

## 灰色分析を適用した階層的クラスター分析法の提案

山口 大輔<sup>†</sup> 小林 俊裕<sup>†</sup> 水谷 晃三<sup>†</sup> 永井 正武<sup>‡</sup>

クラスター分析は類似度を手がかりにいくつかの集落に分類する方法である。しかし、アルゴリズムが多様であり、適用したアルゴリズムによって分析結果が異なる。そのため、分析者は一度に複数のアルゴリズムを適用しているのが現状である。本論文は灰色分析を基礎理論とし、クラスター分析と同等の機能を持つアルゴリズムを提案する。本提案手法は階層的クラスター分析法であり、分析結果を樹形図で表現する。樹形図にはクラスタリング情報と新たにケースの序列情報を付加する。そのため、樹形図の情報量が増え、分析結果からの解釈の幅を広げる可能性を示している。四つの適用例を示し、本提案手法は広範囲の分野に適用可能なことを示した。

## Proposal of Cluster Analysis Based on Grey Theory

Daisuke Yamaguchi<sup>†</sup> Toshihiro Kobayashi<sup>†</sup> Kozo Mizutani<sup>†</sup> and Masatake Nagai<sup>‡</sup>

Cluster analysis is one of the important methods which is used in many fields. At present, there are many algorithms about cluster analysis. Different algorithms have different results in cluster analysis. For acquiring enough useful results, we need use some algorithms. This paper proposes an algorithm based on grey theory, which has the same function as cluster analysis. In this proposal algorithm, not only the analysis results but also rank information can be increased in dendrogram. The proposal algorithm can be used in many fields. Four examples are given in this paper.

### 1. まえがき

クラスター分析はデータの分類やデータ構造を発見することが目的であり、そのために各データの類似度を評価する分析法である<sup>[1][2]</sup>。クラスター分析は汎用的な統計処理手法のため、今日では生命科学、医学、行動科学、社会科学、地球科学、政治・経済など多岐にわたり利用されている。特に、工学においてもパターン認識、人工知能、システム工学、情報検索、ORなどで使われている。近年のクラスター分析に関する研究は、ファジー理論<sup>[3]</sup>やニューラルネットワーク<sup>[4]</sup>を用いたアルゴリズムが報告されている。しかし、一般的に用いられているのは統計処理手法に含まれるクラスター分析法であり、アルゴリズムも多種多様である。そのため、分析者は一度の分析において複数のアルゴリズムを適用しなければならない。さらに、使用

したクラスター分析アルゴリズムによって結果が異なることがある。

筆者らは、1982年に鄧聚龍によって提案された灰色理論<sup>[5]-[7]</sup>の工学的応用について研究している<sup>[8][9]</sup>。灰色理論は、ある主題に対する情報が不足している状態を数学的に取り扱う理論であり、日本では主に経営工学に利用されている<sup>[6]</sup>。本論文は、灰色理論に含まれる分析法である灰色分析を基礎理論とし、クラスター分析と同等の機能を持つアルゴリズムを提案する。

伝統的なクラスター分析法は以下のステップで処理されている<sup>[2]</sup>。

Step1: データマトリクスの獲得

Step2: データマトリクスの標準化

Step3: 類似マトリクスの計算

Step4: クラスター化の実行

Step5: データマトリクスと類似マトリクスの再配列

Step6: コーフェン相関係数の計算

本提案クラスター分析法は、後述する灰色分析を基礎理論として Step3 および Step4 に相当するアルゴリ

<sup>†</sup> 帝京大学大学院理工学研究科

<sup>‡</sup> 帝京大学理工学部

<sup>†</sup> Graduate school of Science and Engineering, Teikyo University

<sup>‡</sup> School of Science and Engineering, Teikyo University

ズムを開発し、提案している。また、灰色分析の特徴を生かすことで分析結果の情報量が増え、分析者の解釈の幅を広げる可能性を示している。

以下、2.では本提案クラスター分析法のアルゴリズムを構築するための基礎理論である灰色分析について解説し、3.では本提案クラスター分析法のアルゴリズムを示す。4.で本提案分析法の適用例を示し、5.において本提案方法の有効性や評価について考察する。6.はむすびと今後の課題について論述する。

## 2. 灰色関連度分析

灰色関連度分析<sup>[6]</sup>とは、灰色理論に含まれる分析技法であり、単に灰色分析とも呼ぶ。複数の観測値を持つ変量群を数列と定義し、複数の数列間におけるデータの類似度を測る分析法である。分析結果として灰色関連度  $\Gamma$  を $[0,1]$ の実数値として得ることができる。灰色分析は本提案クラスター分析法のアルゴリズムを構築するための基礎理論となっている。以下に灰色分析の計算手順を示す。

