博士（情報科学）　弘　　新太郎

学位論文題名

相対射影追跡法による次元縮小法とその応用に関する研究

学位論文内容の要旨

近年，情報技術の発展とともに収集するデータが大規模化，高次元化し，POS データやゲノムデータといった複雑な構造を内包したデータが増加している．また，データを収集する観測間隔も細かくなり，膨大な情報が蓄積されている．このような大規模な高次元データには，これまで得ることのできなかった有益な情報が内在しており，それらを適切に抽出するための手法の開発が期待されている．

一般に，データが高次元になるほど，様々な要因が複雑に絡み合い，その中から有益な情報を抽出することは困難になる．そこで，多変量データ解析では解釈が容易な低次元空間にデータを射影し，重要な情報を探索する次元縮小法が数多く研究されている．古典的な手法としては，主成分分析，因子分析などがある．最近では，コンピュータ指向型の手法として射影追跡法 (Friedman and Tukey, 1974) が提案されている．

射影追跡法は，正規分布から最も離れた構造を探索的に検出する次元縮小法であり，主成分分析などでは捉えることが困難な曲線的・曲面的非線形構造やクラスタ構造を見つけ出す有効な手法として知られている．ただし，射影追跡法は，正規分布を基準としているため，解析者が正規分布以外のある特定の分布もしくはある特定の標本の分布から離れた構造を抽出したい場合には，あまり有効ではない．

また，外的基準を有する高次元データに対して，応答変数に影響を与える説明変数空間を探索することは非常に難しい．このようなデータに対する次元縮小法としては，主成分回帰分析などがあげられる．最近では，ある要因が応答変数にどのように影響しているかを定量的に推定することを主目的とせず，応答変数に影響している説明変数の次元縮小空間の探索に主眼をおいた層別逆回帰法 (SIR) が注目されている．なかでも，射影追跡法を応用した射影追跡層別逆回帰法 (SIRpp) では，真のモデル構造が非常に複雑であっても，応答変数に影響を及ぼす説明変数空間を検出できることが示されている．しかし，SIR や SIRpp には，説明変数が楕円分布もしくは正規分布に従わなければならないという強い制約条件が存在する．

さらに，観測時点が非常に多い高次元データの解析に対して，新たなアプローチとして，データを関数化し，標本を関数の集合として扱う関数データ解析法が提案されている．関数データ解析の枠組みにおいても，低次元空間へ射影し，ある特定の関数群にみられる関数の特徴を検出する次元縮小法は非常に重要な役割を果たす．最近の研究では，関数射影追跡法などの手法が提案されているが，射影追跡法の拡張であるため，正規分布以外の分布から離れた構造を探索することはできない．

以上の背景により，本研究では射影追跡法を拡張し，比較の基準とする分布を正規分布に限ること

なく，解析者が既知の標本もしくは分布を比較の「参照」として定義し，その「参照」の分布から最も離れた構造を探索する相対射影追跡法を提案する．また，SIRpp 等における説明変数が受ける制約の問題を解決するために，相対射影追跡法を層別逆回帰モデルの上に適用し，新たに相対射影追跡層別逆回帰法 (SIRrpp) を提案する．さらに，関数データ解析法における新たな次元縮小法として，相対射影追跡法を拡張し，関数相対射影追跡法を提案する．

本論文は 6 つの章から構成される．

第 1 章では，本研究の背景，及び目的について説明している．

第 2 章では，本研究で扱う 3 つの従来手法について述べている．まず，本研究の基礎となる射影追跡法について，そのアルゴリズムと正規分布からの距離を測る射影指標について詳細に説明している．次に，外的基準を有するデータに対する次元縮小法である層別逆回帰法について述べている．なかでも，射影追跡法を用いた射影追跡層別逆回帰法について詳しく解説している．さらに，関数データ解析法について，その概念を述べた後に，関数射影追跡法について説明している．

第 3 章では，本研究で提案する相対射影追跡法について述べている．すなわち，相対射影追跡法において解析者が「参照」と定めた標本の分布からの距離を測るために，経験分布関数を使用した Area の相対射影指標，Friedman の射影指標を拡張して作成した Friedman Type 相対射影指標，Hall の射影指標を拡張した Hall Type 相対射影指標を提案している．また，数値実験により，これら 3 つの相対射影指標の有効性について比較検討し，Hall type 相対射影指標が 3 つの指標の中で最も真の次元縮小空間を捉えたことを示している．さらに，アメリカの大学における教員の年俸調査データに対して相対射影追跡法を適用し，博士課程を有する大学は，それ以外の大学とは異なり，大学の規模が小さくても教授の待遇がよい傾向にあるという情報を抽出できたことについて論じている．

第 4 章では，相対射影追跡法を層別逆回帰モデルへ応用した SIRrpp を提案している．SIRrpp の大きな特徴は，従来の層別逆回帰法とは異なり，説明変数の分布の制約を受けないことにある．本章では，この根拠について述べ，その汎用性の高さを説明している．また，数値実験により，正規分布に従っていないような，説明変数のデータ構造が複雑な人工データに対して，SIRrpp は真の次元縮小空間を探索可能であることを示し，分布の制約を受けないことを検証している．さらに，層別逆回帰法の従来法と比較することにより，SIRrpp が総合的に優れていることを示している．

