

多数データのグループ分け 「クラスタ分析」

牧野 浩二, 北野 雄大

クラスタ分析の特徴

属性の分からない多数のデータを、いくつかのグループに分けてくれるのがクラスタ分析です。クラスタ(cluster)には、「塊」、「群れ」、「集団」などの意味があります。

クラスタ分析がよく利用されている場面として市場調査があります。例えば次の用途が挙げられます。

1. 売り上げ実績や施設の稼働状況をもとにプロジェクトを評価
2. アンケート結果を利用した商品評価

例えばフィットネス・クラブを例に説明します。

- 機械をよく使い、かつ風呂にもゆっくり入る人
- ルーム・ランナをよく使い、シャワーだけで済ませる人
- 機械もルーム・ランナも使い、風呂にゆっくり入る人
- 風呂だけの人

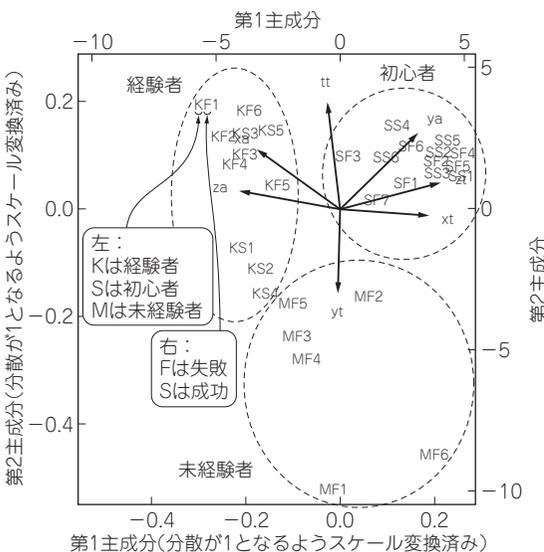


図1 けん玉経験者と初心者/未経験者の動きを主成分分析で見やすいように2次元プロットした

これらを曜日ごと、年齢/性別ごとに分ければ、「土日の利用者を増やすには、駅前のサラリーマンにピラを配ろう」、「シャワーの数を増やし女性客を呼び込めば機械の稼働率が上がる」などといった営業上の対策が打てます。

● 前章の主成分分析と組み合わせると見やすく分類できる

図1は「主成分分析」によって、けん玉経験者と初心者の動きを分類したものです。図1のように初心者、未経験者、経験者が分かれる結果となりました。図1の破線は筆者が単に、初心者、未経験者、経験者に分けるために、後から書き入れたものです。

図1のデータに、今回紹介するクラスタ分析を加えると、図2のように、プログラムが勝手に、何らかのグループ群に分割してくれます。図2の結果の中で、1つだけ(MF5)未経験者なのに経験者のデータ群として分類されていますが、後は見事に振り分けられています。

なお、図1はけん玉の技の1つ「とめけん」を行ったときのデータです(写真1)。とめけんとは、玉を垂直に一気に引き上げて剣先にさす技です。けん玉には

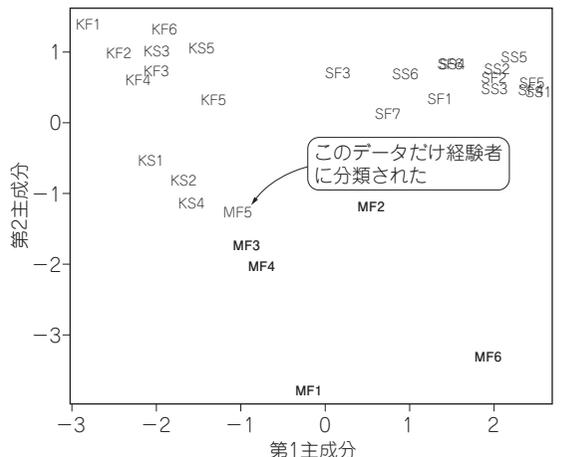


図2 図1にクラスタ分析を加えるとグループ分けできる