灰色分析はデータの類似度を測る基準となる数列、いわゆる基準数列  $x_0$  が必要である。

$$x_0 = \{x_0(1), \dots, x_0(j), \dots, x_0(k)\}, (j=1, 2, \dots, k) \quad (1)$$

そして、基準数列とのデータの類似度合いを測りたい数列を比較数列  $x_i$  とする。原則として、比較数列は  $n$  組用意する。

$$x_i = \{x_i(1), \dots, x_i(j), \dots, x_i(k)\}, (i=1, 2, \dots, n) \quad (2)$$

式(1)、式(2)中の  $x(1) \sim x(k)$  は観測値を持つ変量である。

最初に、基準数列の変量  $x_0(j)$  と比較数列の変量  $x_i(j)$  との差の絶対値  $\Delta_{0i}(j)$  を、次式にて算出する。

$$\Delta_{0i}(j) = |x_0(j) - x_i(j)| \quad (3)$$

次に、灰色関連係数  $\gamma_{0i}(j)$  と呼ばれる値を各  $\Delta_{0i}(j)$  に対して算出する。鄧氏定義の算出式は次式に示す。

$$\begin{aligned} \gamma_{0i} &= \frac{\min_{\forall i} \cdot \min_{\forall j} \{\Delta_{0i}(j)\} + \rho \max_{\forall i} \cdot \max_{\forall j} \{\Delta_{0i}(j)\}}{\Delta_{0i}(j) + \rho \max_{\forall i} \cdot \max_{\forall j} \{\Delta_{0i}(j)\}} \\ &= \frac{\Delta_{\min.} + \rho \Delta_{\max.}}{\Delta_{0i}(j) + \rho \Delta_{\max.}} \end{aligned} \quad (4-1)$$

式(4-1)中の重みを表す  $\rho$  は灰色分析の調節用の係数として、 $\rho = 0.5$  とする。 $\Delta_{\min.} = \min_{\forall i} \cdot \min_{\forall j} \{\Delta_{0i}(j)\}$  は全  $\Delta_{0i}(j)$  の中の最小値であり、 $\Delta_{\max.} = \max_{\forall i} \cdot \max_{\forall j} \{\Delta_{0i}(j)\}$  は全  $\Delta_{0i}(j)$  の中の最大値である。

筆者の一人は式(4-1)について、その物理的意味の吟味性と計算上の利便性などの観点から、次の変形式を使用している<sup>[7]</sup>。

$$\begin{aligned} \gamma_{0i} &= \left[ \frac{(\Delta_{0i}(j) - \Delta_{\min.})(\Delta_{\min.} + \rho \Delta_{\max.})^{-1} + 1}{\Delta_{0i}(j)(\Delta_{\min.} + \rho \Delta_{\max.})^{-1} - \Delta_{\min.}(\Delta_{\min.} + \rho \Delta_{\max.})^{-1} + 1} \right]^1 \\ &= \underbrace{\left[ \frac{(\Delta_{0i}(j) - \Delta_{\min.})(\Delta_{\min.} + \rho \Delta_{\max.})^{-1} + 1}{\Delta_{0i}(j)(\Delta_{\min.} + \rho \Delta_{\max.})^{-1} - \Delta_{\min.}(\Delta_{\min.} + \rho \Delta_{\max.})^{-1} + 1} \right]^1}_{\text{係数部}} \quad (4-2) \end{aligned}$$

次式の処理により、分析結果となる灰色関連度  $\Gamma_{0i}$  を比較数列毎に得ることができる。

$$\Gamma_{0i} = \frac{1}{k} \sum_{j=1}^k \gamma_{0i}(j), (i=1, 2, \dots, n) \quad (5)$$

$\Gamma_{0i}$  は  $x_0, x_i$  間におけるデータの類似度である。 $\Gamma_{0i} \rightarrow 1$  であるほど基準数列  $x_0$  と比較数列  $x_i$  間のデータが類似していることを表す。

## 3. 灰色クラスター分析

本論文で提案するクラスター分析法は階層型分類法に属する。小さいクラスターを合併していくことでより大きいクラスターを構成する方法である。また、構成したいクラスター数は分析結果が出力されるまで不明である。クラスタリング結果は樹形図あるいはデンドログラムと呼ばれる木構造のグラフを出力する。樹形図を作成するために、本提案手法では二つの情報をデータマトリクスから生成する。表1の例を通じて本提案クラスター分析のアルゴリズムを示す。

### 3.1 序列情報の獲得

表1に示すように、本提案クラスター分析ではデータマトリクスとして序列基準数列  $c_0$  および比較数列  $c_i$  を獲得する必要がある。比較数列がクラスタリング対象である。本節で生成する序列情報とは序列基準数列に対する各比較数列の灰色関連度  $\Gamma_{0i}$  のことである。

従来のクラスター分析法で得られる樹形図は、数列の序列ルールとしてデータの類似度または樹形図の見易さを採用している。本提案方法では2.で述べた灰色関連度  $\Gamma_{0i}$  の降順整理を序列ルールとする。表1の序列基準数列は2.で述べた基準数列であり、比較数列は2.における比較数列に当たる。表1を例として序列情報作成方法を示す。2.で述べた灰色分析と同様の処理である。

はじめに式(3)に示す  $\Delta_{0i}(j)$  を求める。表1の例の場合では、

$$\left. \begin{aligned} \Delta_{01}(1) &= 1 - 0.77 = 0.23 & \Delta_{01}(2) &= 1 - 0.77 = 0.23 & \Delta_{01}(3) &= 1 - 0.79 = 0.21 \\ \Delta_{02}(1) &= 1 - 0.42 = 0.58 & \Delta_{02}(2) &= 1 - 0.38 = 0.62 & \Delta_{02}(3) &= 1 - 0.36 = 0.64 \\ \Delta_{03}(1) &= 1 - 0.67 = 0.33 & \Delta_{03}(2) &= 1 - 0.67 = 0.33 & \Delta_{03}(3) &= 1 - 0.71 = 0.29 \\ \Delta_{04}(1) &= 1 - 0.63 = 0.37 & \Delta_{04}(2) &= 1 - 0.65 = 0.35 & \Delta_{04}(3) &= 1 - 0.65 = 0.35 \end{aligned} \right\} (6)$$

