

射影追跡法に基づくデータの非線形構造探索に関する研究

学位論文内容の要旨

多変量データ解析は、事象の識別や予測を行ない、背後に潜む因子を探りデータから有用な情報を引き出す解析法である。しかし、扱うデータの変数の数が増えると、その構造を直観的に把握することや、解釈することが困難となる。そこで多変量データを観察や解釈が容易な次元にまで縮小するということが行なわれ、このための代表的な手法として主成分分析がある。この手法はデータの分布の様子を表現可能な最低次元の線形部分空間を求める手法である。近年、このような線形構造を発見する手法にかわって、コンピュータ指向の手法により非線形な構造を解析したり、データを視覚化することによって潜在する構造を探索的に解析することが重要視されている。しかし、これまでに提案された手法は低次元の非線形構造を検出するに留まり、多変量データに内在するより複雑な構造の抽出とその視覚化は多変量データ解析における重要な課題となっている。

多変量データの非線形性やクラスタなどの興味深い構造を観察しようとする手法として射影追跡法が提案されているが、従来方法では、データを射影する空間が1, 2次元に限られていたため、比較的単純な構造が確認できるだけであった。

一方、探索的データ解析に関連して、各種データの視覚化の方法が研究されてきた。このような視覚化の1つとして代数曲線の当てはめがある。これは、各変数が対等で、複数の説明変数によって1つの目的変数を表わすことが適当でないデータを視覚化する手法であり、当てはめた曲線により変量間の関係を観察するものである。しかし4変量以上のデータへ直接適用しても視覚化は困難である。また従来方法では、データ点との距離が最小となる曲面を求める非線形最適化法に適用する初期曲線の決定の問題と、当てはめる代数曲線の次数に関する問題が残されていた。

以上の背景から、本論文では、多変量データの構造検出とその視覚化の問題に対し、より複雑な構造の検出が可能となる3次元空間への射影追跡法の拡張について考察し、曲線当てはめ法を組み合わせることにより、多変量データの非線形構造を視覚化する手法を提案した。

射影追跡法を3次元へ拡張するにあたり、非線形性やクラスタなどの構造の興味深さを表わす3次元の射影指標を導出した。射影追跡法の基準となる射影指標はデータが正規分布に従うときに最小になり、正規分布から離れるにしたがって大きな値を取るという性質の他に、その計算量や外れ値の影響などが問題になる。そこで、射影指標の性質についてさまざまな例によって考察した。

また、曲線当てはめにおいて非線形最適化法での初期値の改良と情報量規準を導入した代数曲線の次数決定法によりデータ構造を表わす最適な曲線を求める方法を示した。さらに本論文で示した手法の適用例を示し、数値実験的側面から各手法の有効性を示した。

本論文は以下の6章から構成される。

第1章では、本研究の背景と多変量データ解析における射影追跡法と曲線当てはめの役割

について概説し、本研究の目的について述べた。

第2章では、従来の1, 2次元空間への射影追跡と代数曲線・曲面当てはめに関する研究について詳述した。射影追跡法の解析手順を示し、これまでに提案された射影指標を示した。さらに、曲線を当てはめる基準となる近似距離と当てはめのアルゴリズムについて説明した。

第3章では、より複雑な多変量データの構造を探索するために射影追跡法3次元空間へ拡張し、3次元の場合のFriedmanの射影指標、Hallの射影指標およびモーメント射影指標を導出し、その最適化について説明を与えた。さらに数値実験を通して従来の主成分分析法や1, 2次元への射影追跡法では確認できない構造が検出可能であることを示し、その有効性を検討した。また、射影追跡法において射影指標の選択はその性能を決定する重要な要素である。そこで、導出した3次元の各射影指標の計算量、外れ値による影響、射影される3次元空間の回転に関する不変性について考察し、各指標の特徴を示した。

