

氏名	玉谷 充
学位の種類	博士（理学）
学位記番号	総博甲第93号
学位授与年月日	平成26年3月25日
学位授与の要件	学位規則第4条第1項
文部科学省報告番号	甲第519号
専攻名	電子機能システム工学専攻

学位論文題目 Pattern recognition based on naive canonical correlations in high dimension low sample size
(高次元小標本におけるナイーブ正準相関に基づくパターン認識)

論文審査委員	主査 島根大学教授	内藤 貫太
	島根大学教授	中西 敏浩
	島根大学教授	黒岩 大史

論文内容の要旨

パターン認識とは、クラスへの属性が既知のデータがあり、どのクラスに属しているかが未知の観測値が新たに得られたときに、クラスが既知のデータを利用して、その新たな観測値を、属するクラスへ判別する手法である。パターン認識の枠組みは、文字や音声などのオブジェクトをデジタル変換処理する「観測」、オブジェクトの位置や大きさを整えたり、ノイズの除去や色の調節をする「前処理」、オブジェクトに依存した特徴を取り出し、観測値ベクトルに変換する「特徴抽出」、観測値ベクトルから判別するのに有効なものを選択する「特徴選択」、そして、その観測値ベクトルがどのクラスに属するかを決定する「識別」のもとで成り立っている。

パターン認識では、クラスが未知であるデータに対して、誤って判別してしまう確率をなるべく小さくすることが重要であり、様々な手法が提案されている。2クラス判別問題でよく用いられている手法の一つとして、フィッシャーの線形判別関数が挙げられる。この手法は、各クラスの平均ベクトルによる違いから生じるばらつきを表す“クラス間変動行列”と各クラス内のデータのばらつきを表す“クラス内変動行列”の比を最大にする手法であり、各クラスのデータが分散共分散行列の等しい正規分布に従っていれば、他のどんな手法よりも誤判別率が小さいことが一般に知られている。そして、多クラス判別問題においても、正準相関の観点で判別関数を構築することによって拡張を与えることが可能である。しかしながら、これらは次元がデータ数よりも小さいという仮定のもとでの話であり、次元がデータ数よりも大きい場合はクラス内変動行列の推定量が特異行列となってしまうため、この手法を適用することができないという問題が生じる。

博士論文では、ナイーブ正準相関に基づいた多クラス判別関数を新たに提案し、漸近理論を構築した。次元がデータ数よりも大きい場合のパターン認識手法は、2クラス判別問題においては様々な手法が提案され、漸近理論が展開されているが、多クラス判別問題における漸近理論については十分な議論がなされていない。その漸近理論として、誤判別確率の漸近上界を導出するこ

とは、手法の精度保証をする利点となり、大変重要であると言える。また、多クラスにおける特徴選択手法についても提案された手法が少なく、ナイーブ正準相関から得られる固有ベクトル又は判別方向ベクトルを用いたランキング法及び個数選択の基準を新たに提案することは有益であると言える。以下に、博士論文の概要を各節ごとに述べる。

第1節では、次元がデータ数よりも大きい場合のパターン認識手法の背景や本研究の意義を述べた。

第2節では、パターン認識手法の枠組みを述べ、フィッシャーの線形判別関数と正準相関に基づく判別手法が等価であることを述べ、多クラスへの拡張を紹介した。

第3節では、高次元小標本における理論的評価の困難さ及び特徴選択の必要性を述べた。そして、その困難さを回避するナイーブ正準相関に基づく判別手法を新たに提案した。

第4節では、判別関数の推定量の構築方法及び漸近挙動の設定を与え、ナイーブ正準相関によって得られる固有ベクトル及び判別方向ベクトルの漸近挙動を与えるために、適切な条件をいくつか与えた。そのもとで、固有ベクトル及び判別方向ベクトルの一致性については適切なモーメント条件を持つ一般分布のもとで議論を与えた。ベクトルの一致性については、推定ベクトルと目的のベクトルとのなす角度の観点を用いて議論し、次元とデータ数の間の増加具合によって一致性の有無が生じることを示した。そして、一致性が無い場合でも、適切な次元とデータ数の間の増加条件を満たせば、特定の角度へと確率収束することもまた示すことができた。また、誤判別確率については、その漸近上界を正規性の仮定のもとで導出した。この誤判別確率の漸近上界の導出は、次節の特徴選択の基準でも用いられるため、重要な定理の一つであると言える。

