個体のメンバーシップ制約を伴う行列分解型因子分析の研究開発

宇野　光平

因子分析は，心理学において用いられるポピュラーな手法である．多数の測定変数が，観測されていない少数の因子に影響を受けているとする因子分析は，主成分分析と同様に次元縮約の手法とされている．2 つの手法，主成分分析と因子分析の大きな違いは，因子分析が変数独自の変数である独自因子をパラメータとしている点にある．しかし，独自因子をパラメータとすることで共通因子には不定性が生じる．これは因子の不定性と呼ばれ，因子分析の有用性に関わる根本的な問題である(2010，市川)．本研究では個体に対しメンバーシップ制約を課すことで，共通因子が数値的に一意に定まり，因子の不定性が解消されることを示した．

近年 de Leeuw(2004,2008)によって提案された行列分解型因子分析の目的関数は

$$\mathrm{f}(\mathbf{F},\mathbf{\Lambda},\mathbf{U},\mathbf{\Psi}|\mathbf{X}) = \|\mathbf{X}-\mathbf{F\Lambda}-\mathbf{U\Psi}\|^2$$

である．ここで，$\mathbf{X}$は$(n\times m)$のデータ行列で，$\mathbf{F}$は$(n\times p)$の共通因子行列，$\mathbf{\Lambda}$は$(m\times p)$の因子負荷行列，$\mathbf{U}$は$(n\times m)$の独自因子行列，$\mathbf{\Psi}$は$(m\times m)$の独自対角行列である．nは個体の数，mは変数の数，pは因子の数とする．本研究では，共通因子行列$\mathbf{F}$を 2 つの行列$\mathbf{G}(n\times k)$と$\mathbf{C}(k\times p)$と分解することで，メンバーシップ制約を設ける．ここで，kは群の数とし，行列$\mathbf{G}$は既知の行列とする．また$\mathbf{G}$の要素は0もしくは1で，各行に要素が1となる列は 1 つのみの，メンバーシップ行列とする．提案手法の目的関数は，

$$\mathrm{f}(\mathbf{C},\mathbf{\Lambda},\mathbf{U},\mathbf{\Psi}|\mathbf{X},\mathbf{G}) = \|\mathbf{X}-\mathbf{GC\Lambda}-\mathbf{U\Psi}\|^2$$

となる．推定する行列$\mathbf{C},\mathbf{\Lambda},\mathbf{U},\mathbf{\Psi}$でこれらを順番に推定していく交互最小二乗法を用いる．

提案手法の有用性を示すためシミュレーションと実データの解析を行った．シミュレーションは個体数を 3 条件，クラスター数を 3 条件の計 9 条件をそれぞれ 1000 回試行し，各行列の再現率を算出した．また OECD の経済指標データに提案手法を適用した．結果，単なる次元縮約の結果ではなく，メンバーシップ制約に基づき，各群をわける要因となった因子を抽出することができた．

本研究は，因子の不定性という因子分析の数学的問題の解決を目指したものであるが，因子分析において解釈に用いられる因子負荷行列に特徴があらわれることから，判別の手法としての利用も考えられる．たとえば，大学生の成績をデータを学生の所属学部を制約とすることで，所属する学部による成績の違いを明らかにすることが可能である．(行動データ科学)