

因子得点のクラスター化を伴うペナルティ付き行列因子分析

三田村 俊

近年、情報技術の発展や計算機性能の向上により、大規模データの統計的分析とその結果の活用が盛んになっており、多変量データ解析の需要が高くなっている。多変量データ解析は、個体と変数がそれぞれ行と列に格納された行列型のデータを分析対象とし、特に[1]複数個体をより少数の群に分類するクラスター分析、[2]複数変数をより少数の共通因子に要約する因子分析、の2つはメジャーな分析方法である。本研究はクラスター分析と因子分析を融合した新手法を提案し、個体の分類と変数の要約を同時に達成することを目指した。個体の分類と変数の要約を同時に行う手法はいくつか存在し、例えばReduced K-means法(De Soete & Carroll, 1994)や、タンデム分析法がある。しかし、Reduced K-means法(以下RKM)は、変数の要約に主成分分析を用いており、因子分析で推定できる独自分散を推定することができない。独自分散とは、共通因子だけでは説明できなかった各変数に固有の変動であり、変数の特徴や変数同士の関係を把握するために重要である。また、タンデム分析法は、因子分析を適用した後、その結果を用いてクラスタリングを行うという、異なる分析法を段階的に使用するという手法であり、誤差の累積などの問題点が指摘されている(Arabie & Hubert, 1994)。そこで、変数の次元圧縮だけではなく独自分散まで推定できる因子分析とクラスター分析を単一の目的関数で表現することにより、上記の問題点を解決することを目指した。具体的には、因子分析の目的関数に共通因子得点のクラスター化を促すペナルティ関数を組み込む。このことにより、因子負荷量・独自分散の推定に加え、個体の分類まで同時に達成することが可能である。三田村(2021)は、因子分析とクラスター分析を用いた類似の分析手法を提案している。しかし、三田村(2021)の提案手法では、パラメーター推定の計算上の困難により、因子分析最大の特徴である共通因子と独自因子が無相関であるという制約条件を無視したモデルとなっている。そのため、狭義の因子分析を用いた手法とは言えず、不適切な解の導出を招く可能性がある。そこで、本研究ではパラメーター推定の計算に工夫を施し、共通因子と独自因子の無相関を考慮したモデルを提案した。そして、いくつかの実データ解析により提案手法の有用性が示された。性格特性や知能テストに関するデータの解析では、変数を要約したうえで個体を分類し、各群の特徴を推定することに成功した。例えば学校教育や就職活動の選考などにおいて、性格特性や知能テストのようなデータは質問項目やテストの種類といった変数を要約することに加え、その結果に基づいた個体の分類まで重要であるケースが多く、提案手法の応用的な側面が期待される結果となった。また、事前に個体が分類されている教師ありデータの解析では、推定した個体の所属群と真の所属群を比較することで正分類率を算出した。この実験で、RKM およびタンデム分析法の結果と比較し、算出した正分類率から提案手法の分類精度が最も高いことがわかった。さらに、RKM では推定できない独自分散を提案手法では推定することができ、提案手法の有用性が示された。研究過程において、例えば画像データのようなさらに幅広い実データへの適用や、提案モデルに用いているチューニングパラメーターの最適化など、多くの展望と課題が残されていることが判明し、今後さらなる発展が期待される研究になったと言えよう。(行動統計科学)