第5節では、ナイーブ正準相関から導かれる特徴選択手法を提案した。特徴選択手法は主にランキング法、個数選択の基準が重要となる。2クラスにおいては固有ベクトルに基づくランキング法と先行研究の等価性を与え、判別方向ベクトルに基づくランキング法を新たに提案した。個数選択については先行研究の手法を適用した。多クラスにおいてはランキングに用いるべきベクトルが複数本得られるため、そのことを含めた新たなランキング法を提案した。また、個数選択についても誤判別確率の漸近上界を用いた新たな基準を与えた。

第6節では、第4節で与えた条件を満たすいくつかの適切な分布のもとで、ベクトルの一致性に関する理論結果の数値実験を検証し、誤判別確率の上界についても正規分布のもとで数値実験による検証をした。また、特徴選択手法の性能についても数値実験によって先行研究と比較した。

第7節では、特徴選択手法を2クラス及び多クラスの実データに適用した結果を述べ、考察を与えた。

第8節では、博士論文で述べた補題、定理及び系の証明を載せている。

第9節では、博士論文で扱った内容を総括し、今後展開される研究として帯行列を用いた判別関数の手法を、帯行列の高次漸近挙動に関する先行研究と併せて述べた。

論文審査結果の要旨

本申請論文で議論されているパターン認識は、複数の群への属性がわかっているデータを用いて、属性のわからない新たなデータを群に判別する統計手法であり、画像解析、信号処理、医学など、広範な領域で用いられる。マイクロアレイなどの遺伝子発現量のデータの出現により、高次元のデータに基づくパターン認識の研究が近年重要となっている。データの高次元化により、データの次元がデータ数を大きく超えてしまう状況が起こり、このようなデータを高次元小標本と呼ぶ。高次元小標本の設定では、多変量解析における重要な要約統計量である標本共分散行列が特異になるという問題が生じ、これが高次元小標本における多変量解析の理論研究を難しくしている本質的要因であった。

本申請論文では、高次元小標本の設定の下、標本共分散行列についてその対角成分のみを用いることで、その特異性を回避して理論構築を行っている。特に、正準相関分析と判別分析との関連に注目し、ナイーブ正準相関と呼ばれる手法を定義し、それに基づくパターン認識手法を提案、その振る舞いを理論的に調べる研究となっている。

本申請論文に含まれる重要な貢献は、多群かつ次元とデータ数が共に発散する設定の下での提案手法に関する漸近理論の構築とその有効性の評価である。特に、

1. ナイーブ正準相関ベクトルの漸近挙動を、角度の一致性の観点から明らかにした (4 節)
2. ナイーブ正準相関から得られる判別方向ベクトルの漸近挙動を明らかにした (4 節)
3. 提案手法の誤判別確率の漸近上界を導出した (4 節)
4. ナイーブ正準相関ベクトルに基づく特徴選択手法を多群に拡張した (5 節)
5. 判別方向ベクトルに基づく特徴選択手法を新たに提案した (5 節)
6. 数値実験および実際のマイクロアレイデータへの適用を通して、提案手法の有効性を確認した (6 節・7 節)

とまとめられる。

1 と 2 は、一般の分布族の下で得られたものであり、大変強い結果である。多群におけるこのような角度・内積を用いた結果は、他に類を見ない理論的結果であり、高く評価されるものである。また、3 は正規性を仮定するが、多群の下で導出されており、2 群での既存の結果を含むものとなっている。多群の下での煩雑な計算と漸近上界の表現を、申請者は見通し良くまとめている。4 と 5 は、多群の下での特徴選択であり、先行研究は不十分なものがほとんどであった中、申請者は判別手法に忠実でありかつ包括的な特徴選択手法を構築していると言える。6 に挙げたように、本申請論文で提案された多群のパターン認識手法は、高次元小標本に対し有効に機能することが確認された。

このように、申請者のこれまでの一連の研究および本申請論文にまとめられた研究内容は、高い独創性を含む確立された研究であると言える。本申請論文は掲載済論文 2 篇を関連論文としており、申請者の活発な研究活動の成果である。

以上の理由により、本申請論文は博士（理学）の学位授与に相応しいと判断した。