

# 少数データを丁寧に分けられる 「階層型クラスタ分析」の基本原則

牧野 浩二, 北野 雄大

## 階層型クラスタ分析の特徴

クラスタ分析には「階層型」と、「非階層型」があります。階層型は、結果を樹形図として描くことができます。図1と図2の違いを示します。

### ● 少数のデータが得意

階層型は少数のデータを丁寧に分割したいときに適しています。階層型は樹形図になっていますので、1つ1つのデータのつながりに着目できます。しかし、データ数が多いと下の階層がごちゃごちゃしてしまい、このメリットがなくなってしまいます。

非階層型は大まかに設定した数に分割するため、大局的な(おおざっぱに)傾向を見たい場合に適していると言えます。ビッグデータと呼ばれる膨大なデータを扱う場合は非階層型の方がよい場合が多いです。

### ● 結果の再現性が高い

階層型では同じデータに対して同じ処理を行うため、何度繰り返しても同じ結果が得られます。

非階層型は初期値に影響されてしまいますので、毎回同じ結果が得られるとは限りません。

### ● 適切な分割数分かる

階層型は分割の状態を把握できるので、処理結果からどの分割数がいかに判断できます。

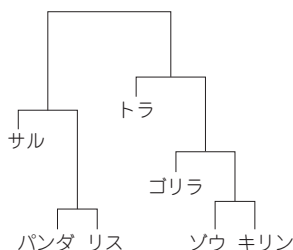


図1 分析対象の個々のデータを樹形図として関係づけてくれる「階層型クラスタ分析」は少数データをていねいに分けられる

非階層型は初めに何個の集合に分けるかを指定するので、分割数を指定しないとうまく分けられません。

### ● 計算量は多くなってしまふ

階層型は比較→階層決定→比較→階層決定…と処理を繰り返します。データの総数が多い場合、階層型の比較回数が多くなり、計算量が多くなります。

非階層型は分割数を初めに決めるため、比較は数回で済みます。

### ● 応用分野

階層型/非階層型ともに次の分類によく使われています。どちらを使うかは上に書いた特徴にマッチしているかどうかで決めます。

- クレジット・カードの利用頻度や額からのダイレクト・メール送信頻度の判別
- 購買・閲覧履歴からのおすすめ商品の判別
- 販売商品の分類による新規商品の差別化戦略策定
- 売り上げ実績や稼働状況からのプロジェクトの評価
- アンケート結果を利用した商品評価

## 使い方のイメージ

### ● メリット…キモとなるデータの解釈をいろいろ検討できる

上記のように、階層型クラスタ分析の特徴は、集団

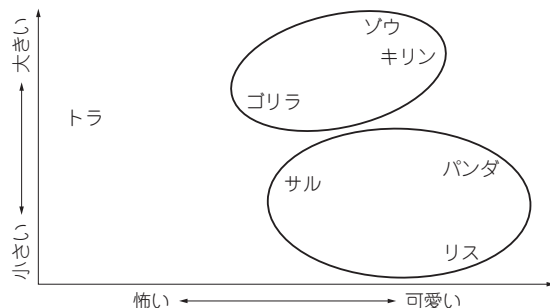


図2 図1と同じ元データを非階層型クラスタ分析で解析(クラスタ数:3)…大ざっぱな傾向しか分からない