

## 仮想要素追加法による階層的クラスタリングの安定性の可視化

一宮 和正<sup>†</sup> 渡部 秀文<sup>‡</sup> 宮村(中村) 浩子<sup>‡</sup> 斎藤 隆文<sup>‡</sup>

<sup>†</sup>東京農工大学 工学部情報コミュニケーション工学科

<sup>‡</sup>東京農工大学 大学院生物システム応用科学府

### 1 はじめに

本研究では、階層的クラスタリングによって得られる階層構造の安定性を可視化する。

階層的クラスタリングを用いたクラスタ分析法は、複数の相関を持つデータをその類似性に基づいて外的基準なしに一意に分類するための手法である。これまでに様々な手法が提案されており、生物学や社会科学などの分野で利用されている。クラスタ分析法は純粋に数学的な手法であり、その性質から、データのわずかな違いによって得られる結果が大きく異なることがある。そのため、クラスタ分析を仮説の科学的裏付けなどに使う場合には、クラスタリング分析結果の安定性を考慮に入れる必要がある。安定性を測定する手法として Ben-Hur らの手法[1]があげられるが、統計的に扱う必要があるという問題がある。そこで、統計的基準によらず幾何学的に安定性を求める手法として仮想要素追加法[2]が提案された。しかし、ここで求まる安定性は数値として表されているため階層クラスタリングのどの領域がどの程度安定/不安定か認識することが困難である。そこで階層構造を表す樹形図上に、安定性も併せて表示する可視化手法を提案する。これによって階層的クラスタリングの安定性を直感的に認識できる。さらに、任意の階層におけるクラスタ間安定性を表で可視化し、クラスタ分析に役立つ情報を示す。

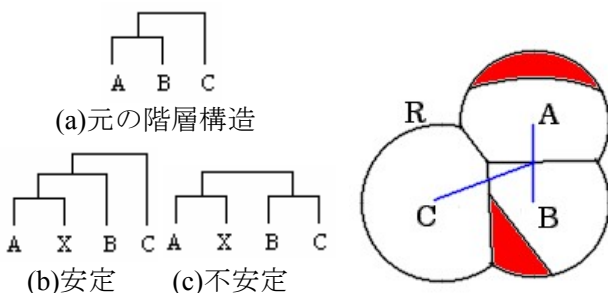


図 1：階層構造の変化 図 2：安定度計算領域

### 2 仮想要素追加法による安定性

本節では、仮想要素追加法[2]における、階層的クラスタリングの安定性の定義とその計算方法について述べる。

#### 2.1 安定性の定義

クラスタリング対象が 3 点である場合を考える。ここに仮想要素を 1 つ追加しクラスタリングを行なうと、その座標によって階層構造が変化する。例えば、図 1(a)のような階層構造に要素 X を追加してクラスタリングを行なうと、図 1(b), (c)のような結果が得られる。図 1(b), (c)ともに追加した X が A とクラスタを形成し階層構造が変化している。特に図 1(c)に注目すると B が A より先に C とクラスタを形成している。このように、追加要素と直接関わらない部分の構造が変化する場合を階層構造が安定でないとして定義する。

#### 2.2 安定性の計算方法

次に要素の追加による安定性の測定法について述べる。図 2 では A, B, C を元ある要素とすると、まず A と B がクラスタを形成する。そこでそれぞれの要素が半径|AB|の領域を持つとし、これら全てを合わせた領域を R とする。要素を領域 R 内に追加した場合、いずれかの要素とクラスタを形成し、階層構造の変化が起こる可能性がある。その領域 R 中で、階層構造に変化が起こる領域は図 2 の着色された領域となる。

そこで、仮想要素の追加により階層構造が変化する可能性のある領域 R に対し、実際に変化が起こる領域の比率が計算でき、この比率を階層安定度として計算できる。この階層安定度は最も不安定な場合で 1/3, 安定な場合で 1 となる。

### 3 安定性の可視化

本項では、仮想要素追加法により求められる階層安定度を樹形図上で可視化する手法と、任意の深さにおけるクラスタ間の安定性を表示する手法について述べる。また、計算した安定度は図 3 に示す色相、輝度にそれぞれマッピングする。

Visualization of Stability of Hierarchical Clustering by Adding a Temporary Element

Kazumasa ICHIMIYA<sup>†</sup>, Hidehumi WATANABE<sup>‡</sup>, Hiroko Nakamura MIYAMURA<sup>‡</sup>, Takafumi SAITO<sup>‡</sup>

<sup>†</sup>Department of Computer, Information and Communication Sciences, Tokyo University of Agriculture and Technology

