述べた分割最適型手法を適応することができる。 α を1.0から0.0まで逐次的に減少させながら各 α 値で最適となる分割を次の α 値における初期分割として与え、最適分割系列を求める。

4. 比較実験

k-means法、組み合わせの手法および総合距離尺度法による分割を比較する実験を次のように行った。

分割対象は、[0, 1)の正方領域の点配置パターンである。このパターンは、図1に模式図で示すような2、3、5、6分割の視覚的に認知できる塊の構造を持たせて個体(点)を配置したものである。図2に点配置パターンの一例を示す。

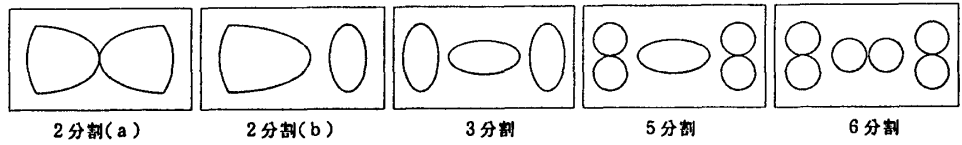


図1 分割状態模式図

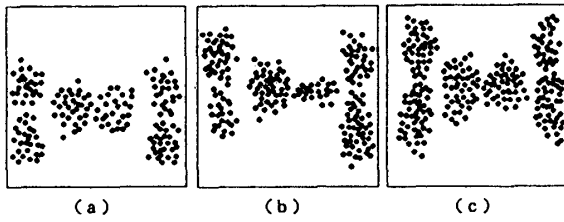


図2 点配置パターン

各手法での2、3、5、6分割状態を出力し、分割が図1に示すような状態になっているかどうかを観測する。

5. 実験結果および考察

図3に図2(a)のパターンについて図1に対応した分割例を示す。表1に実験結果を示す。表1において各手法での分割結果が図2の分割状態ようになったものには○、ならなかったものには×で示した。

表1よりk-means法では2分割(a)のみしか得られない、また組み合わせの手法のいれも2分割は、当然であるが(2aまたは2b)のいずれか一方の分割しか得られていない。これに対し、総合距離尺度法では、(a)のパターンではすべての分割状態が得られているが、(b)や(c)のパターンでは5分割状態が得られていない。しかし、総合距離尺度法は2通りの2分割状態を得ている。このことは、階層的分類と非階層的分類の中間的な働きがあり、同一分割数のものを複数含む分割系列をえることができることを示している。すなわち、総合距離尺度法では、単純入れ子状でない

分割系列を得ることができる。また、各分割が得られる α の範囲は、3分割、5分割では他の分割に比べ狭い結果を得た。これは3分割、5分割が分割し難い分割であることを示していると考えられる。

以上より総合距離尺度は、在来の手法にはない同一分割数での構造おも把握できるクラスタリング手法であるといえる。

6. おわりに

本報告では、階層的な分類と非階層的な分類の両者の特徴をもった総合距離尺度法を提案した。在来手法との比較実験を行い、その結果より在来の手法では把握できない構造を把握することができることを確認した。今後の課題としては、どのような構造が把握できるか、また、その構造あるいは構造の強度とパラメータ α との関係を明らかにするなどが挙げられる。

参考文献

- (1) Anderberg, M. R. (西田英朗 監訳): クラスタ分析とその応用, p. 442, 内田老鶴圃, 東京(1988).
- (2) 岡田, 加藤, 小沢: 総合距離尺度によるクラスタ分割(C L A I M法), 情報処理学会論文誌, Vol. 35, No. 11, pp. 2291-2299(1994).

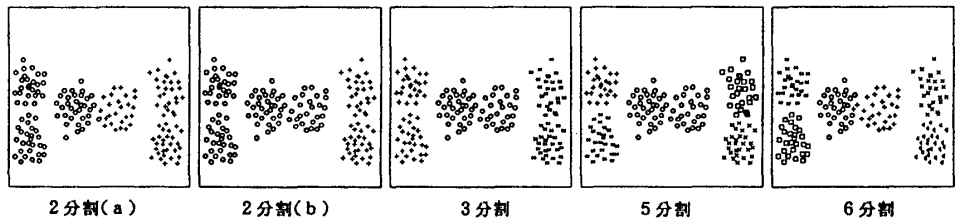


図3 図2(a)パターンの分割例

表1 在来手法および総合距離尺度法による分割結果

パターン No.		a					b					c				
分割数(分割状態)		2a	2b	3	5	6	2a	2b	3	5	6	2a	2b	3	5	6
非階層的分類	k-means法	○	×	×	×	×	○	×	×	×	○	○	×	×	×	×
	最短距離法	×	○	○	×	×	×	○	○	×	×	×	○	○	×	×
階層的分類 (組み合わせの手法)	最長距離法	×	○	○	×	×	○	×	×	×	×	○	×	×	×	×
	群平均法	×	○	○	○	○	×	○	○	○	○	×	○	○	○	○
	加重群平均法	×	○	○	○	○	○	×	×	×	×	×	○	×	○	○
	重心法	×	○	○	×	×	○	×	×	×	×	○	×	×	×	○
	メジアン法	×	×	×	×	×	○	×	×	×	×	○	×	×	×	○
	ワード法	×	○	○	○	○	×	○	○	○	○	×	○	○	○	○
構造的分類	総合距離尺度法	○	○	○	○	○	○	○	○	×	○	○	○	○	×	○