わりとよく使われるタイプは動かしてガッテン！

ダウンロード・データあります

人工知能アルゴリズム探検隊

第6回 少数データを丁寧に分けられる「階層型クラスタ分析」の基本原理

ご購入はこちら

牧野 浩二，北野 雄大

人工知能のアルゴリズムとしては，第5回（2017年2月号）に引き続きクラスタ分析を使います．今回紹介するのは「階層型クラスタ分析」です．

クラスタの意味は「塊，群れ，集団」などで，データをグループに分けて分析をする手法です．

階層型クラスタ分析の特徴

クラスタ分析には「階層型」と，第5回で紹介した「非階層型」があります．今回紹介する階層型は，結果を樹形図として描くことができます．**図1**と**図2**に違いを示します．

● 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します．

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良い場合が多いです．

● 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．

● 適切な分割数が分かる

階層型は分割の状態が把握できるので，処理結果からどの分割数が良いかを判断できます．

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う

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／非階層型ともに以下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まります．

- 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

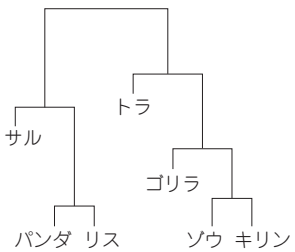

図1　分析対象の個々のデータを樹形図として関係づけてくれる「階層型クラスタ分析」は少数データをていねいに分けられる

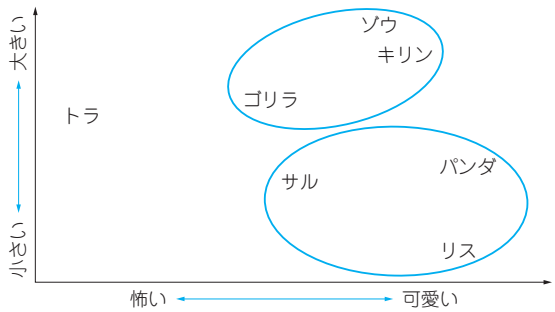

図2　図1と同じ元データを前回の非階層型クラスタ分析（2017年2月号）**で解析**（クラスタ数：3）**…大ざっぱな傾向しかわからない**

第1回　PICで試せちゃう！なんとなーく思い出せる人工知能記憶装置「アソシアトロン」（2016年9月号）
第2回　パターン認識でよく使われる「サポート・ベクタ・マシン」（2016年10月号）
第3回　ニューラル・ネットの基本学習法「バックプロパゲーション」（2016年11月号）