となる。

次に、式(4-1)に示す灰色関連係数  $\gamma_{0i}(j)$  を求める。式(4-1)より、灰色関連係数の算出式は、

$$\begin{aligned} \gamma_{0i}(j) &= \frac{0.21 + 0.5 \times 0.64}{\Delta_{0i}(j) + 0.5 \times 0.64} \\ &= \frac{0.53}{\Delta_{0i}(j) + 0.32} \end{aligned} \quad (7)$$

と得られる。ただし、

$$\left. \begin{aligned} \Delta_{\min.} &= \min_{\forall i} \cdot \min_{\forall j} \{ \Delta_{0i}(j) \} = \Delta_{01}(3) = 0.21 \\ \Delta_{\max.} &= \max_{\forall i} \cdot \max_{\forall j} \{ \Delta_{0i}(j) \} = \Delta_{02}(3) = 0.64 \\ \rho &= 0.5 \end{aligned} \right\} (8)$$

である。

式(6)で算出した各  $\Delta_{0i}(j)$  を式(7)に代入すると、各  $\gamma_{0i}(j)$  は、

$$\left. \begin{aligned} \gamma_{01}(1) &= 0.964 & \gamma_{01}(2) &= 0.964 & \gamma_{01}(3) &= 1.000 \\ \gamma_{02}(1) &= 0.589 & \gamma_{02}(2) &= 0.564 & \gamma_{02}(3) &= 0.552 \\ \gamma_{03}(1) &= 0.815 & \gamma_{03}(2) &= 0.815 & \gamma_{03}(3) &= 0.869 \\ \gamma_{04}(1) &= 0.768 & \gamma_{04}(2) &= 0.791 & \gamma_{04}(3) &= 0.791 \end{aligned} \right\} (9)$$

となる。

式(5)より、灰色関連度  $\Gamma_{0i}$  を算出する。

$$\left. \begin{aligned} \Gamma_{01} &= \frac{1}{3} (0.964 + 0.964 + 1.000) = 0.976 \\ \Gamma_{02} &= \frac{1}{3} (0.589 + 0.564 + 0.552) = 0.568 \\ \Gamma_{03} &= \frac{1}{3} (0.815 + 0.815 + 0.869) = 0.833 \\ \Gamma_{04} &= \frac{1}{3} (0.768 + 0.791 + 0.791) = 0.783 \end{aligned} \right\} (10)$$

式(10)で得られた灰色関連度  $\Gamma_{0i}$  を降順に整列することで次式に示す序列情報とする。樹形図の作成には数列番号および灰色関連度の双方を使用する。

$$\left. \begin{aligned} \Gamma_{01} = 0.976 & \quad \Gamma_{03} = 0.833 & \quad \Gamma_{04} = 0.783 & \quad \Gamma_{02} = 0.568 \\ \Gamma_{01} \succ \Gamma_{03} \succ \Gamma_{04} \succ \Gamma_{02} \end{aligned} \right\} (11)$$

### 3.2 クラスタリング情報の獲得

クラスタリング情報の獲得には表1の比較数列だけをを用いる。はじめに、灰色関連マトリクス  $R_{n \times n}$  と呼ばれる行列を構築する。灰色関連マトリクスとは、従来のクラスター分析における類似マトリクスに相当する

表1 数値例

	$c_i(1)$	$c_i(2)$	$c_i(3)$	
序列基準数列	$c_0$	1.00	1.00	1.00
比較数列	$c_1$	0.77	0.77	0.79
	$c_2$	0.42	0.38	0.36
	$c_3$	0.67	0.67	0.71
	$c_4$	0.63	0.65	0.65

行列のことであり、各比較数列を一度ずつ基準数列に設定し、灰色分析することで得られる灰色関連度  $\Gamma_{ij}$  を行列表現したものである。

$$R_{n \times n} = \begin{bmatrix} \Gamma_{01} & \cdots & \Gamma_{0j} & \cdots & \Gamma_{0n} \\ \vdots & \ddots & & & \vdots \\ \Gamma_{i1} & & \Gamma_{ij} & & \Gamma_{in} \\ \vdots & & & \ddots & \vdots \\ \Gamma_{n1} & \cdots & \Gamma_{nj} & \cdots & \Gamma_{nn} \end{bmatrix} \quad (12)$$

比較数列が  $n$  ある場合、灰色関連マトリクスは  $n \times n$  行列となる。表1の比較数列を用いると、 $R$  は、

$$R_{4 \times 4} = \begin{bmatrix} 1.000 & 0.356 & 0.698 & 0.618 \\ 0.356 & 1.000 & 0.423 & 0.458 \\ 0.653 & 0.374 & 1.000 & 0.819 \\ 0.522 & 0.364 & 0.790 & 1.000 \end{bmatrix} \quad (13)$$

となる。 $R_{n \times n}$  は対称行列である必要はない。

次に、 $R_{n \times n}$  から連関尺度となる灰色相関係数  $l(c_i, c_j)$  を算出する。灰色相関係数とは、従来のクラスター分析における類似係数に相当する値であり、二つの比較数列  $c_i, c_j$  間におけるデータの類似度を表す値である。次式にて  $l(c_i, c_j)$  を算出する。

$$l(c_i, c_j) = \frac{\Gamma_{ij} + \Gamma_{ji}}{2} \quad (14)$$

ただし、

$$\left. \begin{aligned} i &< j \\ i &= 1, 2, \dots, n-1 \\ j &= 2, 3, \dots, n \\ 0 &\leq l(c_i, c_j) \leq 1 \end{aligned} \right\} (15)$$