第 5 章では，相対射影追跡法を関数データ解析へ拡張した関数相対射影追跡法を提案している．一般に，高次元空間上のデータの線形射影は，重みベクトル (射影ベクトル) とデータの内積で表される．それに対し，関数データは無限次元空間上のデータとして扱われるため，その射影は重み関数 (射影関数) と関数データの積の積分値で表される．この性質に基づき，解析対象とする関数データと解析者が定めた「参照」とする関数データを低次元空間に射影し，射影された空間上での両データの分布の距離を測る関数相対射影指標について説明している．また，数値実験において従来法である関数射影追跡法と比較し，提案手法が真の構造を検出できたことを示している．さらに，イギリスで行われた National study of health and growth (Holland *et al.*, 1999) の小児の成長曲線のデータに対して，5 歳時点で低身長の女児と標準身長以上の女児に分類し，低身長女児にみられるその後の成長の特徴を検出する目的で，従来法と提案手法を用いて解析を行っている．提案手法による解析の結果，低身長の女児には途中から標準身長に追いつく女児や年齢とともに成長速度が落ちていく女児などが存在し，標準身長以上の女児にはみられない成長の傾向を捉えたことについて論じている．

最後に，第 6 章で本研究の成果と今後の課題について述べている．

学位論文審査の要旨

主　査　教　授　水　田　正　弘
副　査　教　授　栗　原　正　仁
副　査　教　授　大　宮　　学
副　査　助教授　南　　弘　征

学位論文題名

相対射影追跡法による次元縮小法とその応用に関する研究

高次元データの解析において，次元を縮約させることは重要な課題である．古典的な多変量解析では，線形計算の範囲で実行可能な次元縮小法が数多く開発されてきた．例えば，主成分分析や因子分析がその代表である．これらの手法により，データが有する低次元な線形構造を見出すことができるが，複雑な構造や非線形な構造を抽出することは困難である．

線形計算に限定しない計算機指向型の次元縮小として，射影追跡法が注目されている．これは，正規分布を構造のない分布と想定し，低次元空間に射影されたデータの分布の非正規性を基準として次元縮小の射影を見つける手法である．また，射影追跡法は，射影追跡回帰モデルや射影追跡による層別逆回帰法 (SIRpp) をはじめとする非線形回帰分析法における説明変数空間の次元縮小にも適用されている．さらに，新しいデータ解析法として注目されている関数データ解析法においても射影追跡法が提案されている．しかし，射影追跡法は，正規分布との差をもとに射影方向を探索しているので，データの解析者が正規分布以外の分布もしくはある特定のデータの分布から離れた構造を抽出したい場合には利用できないという問題点がある．

このような背景から本論文では，射影追跡法を拡張し，比較の基準とする分布を正規分布に限ることなく，解析者が定義した既知の標本もしくは分布を比較の“参照”として，そこから最も離れた構造を探索する相対射影追跡法について述べている．また,SIRpp 等における説明変数が受ける制約の問題を解決するために，相対射影追跡法を層別逆回帰モデルの上に適用し，新たに相対射影追跡層別逆回帰法 (SIRrpp) を提案している．さらに，関数データ解析法における新たな次元縮小法として，相対射影追跡法を拡張し，関数相対射影追跡法を提案している．

本論文の主要な成果は以下のとおりである．

相対射影追跡法において，解析者が“参照”と定めた標本の分布からの距離を測るための指標として，経験分布関数を使用した Area の相対射影指標,Friedman の射影指標を拡張して作成した Friedman Type 相対射影指標,Hall の射影指標を拡張した Hall Type 相対射影指標が提案されてい

る. また, これら 3 つの相対射影指標の有効性について, 数値実験による比較検討を行い,Hall type 相対射影指標が 3 つの指標の中で最も妥当な次元縮小空間を捉えるという知見を得ている. また, アメリカの大学における教員の年俸調査データに対して相対射影追跡法を適用し, 従来の射影追跡法では捉えることのできなかった新たな情報を抽出可能であることが示されている.

SIRrpp の提案においては, 従来の層別逆回帰法で強い制約条件となっていた, 説明変数の分布に関する制約を提案手法により取り除くことができる根拠を述べるとともに, その汎用性の高さを示した. また, 説明変数のデータ構造が複雑であり, 正規分布に従っていないような人工データに対して数値実験を行い, 提案手法は真の次元縮小空間を探索可能であることが示されている. さらに, 従来の層別逆回帰法との比較も行われており, SIRrpp が総合的に優れていることを示している.

また, 関数データに対する射影の概念についてまとめ, 解析対象とする関数データと解析者が定めた“参照”とする関数データを低次元空間に射影した場合に, 射影された空間上での両データの分布の距離を測る関数相対射影指標が提案されている. 数値実験によって, 従来法である関数射影追跡法と比較し, 提案手法が真の構造を検出可能なことが示されている. また, イギリスで行われた National study of health and growth (Holland *et al.*,1999) の小児の成長曲線のデータに対して,5 歳時点で低身長の小児と標準身長以上の小児に分類し, 低身長小児にみられるその後の成長の特徴を検出する目的で, 従来法と提案手法を用いて解析を行い, 提案手法によって, 低身長の小児には途中から標準身長に追いつく小児や年齢とともに成長速度が落ちていく小児などが存在し, 標準身長以上の小児にはみられない成長の傾向を捉えたことについて論じている.

これを要するに, 著者は, 相対射影追跡法を提案, 発展させることにより, さまざまな高次元データから新たな情報を抽出するための手法について有効な知見を得たものであり, 情報科学ならびに計算機統計学に貢献するところ大なるものがある. よって著者は, 北海道大学博士 (情報科学) の学位を授与される資格あるものと認める.