<sup>‡</sup>Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology

### 3.1 樹形図上での可視化

階層構造を樹形図上に可視化するにあたって、クラスタ間の距離を考慮して、ノードの配置を決定する。図 4(a)では、先に結合する A, B を 1 つのクラスタとして、図 4(b)の 3 ノードの配置を考える。このとき、C, D のうち A+B のクラスタに距離が近い方を A+B に近い左側に配置する。このように配置することで、樹形図上でもクラスタ間の距離関係の把握が可能になる。さらにその 3 要素に注目すれば、どのような構造変化が起きやすいのかがわかる。

さらに安定度は 3 要素から計算されるため、樹形図上で可視化するにあたっていずれかの 3 ノードを選択する必要がある。そこで、同じくクラスタリングの結合順序により選択しノード間で計算して可視化する。図 4(a)の場合、A と B が先にクラスタを形成するため、図 4(b)のように A+B のクラスタと要素 C, D の間で計算し可視化する。可視化結果は、どの 3 点から階層安定度が計算されたかわかるように、3 点を頂点とした三角形で可視化する(図 5)。図 5(b)のクラスタ A, B, C に着目すると、その間で安定度が低いことがわかる。この場合、A+B というクラスタの安定度が低く、先に B と C がクラスタを形成することも考えられることを意味している。図 5(a)の色相マッピングでは安定度の値の読み取りなど、個々のクラスタの分析に利用できる。図 5(b)の輝度マッピングでは、全体から安定度の低い部分の発見が容易になり、全体的な構造の分析に利用できる。

### 3.2 任意階層での可視化

任意の階層を指定し、その階層でクラスタを形成する 2 要素とその他の要素の間でそれぞれ安定度を計算し可視化する。行をその階層に存在する要素、列を 1 つのクラスタに形成される 2 要素の組として表の形で表示する。また、行、列ともに並び順は 3.1 項で示した樹形図に合わせることで樹形図との対応が容易に行なえる。図 6(b)の矢印で示した部分は図 5(b)のクラスタ対 E とクラスタ D との安定度が低いことを示している。すなわち、E の一方が D とクラスタを形成する可能性がある。図 6(a)の色相マッピングでは安定度の読み取りと、クラスタ間の安定度の分布がわかる。図 6(b)の輝度マッピングでは安定度の低い部分の発見が容易で、不安定なクラスタに対して影響を及ぼす可能性のある要素が容易に発見できる。

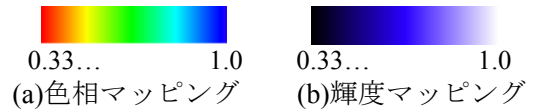


図 3：階層安定度のマッピング

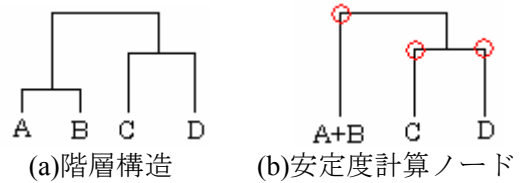


図 4：安定度計算におけるノード選択

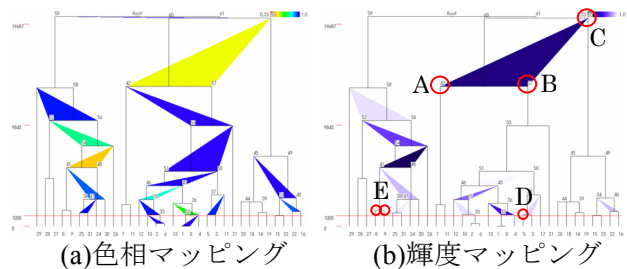


図 5：樹形図上での可視化

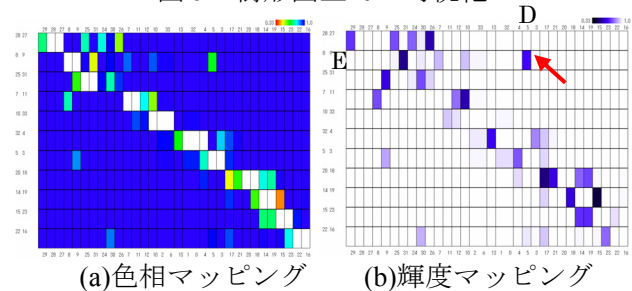


図 6：任意階層での可視化

## 4 おわりに

仮想要素追加法による安定性の樹形図上への可視化を実現し、さらに任意階層におけるクラスタ間の安定性を可視化した。これは、安定性が低いクラスタや階層構造に影響を与える可能性のある重要なクラスタの発見に役立つことができる。今後の課題として、最適なクラスタ数の発見を支援する可視化の実現があげられる。

## 参考文献

- [1] A. Ben-Hur, A. Elisseeff, and I. Guyon, "A stability based method for discovering structure in clustered data," In *Proc. Pacific Symposium on Biocomputing 2002*, pp. 6-17, 2002.
- [2] 南雲 拓, 斎藤 隆文, 宮村(中村) 浩子, "仮想要素追加法による階層的クラスタリングの安定性の解析," 情報処理学会研究報告(MPS-58), pp.31-34, 2006.