

氏 名 中 村 好 宏

学位（専攻分野） 博士(学術)

学 位 記 番 号 総研大甲第417号

学位授与の日付 平成11年9月30日

学位授与の要件 数物科学研究科 統計科学専攻

学位規則第4条第1項該当

学 位 論 文 題 目 等質性に基づく多変量データの低次元空間への射影とその応用

論 文 審 査 委 員 主 査 教 授 大隅 昇  
教 授 中村 隆  
教 授 馬場 康維  
教 授 田中 豊（岡山大学）  
教 授 佐藤 義治（北海道大学）  
教 授 柳井 晴夫（大学入試センター）

## 論文内容の要旨

複数の観測特性の多変量データを解析する手法として様々な多変量解析の手法が提案されており、実際に医学、生物学、心理学、社会学等、広い分野で利用されている。データ構造の視覚化に関わる代表的な解析法として、主成分分析、正準相関分析、数量化法、対応分析、等質性分析、双対尺度法等が挙げられる。その中の等質性分析は、Gifi, A. (1990)によりまとめられたもので、正準相関分析、主成分分析等を包含する応用範囲の広い興味深い手法である。

等質性分析では、変数間の等質性を測る基準として、特性ベクトルの線形結合により得られた「ベクトル」と、その平均ベクトルとの距離の平方和に基づく損失関数が用いられる。この損失関数の最小化により得られる空間に「ベクトル」を射影することによってデータ構造を記述する手法が等質性分析である。なお、ここで、「特性ベクトル」を、個体のプロフィールを表す「個体ベクトル」と置き換えれば、個体間の等質性を測ることができる。したがって、以下、「ベクトル」はデータ行列を形成する「特性ベクトル」あるいは「個体ベクトル」を示すものとする。

本論文では、この等質性基準を拡張し、より一般的な損失関数を提案している。Gifiによる等質性基準は、ベクトルとベクトルとの距離、すなわち点と点との距離に基づいたものである。本論文で提案した基準は、ベクトルと直線との距離あるいはベクトルと平面との距離、さらには、ベクトルとベクトルの射影空間との距離に基づくものであり、Gifiの等質性基準の自然な一般化になっている。具体的には、ベクトルの線形結合と射影空間との距離の平方和を損失関数として用いる。この損失関数の最小化によりベクトルの射影空間を得ることができる。特性ベクトルを射影空間に射影することによって、変数間の関係を視覚的に把握することが出来る。また、個体のプロフィールを表す「個体ベクトル」を直線、平面等へ射影すればデータ構造の視覚化が可能になる。

本論文では、この射影による変数の記述法を PML (Projection method by minimizing loss function) と呼ぶ。この PML は主成分分析法、正準相関分析法等の既存の多変量解析法の一般化と考えることができる。既存の多変量解析法の多くは、特異値分解に基づく直交系へのデータの射影である。PML では射影する空間の基底に対し直交の制約を与えない。したがって、解を求める方法は必ずしも特異値分解にはならない。本論文では、特異値分解にならない場合の解法として交互最小二乗法を用いた結果を示している。

本論文では、PML を応用したクラスタリングの一方法を提案している。ここで提案した方法は分割最適化型のクラスタ分析の一種である。分割最適化型のクラスタ分析とは、まず分類対象を指定した群の数にクラスタ化し、次にその分割した群のまとまりの良さを損失関数に基づく基準により評価し、適切な分割を探索的に求める分類手法である。

本論文では損失関数を用いて変数の分類を行うために、その損失関数を分割の評価基準としても用いている。分類対象の分割の組み合わせを入れ替えながら最適な分割を求め、その分割を良く表現する空間を求める。その空間にベクトルを射影することにより、分割を考慮したベクトルの射影を得ることが出来る。

実際のデータ解析においては視覚化されたデータ構造の解釈の方法が重要である。たとえば、因子分析では求めた因子を回転させることにより、因子の解釈を容易にする方法が

研究されている。本論文では、PMLにより求めた空間の基底を回転し、より解釈の容易な基底を求めるための基準も提案している。

論文は6つの章からなっている。第1章は、研究の歴史的な背景および本論文の構成の説明である。第2章は従来の等質性分析およびその関連手法の概観に当てられている。

第3章では、新たな手法としてPMLと名付けた、「等質性基準」に基づく一般的な射影法を提案し、さらにその性質を整理し、他の類似の方法との関係を論じている。前述のように所与のデータのベクトルの射影空間は一般化された損失関数の最小化によって求められる。また、損失関数最小化問題は、特定の条件の下では、固有値問題になることを示し、主成分分析法、正準相関分析法との関係を論じている。さらに、固有値問題となる場合も含むより一般的な損失関数最小化問題の解法アルゴリズムとして、交互最小二乗法が有効であることが示されている。また、人工データや実際データによるPMLの適用・分析例を示し、提案の方法の有効性を検証している。

