

次元縮約手法における負荷行列の解釈可能性向上に関する研究

橋本 翔

第1章 序論

心理学データ, 社会学データ, 行動学データなどに対する分析手法として次元縮約法は広く用いられている分析法である。

その代表的なものは因子分析と主成分分析であり, これらの手法は顕在変数に影響を与えている潜在変数を測定するモデルとして用いられている。

特に, 質問紙調査の分析においては, リッカート法やSD法で得られたデータから因子分析を行うことによって, 尺度を構成するということが, 広く行われている。

しかしながら, 因子分析あるいは主成分分析の因子分析的用法では, 分析者は結果の解釈に多大な時間と労力が必要になっている。例えば, 共通因子に影響を受けている変数のみを分析に用いるため変数選択を行うが, 変数を取捨選択しながら分析を反復せねばならない。変数のクラスタリングや因子の命名を容易にする構造の探索を行いたいのに, 変数の値や構造が回転法や変数選択の結果変化するので, 容易には定めることができない。

そこで因子分析の負荷行列について分析者の目的ごとに解釈の容易な構造を想定した。変数選択という問題に対しては負荷行列の一部の行はすべてが 0 であることが望ましいと考えられる。変数のクラスタリングに対しては, 負荷行列は各行が数種類の行ベクトルからなっていることが望ましいと考えられる。因子の命名が可能な構造の探索に対しては, 負荷行列が絶対値の大きい値と小さい値からなることが望ましいと考えられる。

本論文の目的はこれらの構造を達成するための新たなデータ分析法を開発することである。

第2章 数学的基礎

第2章では, 本論文で必要となる線形代数と数値最適化, および確率分布をまとめた, 以降の章のための基礎を築いた。始めに, 行列とベクトルの表記について述べ, 基本的な演算と, 特殊な行列について説明を行った。そののちに, 最小二乗基準の説明し, そこで用いられているトレースの最大化について説明を行った。

またベイズ推定法について簡単に説明を行い, 分析に用いた確率分布を紹介した。

第3章 反復最小二乗法を用いた共通因子項目の変数選択法の開発

第3章では因子分析や主成分分析を用いるに当たり問題となっている, 「共通因子で説明されるべき変数の選択」を目的に, 共通因子から強く影響を受けている変数の選択と, それによる次元縮約を同時に行う分析モデルを開発した。

提案モデルでは, 一つの変数は共通因子のみに影響を受けているか独自因子のみに影響を受けているかのどちらであると仮定し, それぞれの変数を共通因子に説明される群と独自因子のみに影響を受けている群とに分類する。また, その過程で共通因子の因子負荷量と独自因子の因子負荷量が推定され, それを分析に用いることができる。

モデルに含まれるパラメータの最適化は交互最小二乗法によって行われ, それぞれのパラメータの更新アルゴリズムを示した。

提案モデルを用いた変数選択の適切さを測定するために, 数値シミュレーションを行った結果, サンプルサイズは 200 程度であり独自因子にのみ説明される変数が 2 つ以上の時に非常に高い精度で変数選

扱が可能であるということが示された。

実データ分析の結果からは、提案モデルを分析に用いることによって内的整合性の高い変数群を抽出できる可能性が示唆された。

第4章 非計量主成分分析を用いた変数のクラスタリング

第4章では同一クラスタ内の変数がすべて同一の負荷を持つというのは現実的な仮定ではないため、同一の負荷を持つことはそのままに、各変数の各カテゴリ異なったスケールで潜在変数を測定していると仮定し、アンケートデータをカテゴリカルデータとして扱うことを考えた。

そこで非計量主成分分析モデルの下で、各クラスタに所属する変数はカテゴリ数量が同一次元上に存在するという制約を利用し、複数のカテゴリカル変数をより少数の直線を用いて表現する手法を提案した。また提案手法のパラメータについて、反復最少二乗法を用いた推定方法について記述した。

実データの解析では、野球データの解析例からは、打撃積極性と走力についてがほぼ独立の関係であることや、ホームランバッターの性質について変数間の有用な関係性が得られ、これらのことから提案手法の妥当性が確認された。

また、疫学データの解析例からは、通常非計量主成分分析では見出すことが難しかった変数間のクラスタ構造を発見することができ、さらに症状を見分けるのに特に有用な変数を見出すという提案手法の応用可能性を見出すことができた。

さらに、得られた知見は従来法では得ることのできない、妥当性の高いものであった。就職希望者データの解析例からは、次元数を変化させても結果はほとんど変わらないことを示し、またこれらの結果は先行研究と比較しても妥当性があることが示された。

これらの結果から、本研究で提案された複数のカテゴリカル変数を一本の直線として表すという手法の有用性、そしてその結果の妥当性が示されたと考えられる。

第5章 ベイジアン因子分析を用いた解釈可能な負荷行列の探索

第5章では、探索的因子分析では回転の選択に関して不明瞭な研究者の恣意性が含まれる問題を解決するために、負荷行列が単純構造を持つことを事前情報として積極的に用いるベイジアン因子分析のモデルを提案した。

提案モデルでは、負荷行列の事前平均は0とそれ以外の比較的大きな値からなると定め、その事前情報を事前分布の形で分析に取り入れつつパラメータを推定するためのアルゴリズムを示した。

シミュレーション分析の結果から、提案モデルでは回転手法よりもより単純構造が強い負荷行列を推定することができることが示され、既存データの分析からはサンプルサイズが少数でも適切な解を推定可能であることや、既存手法では得られなかった有用な解釈を得られる可能性があることが示された。

提案モデルの利点としては、シンプリマックスと同様のターゲット分割を用いているにも関わらずターゲットの持つ0の個数を設定する必要がないこと、既存手法よりも単純性の高い負荷行列を求めることができること、事前情報や分析者の考えが分析に用いられるために異なる事前情報や知識を比較できることなどが挙げられる。

第6章 総合考察

総合考察においては、本論文で提案した3つの手法のそれぞれが、因子分析の負荷量の解釈にどのように役に立っているのかを改めて説明した。

3章で提案した変数選択モデルを用いることによって、内的整合性の高い因子に関わる変数が抽出に成功した。このことは、負荷行列の解釈を行う際に何度も繰り返し分析を行わなくても良いデータ解析法

であることが示され、変数選択問題の解決に役立つと言える。

4章で提案した変数のクラスタリング手法は、有用な変数のクラスタリングを行う可能性が示唆された。明示的にデータから変数のクラスタが抽出されており、負荷行列を解釈するよりもより直接的に変数クラスタが求められている。

5章で提案したベイズ推定により解釈可能な解をデータから直接探索するデータ解析法では単純構造を達成するような負荷行列の推定が成功し、さらにその解は従来の回転法よりも単純なものであった。このことから提案手法は、因子回転によって得られた負荷行列はデータとの照合が付かないという問題の、解決法の一つである考えられるだろう。(行動統計科学)