灰色関連度  $\Gamma_{ij}$  は、数列  $c_i$  を基準とした場合の数列  $c_j$  との類似度であり、 $\Gamma_{ji}$  は数列  $c_j$  を基準とした場合の数列  $c_i$  との類似度である。従って、式(14)より  $l(c_i, c_j)$  は  $\Gamma_{ij}$  と  $\Gamma_{ji}$  の平均値とすることで二つの比較数列  $c_i, c_j$  間のデータの類似度としている。 $l(c_i, c_j) \rightarrow 1$  であるほど  $c_i, c_j$  間のデータが類似していることを表す。

灰色相関係数は各比較数列の組み合わせであり、総

数は比較数列の数により決まる。灰色相関係数の総数を  $m$  とすると、

$$m = \sum_{k=1}^n (k-1) \quad (16)$$

である。

表1の例の場合、灰色相関係数の総数  $m$  は6であり、灰色相関係数  $l(c_i, c_j)$  を計算すると

$$\left. \begin{aligned} l(c_1, c_2) &= 0.356 & l(c_1, c_3) &= 0.675 & l(c_1, c_4) &= 0.570 \\ l(c_2, c_3) &= 0.398 & l(c_2, c_4) &= 0.411 & l(c_3, c_4) &= 0.804 \end{aligned} \right\} (17)$$

となる。

式(17)を  $l(c_i, c_j)$  の降順に整理し、 $c_i$  および  $c_j$  による比較数列の組み合わせ情報および  $l(c_i, c_j)$  の値をクラスタリング情報とする。表1の例では次式がクラスタリング情報となる。

$$\left. \begin{aligned} l(c_3, c_4) &= 0.804 & l(c_1, c_3) &= 0.675 & l(c_1, c_4) &= 0.570 \\ l(c_2, c_4) &= 0.411 & l(c_2, c_3) &= 0.398 & l(c_1, c_2) &= 0.356 \end{aligned} \right\} (18)$$

### 3.3 樹形図の描画

樹形図は比較数列またはクラスター同士を結線したグラフである。本提案クラスター分析では類似度の高い  $l(c_i, c_j)$  における比較数列の組み合わせ  $c_i, c_j$  から順にクラスタリングしていく。結線する前に予め序列情報を基に各数列を整理し描画しておく。

新たに作られるクラスターを  $t$  とすると、 $t$  は比較数列または既存のクラスターの集合と捉えることができる。比較数列同士をクラスタリングする場合、クラスター  $t$  は、

$$t = c_i \cup c_j, \quad \text{if } \max_{\forall l} \{l(c_i, c_j)\} \quad (19)$$

により構成される。グラフは  $c_i$  と  $c_j$  を結線する。

$c_i$  あるいは  $c_j$  のどちらか一方が既存のクラスター  $p$  に属している場合、 $t$  は、

$$\left. \begin{aligned} t &= p \cup c_j \quad (c_i \in p) \\ t &= c_i \cup p \quad (c_j \in p) \end{aligned} \right\}, \quad \text{if } \max_{\forall l} \{l(c_i, c_j)\} \quad (20)$$

とすることで、既存のクラスターに数列が包含される。グラフは既存のクラスターと数列を結線する。

$c_i$  および  $c_j$  の両者がそれぞれ既存のクラスター  $p, q$  に属している場合、新たに構成されるクラスター  $t$  は、

$$t = p \cup q, \quad \text{if } (c_i \in p, c_j \in q), \max_{\forall l} \{l(c_i, c_j)\} \quad (21)$$

とすることで、クラスター同士で新たなクラスターを

構成する。グラフはクラスター同士の結線となる。一度参照した  $\max_{\forall l} \{l(c_i, c_j)\}$  は破棄する。次回のクラスタリング時には次の  $\max_{\forall l} \{l(c_i, c_j)\}$  を参照する。式(19)、式(20)、式(21)のいずれかの処理を繰り返し行くと、最終的にクラスター  $t$  は、

$$t = \{c_1 \cup c_2 \cup \dots \cup c_n\} \quad (22)$$

となり、全比較数列を含む一つのクラスターとなる。

表1の例を実行すると、式(11)の序列情報および式(18)のクラスタリング情報を用いて、図1に示す樹形図が作成される。灰色クラスター分析で得られる樹形図は、図の左側に序列情報、右側にクラスタリング情報の計二つの情報を表示する。従って、従来の樹形図よりも情報量は増加していることになる。

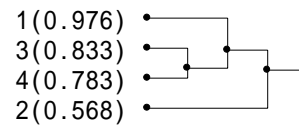


図1 数値例のクラスタリング結果

### 3.4 アルゴリズム要約

本稿で提案する灰色クラスター分析のアルゴリズムを以下に要約する。

- Step1: データマトリクスとして序列基準数列  $c_0$  および比較数列  $c_i$  の獲得
- Step2: データマトリクスを灰色分析し、式(11)に示す形式の序列情報を獲得
- Step3: 式(18)に示す形式のクラスタリング情報の獲得
  - 3-1: 灰色関連マトリクス  $R_{n \times n}$  の獲得
  - 3-2: 灰色相関係数  $l(c_i, c_j)$  の算出と降順整理
- Step4: クラスタリングの実行および樹形図の描画