第4章では、PMLを用いた分割最適化型のクラスター化法の一方法を提案している。ここでは、損失関数を最小にする意味で、データに内在する潜在的なクラスター構造を適切に表現する射影空間を探索するためにPMLを用いている。提案手法により、分類対象の所属クラスターの特定と、クラスター相互の相対的な位置関係の視覚的表現が同時に可能となる。

第5章では、PMLでは空間は一意に定まるが基底の取り方に不定性が生じるという性質から、PMLによる射影空間上での座標の回転について論じ、第3章で用いたデータについて座標回転の例示を行い、その効用を議論する。

第6章では、提案した方法の今後の発展の方向について述べている。以上の章に加えて、損失関数最小化と固有値問題の数理的な関係を付録として整理している。

## 論文の審査結果の要旨

審査委員会は数物科学研究科統計科学専攻の中村好宏氏の論文について、数物科学研究科における課程博士の授与に係わる論文審査等の手続き等に関する規定に基づき、公開の論文発表会を開催し審査を行った結果、下記の理由により博士論文として十分な内容を備えていると判断した。

(1) 論文概要：本論文は、「等質性基準」に基づく射影法（正射影）の観点から各種の多変量解析手法を数理的・総合的に整理し、新しい視点に着目した方法論を提案したものである。

論文は6つの章からなっている。第1章は、研究の歴史的な背景および本論文の構成の説明である。第2章は従来の等質性分析およびその関連手法の俯瞰に当てられている。

第3章では、新たな手法として PML(Projection method by minimizing loss function) と名付けた、「等質性基準」に基づく一般的な射影法を提案している。さらにその性質を整理し、他の類似の方法との関係を論じている。射影空間は一般化された損失関数の最小化によって求められる。損失関数最小化問題は、特定の条件の下では、固有値問題になることが示され、主成分分析法、正準相関分析法との関係が論じられている。さらに、固有値問題となる場合も含む一般的な損失関数最小化問題の解法アルゴリズムとして、交互最小二乗法が有効であることが示されている。また、人工データや実際データによる PML の適用・分析例が示され、提案の方法の有効性が検証されている。

第4章では、PML を用いた非階層的クラスタリングの一方法を提案している。この方法において、PML は損失関数を最小にする意味で潜在的なクラスター構造を最も良く表現できる射影空間を探索するために用いられている。ここで提案された手法により、分類対象の所属クラスターの特定と、クラスター相互の相対的な位置関係の視覚的表現が同時に可能となる。

第5章では、PML では空間は一意に定まるが基底の取り方に不定性があるという性質から、PML による射影空間上での座標の回転について論じ、第3章で用いたデータについて座標回転の例示が行われ、その効用が議論されている。

第6章では、提案した方法の今後の発展の方向について述べている。

以上の章に加えて、損失関数最小化と固有値問題の数理的な関係が付録に整理されている。

(2) 論文の評価：本研究は、多変量解析法における新しい視点に立った方法論を提案したものである。ここで提案された方法論は、「等質性」を広く解釈することによって導出された損失関数の最小化に基づく射影法（PML）である。

本論文では、距離をユークリッドノルムに限定しているが、加重ノルムに拡張することによって、さらに一般化の可能性があり、これは PML の今後の発展の可能性を示唆するものである。

また、本研究では、PML を適用することで、非階層的クラスタリングの一手法を提案している。このように PML をクラスタリングに導入することは新しい発想として評価さ

れる。非階層的クラスタリングに PML を導入することにより、損失関数最小の意味でクラスタ構造を最も良く表現する射影空間の探索と、分類対象に内在するクラスタ構造の検出を効果的に行うことができるという点で、注目に値する。

本論文の一部は、既に統計学の学術誌に掲載されることが確定している。

以上から、総合的に判断して、本研究は多変量解析の分野における有意義な貢献をしているものと評価できる。