第4章では、代数曲線当てはめの改良を示すとともに代数曲線の次数決定法を提案した。データ点との距離が最小となる曲線を求める最適化の計算では、多くの非線形最適化法の計算と同様に初期値の選択がその後の計算に大きな影響を及ぼす重要なものである。そこで同次座標系を導入した新たな初期値決定法を考案し、数値実験によりその有効性を示した。また、代数曲線の次数を高くするとデータへの当てはまりは良くなるが、次数の増加に伴いパラメータの数が急激に増加し、データのわずかな変化に過敏に反応し、曲線が不安定になり、データの構造を表わす曲線が得られない場合があり、計算量も増加する。この問題に対し情報量基準を導入した代数曲線の次数決定法を提案した。

第5章では、多変量データの構造を視覚化するために、直接、多次元のデータに代数曲線・曲面を当てはめるのではなく、3次元への射影追跡によって多変量データの非線形構造を3次元空間へ射影し、射影された3次元空間内のデータに対し代数曲面を当てはめ、多変量データの構造を視覚化することを提案した。射影追跡法によって3次元空間へ射影されたデータを散布図行列によって直接観測するよりも代数曲面を当てはめることで非線形構造が明確になり、変量間の関係を捉えることができる。このことを数値実験により示した。

最後に第6章では、結論として、本手法により従来検出が困難であった多変量データの非線形構造を見出すことができることを示し、その有効性を論じた。さらに現象解析においてデータの視覚化が重要な手段であるという立場から本研究の有効性を述べ本論文の総括とし、今後に残される課題について述べた。

学位論文審査の要旨

主 査 教 授 佐 藤 義 治
副 査 教 授 新 保 勝
副 査 教 授 伊 達 惇
副 査 助 教 授 水 田 正 弘

学位論文題名

射影追跡法に基づくデータの非線形構造探索に関する研究

多変量データ解析は、複雑な現象の機巧を解明するための方法論の一つとして、近年計算機の発達と共に急速に発展している分野である。現象に適したモデルを探索的に構築する過程において、観測データのもつ構造を把握することは不可欠であり、データを何らかの方法で視覚的に表現することが重要な課題となっている。

多変量データの視覚化や構造推定のための方法として、射影追跡法や曲面(曲線)当てはめ法などが提案されている。射影追跡法は、多変量データのもつ非線形性やクラスターなどの構造を視覚的に表現しようとする方法であるが、従来の方法では射影する空間が1, 2次元に限られ、比較的単純な構造が検出できにすぎなかった。一方、曲面当てはめ法は、データに当てはめた曲線や曲面によって、データの構造を変量間の関係として捉えるものであるが、従来の方法では、データ点との距離が最小となる曲面を求める非線形最適化法に適用する初期曲線の決定の問題と、当てはめる代数曲線の次数に関する問題が残されていた。

本論文はこのような背景に基づき、多変量データの構造検出とその視覚化の問題に対し、より複雑な構造の検出が可能となる3次元空間への射影追跡法の拡張について考察し、代数曲面当てはめ法を組み合わせることにより、多変量データの非線形構造を視覚化する手法を提案している。

本論文では、射影追跡法を3次元へ拡張するにあたり、非線形性やクラスターなどの構造の興味深さを表す3次元のFriedman, Hallの指標およびモーメント指標を導出し、これらの指標の計算量、外れ値の影響、射影される3次元空間の回転に関する不変性について考察し、各指標の特徴をさまざまな例によって示している。また、代数曲面当てはめ法での非線形最適化法において、同次座標系を導入した新たな初期値決定法を考案し、また情報量規準を導入した代数曲線の次数決定法により、データ構造を表現する最適な曲面を求める方法を確立し、数値実験によりその有効性を示している。

さらに、本論文において、射影追跡法によって射影された3次元空間内のデータに対し代数曲面を当てはめ、多変量データの構造を視覚化することを提案し、数値実験により、データの非線形構造が明確になり変量間の関係を捉えることができること

を示している。

これを要するに、著者は、多変量データにおける非線形構造の抽出とその視覚化に関して、射影追跡法の3次元空間への拡張および代数曲線・曲面当てはめ法の改良、さらに両手法を融合した効果的な探索的解析法を提案し、その有効性を示したものであり、情報解析学および情報処理工学の進歩に寄与するところ大である。

よって著者は、北海道大学博士(工学)の学位を授与される資格あるものと認める。