## 4. 適用例

### 4.1 適用例(1)-生物の特徴分類

クラスター分析あるいは数理分類学の典型的な例は動物や植物といった生物学における分類である。本適用例は15種類の動物を、表2に示すカテゴリを基準に分類する<sup>[12]</sup>。

#### 4.1.1 データマトリクス

表2の各項目はいずれも生物学からみたカテゴリであり、各生物の進化の過程を調査する試みとなってい

表 2 生物の分類変量<sup>[12]</sup>

$c_i(1)$	体のつりあいの形式	$c_i(13)$	消化器官 1
$c_i(2)$	生活様式 1	$c_i(14)$	消化器官 2
$c_i(3)$	生活様式 2	$c_i(15)$	消化器官 3
$c_i(4)$	生活様式 3	$c_i(16)$	消化器官 4
$c_i(5)$	生殖方式 1	$c_i(17)$	神経器官 1
$c_i(6)$	生殖方式 2	$c_i(18)$	神経器官 2
$c_i(7)$	脊椎・無脊椎	$c_i(19)$	神経器官 3
$c_i(8)$	節の分け方	$c_i(20)$	体の型 1
$c_i(9)$	循環方式	$c_i(21)$	体の型 2
$c_i(10)$	恒温・変温	$c_i(22)$	体の型 3
$c_i(11)$	呼吸方式 1	$c_i(23)$	体の型 4
$c_i(12)$	呼吸方式 2	$c_i(24)$	体の型 5

る。データマトリクスは表 2 の各項目を基に作成された表 3 に示す質的データとなっている。本適用例では伝統的なクラスター分析法と灰色クラスター分析の比較を行う。

灰色クラスター分析に用いる序列基準数列  $c_0$  は、ヒトのデータ  $c_1$  と同値にする。最も進化した生物とされているヒトを基準とすることで、樹形図における各生物の序列は高度な生態系から原始的な生態系の生物となり、各生物の進化の過程を樹形図から捉えやすくなることができる。

4.1.2 分析結果

図 2 は伝統的なクラスター分析法による樹形図<sup>[12]</sup>であり、図 3 は灰色クラスター分析による樹形図である。

図 2 および図 3 を比較すると、クラスタリング情報に大きな差異は認められない。しかし、灰色クラスター分析では序列情報から各生物の進化の度合いをヒトと数値的に比較することができる。灰色クラスター分析を用いることで、各生物の生物学的な分類および進化の順序を数理的に調査することが可能となっている。

4.2 適用例(2)-店舗の営業成績分類

近年はデータマイニングが盛んに行われ、各企業は長年蓄えてきたデータを科学的に処理し、経営に生かしている<sup>[13]</sup>。データマイニングで主に使われるのは統計処理手法であり、クラスター分析も含まれている。

本適用例では経営状態の評価に灰色クラスター分析の適用を試みる。

4.2.1 データマトリクス

表 4 は全国規模で展開しているある雑貨店 30 店舗のデータである<sup>[13]</sup>。各店舗を表 5 に示す項目で評価し、点数を[0,1]の実数値に収めている。表 4 のデータから

表 3 各生物の特徴データ<sup>[12]</sup>

	$c_i(1)$	$c_i(2)$	...	$c_i(23)$	$c_i(24)$
$c_1$ . ヒト	1	0	...	4	5
$c_2$ . 猿	1	0	...	4	4
$c_3$ . 犬	1	0	...	4	4
$c_4$ . 牛	1	0	...	4	4
$c_5$ . 馬	1	0	...	4	4
$c_6$ . 鯨	1	0	...	5	4
$c_7$ . 鳥	1	0	...	3	3
$c_8$ . 蛇	1	0	...	3	3
$c_9$ . 蛙	1	0	...	2	2
$c_{10}$ . 鯉	1	0	...	1	1
$c_{11}$ . エビ	1	0	...	0	0
$c_{12}$ . イナゴ	1	1	...	0	0
$c_{13}$ . ドブガイ	1	0	...	0	0
$c_{14}$ . マキガイ	2	0	...	0	0
$c_{15}$ . ヒドラ	0	0	...	0	0

