

氏 名 樋 口 勇 夫

学位（専攻分野） 博士(学術)

学 位 記 番 号 総研大甲第363号

学位授与の日付 平成11年3月24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 自己組織化学習法則による多変量解析

論 文 審 査 委 員 主 査 教 授 松 縄 規
教 授 伊 藤 栄 明
教 授 江 口 真 透
村 田 昇 (理化学研究所)

論文内容の要旨

本論文は学習理論を応用した主成分分析の手法について理論的な考察と関連する学習アルゴリズムについての研究を行っている。近年、生物の脳をモデル化したニューラルネットワークの理論は様々な統計手法に応用されている。しかし、そのアプローチの性能はコンピュータによる数値実験でのみ示されている場合が多く、理論的な考察は十分とは言えない。ニューラルネットワークの特徴としては、並列的な情報処理と学習のプロセスがあげられる。ニューラルネットワークでは、複数のニューロンが結合して、個々のニューロンはそれぞれ独立して情報処理を行う。各ニューロンは結合比重というパラメータを持つ。ネットワーク全体の情報処理はそれぞれのニューロンが学習によって結合比重を修正することによって行われる。

本研究について、著者はつぎのような一般的な理解の下に研究を論文をまとめている。主成分分析は多変量データの縮約の代表的な方法として幅広く用いられている。これはデータ分散行列の固有値問題に帰着される。線形ニューロンによるネットワークは学習によって主成分分析を行うことができる。その時、結合比重が主成分ベクトルに相当する。従って学習のアルゴリズムを主成分分析に応用することができる。また実際面からの有利さとして、著者はニューラルネットワークによる主成分分析はオンラインデータに対する応用が容易であることも指摘している。このような認識を基盤として、著者は以下のような具体的な問題に取り組みその理論的解決を図っている。

データ分散行列は外れ値の影響を強く受ける。従って分散行列に基づいている主成分分析もロバストではない。Xu and Yuille は外れ値の判定を行うバイナリーフィールドを導入し、ニューロンの自己組織化法則を応用することで主成分分析のロバスト化を行ったが、そのロバストネスについては数値計算の結果を示したのみで理論的な考察は行っていなかった。主成分分析のロバスト化については様々な方法が提起されていたが、代表的な方法はデータ分散行列のM-推定量を用いる方法である。自己組織化ロバスト主成分分析のアルゴリズムは、ある重みづけ分散行列の固有値問題へと帰着される。

一般に、ロバストネスを示す尺度として、影響関数がある。従来の主成分分析の影響関数を計算すると、データが平均から離れるほど、そのデータに対する影響関数の絶対値が無限に大きくなることが分かる。これはそのようなデータが主成分分析の結果に大きく影響を与えることを示しており、このことから従来の（古典的）主成分分析はロバストでないといえる。自己組織化ロバスト分散行列の影響関数は陽に計算できる。その結果を用いて自己組織化アルゴリズムによる統計量の影響関数も計算することができる。自己組織化アルゴリズムの影響関数は主成分ベクトルの方向を除いて、平均からのデータの距離が遠くなるほどゼロに収束する。これは影響関数が主成分ベクトルに直交する成分が大きくなると急激にゼロになるような重みがついた形になっているからである。このことから自己組織化アルゴリズムは主成分ベクトルに直交する成分に対してロバストであることが分かる。

また、Maronna の行ったM-推定量を用いた方法との比較も行っている。分散行列の

M-推定量は、平均からの距離が大きなデータの重みを小さくした重み付き分散行列である。これに対して、自己組織ロバスト主成分分析の分散行列は、主成分ベクトルの方向に直交する成分に対して重みが小さくなる。平均から見て主成分ベクトルの方向に離れたデータは主成分ベクトルにあまり変化を与えない。このようなデータを排除しないのが自己組織化アルゴリズムの特徴であることを、影響関数の理論から示している。この重み付き分散行列に着目すると、この重みは分散行列の固有ベクトルを含んだ関数であるため、直接計算することは不可能である。そのため、これを反復計算で求めるアルゴリズムを提起している。基になる重み付き分散行列が等しいため、影響関数はXu and Yuille のアルゴリズムと同じであり、従ってこの方法も同様のロバストネスを持つ。数値計算の結果、本論文での方法はこれまでのXu and Yuille の方法と比べて少ない反復回数で収束することが示されている。

関連する研究として数量化理論についてつぎのことがらの考察もなされている。数量化の方法として知られる双対尺度法はデータ行列の特異値分解に帰着される。主成分分析と同様にエネルギー関数を考え、各データに対するエネルギーを定義することで、主成分分析と同様の重み付き行列を考えることができる。そこで、自己組織化法則による双対尺度法を提起し、その性質について議論している。数量化の場合、主成分分析の時のように平均から無限に離れた外れ値というものは考えにくい。この場合、自己組織化は外れ値の判定とは別の動きをすると考えられる。本論文では、数値実験の結果から、自己組織化双対尺度の性能についても考察している。

論文の審査結果の要旨

審査委員会は樋口勇夫君の論文について慎重に検討し、公開論文発表会を開催して審査を行った。下記のように、同君は、ニューラルネットワークの理論を用いた多変量解析の理論と関連する手法において幾つかの重要な結果と種々の興味ある知見を得た。特に主成分分析について集中的な研究がなされ、ニューラルネットワークスで提案された自己組織化による学習法則からの適用が検討された。

(1) 論文概要。七章から成っている。第1章で論文全体の構成、研究内容について紹介されている。第2章では、ニューラルネットワークと自己組織化法則についてその学習法則、特に自己組織化法則が本論文の統計解析に関わる事項について紹介され、検討がなされている。次いで第3章では、高次元のデータを低次元へ有効に射影する代表的な多変量解析手法である方法主成分分析とニューラルネットワークの学習についてそれらの関連性と意義が述べられている。第4章で自己組織化法則を用いたロバスト主成分分析について説明し、Xu and Yuilleの研究について検討している。第5章が本研究の中核であり、自己組織化法則によるロバスト主成分分析の影響関数についての計算とそれに基づくロバストネスの理論的検証がなされている。また影響関数によって従来の他のロバスト化との比較を行っている。第6章で、多変量解析手法へのひとつの応用として、自己組織双対尺度法が示されている。

(2) 論文の評価：本論文は従来のニューラルネットで提案されている方法に対して次の二点の貢献をした。第1点は、この自己組織化による主成分分析の理論的考察をしたことである。従来、データの外れ値に対してロバストであることが数値実験を通して確認されていたが、彼は汎関数（影響関数）の方法論でその理論的根拠を明らかにした。どんな外れ値を与えてもその方向からの汎関数微分が有界になり、そのノルムの上限を明示的に与えた。この結果は数値実験で得られたものと一致するものである。第2点は、勾配学習アルゴリズムの替わりに新しいアルゴリズムを提案したことにある。これは直前の学習によって得られた主成分スコアに応じて、全てのデータを重み付けするステップと重み付けされたデータに対して通常の主成分分析を行うステップを交互に反復させるアルゴリズムである。この交代化アルゴリズムは従来の勾配学習アルゴリズムと比較して安定した収束性を有することが証明された。

以上のように、樋口勇夫君の論文は、統計理論の視座からのニューラルネットの研究に重要な寄与をするものであり、今後のこの方面の研究に一つの手がかりを与えたものと言える。これらの貢献により、同君の論文は博士号取得に十分値する内容を備えていると判断した。