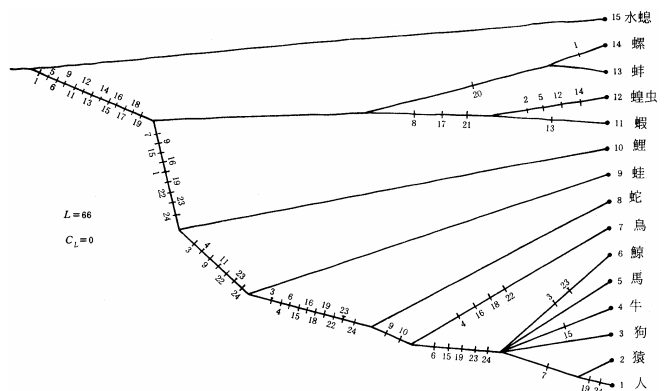


図 2 伝統的なクラスター分析による樹形図<sup>[12]</sup>

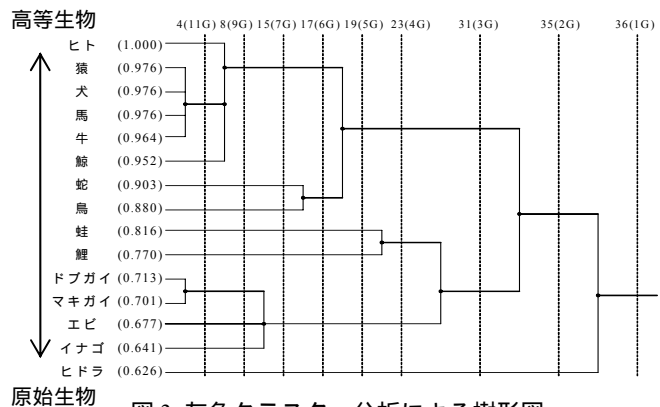


図 3 灰色クラスター分析による樹形図

表 4 営業店舗の評価データ<sup>[13]</sup>

	$c_i(1)$	$c_i(2)$	$c_i(3)$	$c_i(4)$	$c_i(5)$	$c_i(6)$
$c_1$	0.75	0.73	0.77	0.89	0.88	0.83
$c_2$	0.79	0.77	0.80	0.82	0.77	0.75
$c_3$	0.77	0.77	0.79	0.81	0.82	0.83
$c_4$	0.76	0.76	0.72	0.82	0.79	0.75
$c_5$	0.79	0.76	0.77	0.78	0.81	0.85
...	...	...	...	...	...	...
$c_{26}$	0.37	0.40	0.39	0.42	0.41	0.40
$c_{27}$	0.32	0.36	0.36	0.32	0.32	0.34
$c_{28}$	0.39	0.43	0.44	0.40	0.39	0.37
$c_{29}$	0.38	0.40	0.36	0.37	0.35	0.35
$c_{30}$	0.42	0.38	0.36	0.37	0.36	0.33

営業状態の優良な店舗およびそうでない店舗に分類する。序列基準数列  $c_0$  は各評価項目を満点と想定して、

$$c_0 = \{ 1, 1, 1, 1, 1, 1 \} \quad (23)$$

とする。式(23)の値に設定することで、樹形図における各店舗の序列は6項目の総合評価による営業状態ランキングとなる。

#### 4.2.2 分析結果

表4のデータおよび式(23)の序列基準数列をデータマトリクスとして灰色クラスター分析を行ったところ、図4に示す樹形図が得られている。

図4のクラスタリング情報によると、三つのクラスターに分けた場合、優良店舗、不良店舗、中間店舗に分類することができる。さらに、序列情報により営業評価の総合ランキングが示されているため、各店舗の営業成績が順位として表れている。分析者は図4から二つの情報を同時に得ることができるため、各店舗の営業状態を把握することが容易になる。また、図4から各店舗に個別の指導を与えることもできる。

以上のことから、本提案クラスター分析は経営状態の評価に使用することが可能である。

#### 4.3 適用例(3)-首都機能移転候補地域の選定

日本国政府は、国会などの首都機能を東京以外の土地に移転し、東京一極化の排除、災害対応力の強化などを図る計画を進めている。上記の計画で最も難航していることは、東京以外のどの地域に首都機能を移転させるかという問題である。ORではAHP法などを利用した移転先候補地の評価を行っている<sup>[14]</sup>。本適用例は灰色クラスター分析で移転先候補地の評価を行い、AHPと同様に意思決定支援することを試みる。

##### 4.3.1 データマトリクス

首都機能の移転先の選定基準は国会等移転調査委員会報告において定められている<sup>[14]</sup>。表6に選定評価基準を示し、表7に移転先候補となっている10地域の定量評価を示す。評価は[0,1]の実数値であり、1に近いほど評価が高いことを表している。表7を比較数列とすると、序列基準数列  $c_0$  は表6に示す各評価項目の最高評価値として、

$$c_0 = \{ 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1 \} \quad (24)$$

とする。式(24)に示す値を設定すると、序列情報は各候補地の総合評価のランキングとなる。

表5 店舗の評価項目<sup>[13]</sup>

$c_i(1)$	挨拶をしているか
$c_i(2)$	店員同士で私語をしていないか
$c_i(3)$	店員は客に適時声をかけているか
$c_i(4)$	店内は清潔に保たれているか
$c_i(5)$	商品は見やすく配置されているか
$c_i(6)$	商品は常に補充されているか

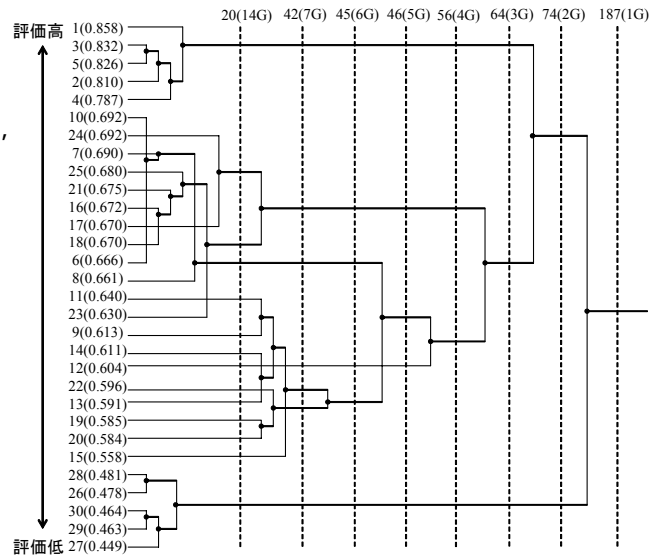


図4 雑貨店30店舗の営業評価

表6 候補地の評価基準<sup>[14]</sup>

$c_i(1)$	おおまかな位置	$c_i(9)$	空港の現状と計画
$c_i(2)$	政令指定都市との連携	$c_i(10)$	空港からの直線距離
$c_i(3)$	地域の主要都市	$c_i(11)$	開発可能性の土地
$c_i(4)$	全国参集時間	$c_i(12)$	起伏量
$c_i(5)$	ネットワーク上の位置	$c_i(13)$	標高
$c_i(6)$	東京からの直線距離	$c_i(14)$	地盤
$c_i(7)$	東京からの所要時間	$c_i(15)$	活火山
$c_i(8)$	新たな交通基盤整備		

表7 候補地域の定量評価<sup>[14]</sup>

	$c_i(1)$	$c_i(2)$	...	$c_i(14)$	$c_i(15)$
$c_1$ 宮城	1.00	0.33	...	0.35	0.33
$c_2$ 福島	1.00	1.00	...	1.00	0.33
$c_3$ 栃木	1.00	1.00	...	1.00	0.33
$c_4$ 茨城	1.00	1.00	...	1.00	1.00
$c_5$ 静岡	0.33	1.00	...	0.19	1.00
$c_6$ 岐阜	0.33	0.33	...	0.35	1.00
$c_7$ 愛知西	0.33	0.33	...	0.19	1.00
$c_8$ 愛知東	0.33	0.33	...	0.19	1.00
$c_9$ 三重	0.33	0.33	...	0.19	1.00
$c_{10}$ 畿央	0.33	0.33	...	0.35	1.00

### 4.3.2 分析結果

図5は表7のデータおよび式(24)の序列基準数列を用いて灰色クラスター分析した結果である。

図5中の序列情報から、移転先となる候補地域を総合評価ランキングとして見ることができる。また、クラスタリング情報から移転先候補地域の絞り込みを行うことができる。例えば、二つのクラスターに分割することで、 $c_2, c_3, c_4$ の地域を有力候補地域として絞り込むことができる。以上のことから、灰色クラスター分析はORの分野に適用可能である。

### 4.4 適用例(4)-色と言葉における感性の測定

感性情報処理研究の一環として、言葉からイメージする色または色からイメージする言葉を調査した報告がある<sup>[15]</sup>。本適用例では色からイメージする言葉を調査する。灰色クラスター分析を用いて、色と言葉の関係を樹形図に表すことを試みる。

#### 4.4.1 データマトリクス

本調査では表8に示す8色および感性語10対を採用し、感性計測システム<sup>[10]</sup>を被験者に操作してもらうことでデータ収集する。データ収集後、表9に示すデータが得られる。

表9のデータは $[0,1]$ の実数値であり、1に近いほどイメージが強いことを表す。本適用例では表8に示す“各色からイメージする言葉”を調査するため、変量は一つとして合計8回クラスタリングを行う。表9に示すデータは比較数列となり、序列基準数列 $c_0$ は最もイメージが強いことを表す値として、

$$c_0 = \{ 1 \} \quad (25)$$

とする。

#### 4.4.2 分析結果

分析結果の代表として、 $c_i(1)$ のデータを比較数列に設定し、式(25)の値を序列基準数列とした樹形図を図6に示す。

図6は“黒からイメージできる言葉”の樹形図である。序列情報から黒を強くイメージする言葉のランキングが得られる。クラスタリング情報から黒をイメージする言葉、イメージできない言葉の分類ができる。

同様のことが他の7色についても行うことができる。以上のことから、灰色クラスター分析は色と言葉についての感性情報処理に適用が可能である。さらに、本

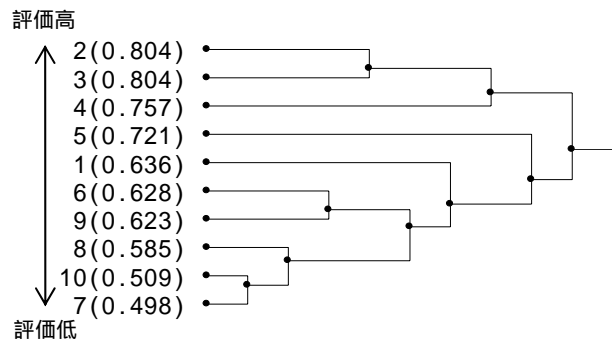


図5 首都機能移転候補地域の総合評価

表8 調査対象色

$c_i(1)$	黒	$c_i(5)$	赤
$c_i(2)$	青	$c_i(6)$	白
$c_i(3)$	緑	$c_i(7)$	黄色
$c_i(4)$	灰色	$c_i(8)$	茶色

表9 色と言葉の感性データ

	$c_i(1)$	$c_i(2)$	...	$c_i(8)$
$c_1$ . 派手な	0.21	0.72	...	0.05
$c_2$ . 上品な	0.27	0.59	...	0.07
$c_3$ . 心地よい	0.76	0.95	...	0.10
$c_4$ . 穏やかな	0.59	1.00	...	0.60
$c_5$ . 自然な	0.97	1.00	...	0.40
$c_6$ . 夢のある	0.02	0.79	...	0.21
$c_7$ . 嬉しい	0.01	0.61	...	0.35
$c_8$ . 怒りをおぼえる	0.33	0.00	...	0.12
$c_9$ . 悲しい	0.74	0.00	...	0.13
$c_{10}$ . 楽しい	0.15	0.85	...	0.21

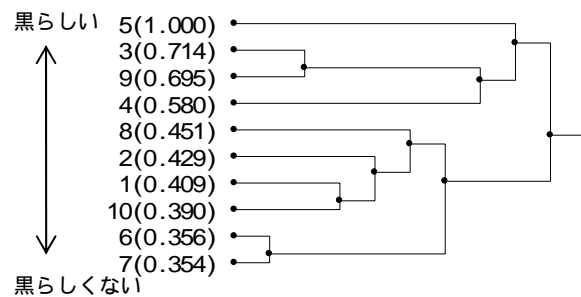


図6 黒からイメージする言葉

調査結果をデータベース化することで、言葉から色を検索するシステム<sup>[11]</sup>を構築することも可能である。

## 5. 考察

本稿では生物学、経営学、OR、感性工学などの分野において、本クラスター分析法が適用可能であることを示した。従来のクラスター分析と同じく、本提案方

法は多分野に適用できる可能性があることを示した。

本提案方法の利点は、樹形図が提供する情報が従来の樹形図よりも多くなっている点にある。新たに追加された序列情報は各比較数列の総合評価ランキングとなっているため、これまでの樹形図よりも多くの情報を分析者に提供している。従って、分析者に解釈の幅を広げる可能性が含まれている。実際に、本稿で示した適用例では序列情報およびクラスタリング情報の双方を見ることにより、分析結果から得られる有益な情報が増えている。

本稿ではクラスター分析のアルゴリズムおよび樹形図の評価を行っていない。基本的なクラスター分析の評価項目として、アルゴリズムや樹形図の正当性や計算量などが挙げられる。しかし、Anderberg はクラスター分析の価値はアルゴリズムの機械的な計算だけでなく、状況や分析者の研究目的やその解釈技術によって決まるため評価は難しいと指摘している<sup>[1]</sup>。アルゴリズムの評価については今後の課題となる。

## 6. むすび

本論文では、灰色理論に含まれる分析法である灰色分析から、クラスター分析と同等の機能を持つ新しいアルゴリズム“灰色クラスター分析法”を提案した。

灰色クラスター分析法は階層型分類法に属するため、分析結果は樹形図として木構造のグラフを出力する。また、灰色分析の特徴を生かし、データの類似度を表す灰色関連度 $\Gamma$ の値および数列の序列情報を樹形図に付加することで、分析者に対し樹形図からの解釈の幅を広げる可能性を示している。

生物学、経営学、OR、感性工学を代表して、四つの例に灰色クラスター分析の適用を行い、灰色クラスター分析法が多分野に活用できる可能性を示している。いずれの適用例においても樹形図から得られる情報量が従来のクラスター分析法よりも増えていることを示している。今後の課題としてはアルゴリズムや樹形図の正当性、また、アルゴリズムの計算量などの評価が挙げられる。

## 文 献

[1] M. R. Anderberg (著), 西田英郎 (監訳), クラス

ター分析とその応用, 内田老鶴圃, Dec. 1988.

- [2] H. C. Romesburg (著), 西田英郎, 佐藤嗣二 (共訳), 実例クラスター分析, 内田老鶴圃, July 1992.
- [3] 高橋正人, 服部和雄, “ファジー-c-means 法と最近傍決定則を用いたクリスプなクラスタリング手法,” 信学論(D-II), Vol.J83-D-II, No.9, pp.1957-1961, Sept. 2000.
- [4] 石原茂和, 石原恵子, 長町三生, “感性工学データ解析のための自己組織化ニューラルネットワークを用いた階層的クラスタリング手法の開発,” 信学論(A), Vol.J82-A, No.1, pp.179-189, Jan. 1999.
- [5] D. Julong, “Control problems of grey systems,” Systems and Control letters, Vol.5, pp.288-294, 1982.
- [6] 鄧聚龍 (著), 趙君明 / 北岡正敏 (共訳), “灰色理論による予測と意思決定,” pp.18-26, 日本理工出版会, 東京, Dec. 1999.
- [7] 永井正武, 灰色理論帝京大学理工学部講義ノート, 2003. (出版予定)
- [8] 山口大輔, 小林俊裕, 水谷晃三, 永井正武, “灰色分析を適用した商品特徴空間構築手法への一提案,” FIT2003, K-063, pp.561-562, Sept. 2003.
- [9] 山口大輔, 小林俊裕, 水谷晃三, 永井正武, “灰色分析を適用した判別モデルの提案,” 信学技報, Vol.103, No.521, HIP2003-82, pp.31-36, Dec. 2003.
- [10] 伊藤貴雄, 水谷晃三, 黒坂功, 永井正武, “自動数量化のための感性情報計測システム構築への提案,” 2000 信学総大, A-15-11, Mar. 2000.
- [11] 山口大輔, 小林俊裕, 水谷晃三, 永井正武, “色と言葉の感性データベース構築法への一提案,” 2004 信学総大, Mar. 2004. (執筆中)
- [12] 徐克學, “生物分類-数量分類學,” pp.249-259, 水産出版社, 台湾, 台北, Nov. 1996.
- [13] 内田治, “例解データマイニング入門,” pp.51-57, 日本経済新聞社, 東京, Oct. 2002.
- [14] 木下栄蔵, “AHP による首都機能移転候補地域選定に関する分析,” オペレーションズ・リサーチ, Vol.45, No.2, pp.67-75, Feb. 2000.
- [15] 金澤律子, “言葉における色のイメージ調査,” 日本色彩学会誌, Vol.27, No.SUPPLEMENT, pp.118-119, May 2003.