

平成 22 年 4 月 26 日現在

研究種目： 基盤研究 (B)
研究期間： 2007 ~2010
課題番号： 19300096
研究課題名 (和文) 超高次元データの分類手法の導出とその理論的性質の解明および実データへの応用の研究
研究課題名 (英文) Research of classification methods of hyper-spectral data, elucidation of the theoretical nature and applications to real data
研究代表者
西井 龍映 (NISHII RYUEI)
九州大学・大学院数理学研究院・教授
研究者番号： 40127684

研究代表者の専門分野： 総合領域
科研費の分科・細目： 情報学・統計科学
キーワード： 分類・パターン認識, カテゴリ分解, モデル選択

1. 研究計画の概要

近年の情報化や計測技術の発展により、従来得られなかった超高次元のデータが観測されるようになり、有効な解析手法が求められている。たとえば特徴変数の次元が 10 個前後のデータを念頭においた従来の統計解析手法をそのまま応用しても十分な解析性能は保証されない。ここでは判別分析、および関連する分野に重きを置いて次の研究を行う。

- (1) 超高次元データの判別手法を提案し、実データでその性能を検証することを研究目的とする。また標本が複数のカテゴリに含まれている場合、そのカテゴリ比率を推定する問題： カテゴリ分解についても考察する。特に空間データの場合、空間相関をモデル化してカテゴリ分解の精度を向上させる手法について考察し、実データで検証する。
- (2) 時系列データの統計解析、およびモデル探索・選択について考察する。
- (3) 高次元データの統計解析やその環境問題への応用について考察する。

2. 研究の進捗状況

(1) 空間データのカテゴリ分解については、1) スペクトルの混合正規性、2) 被覆割合が与えられるとスペクトルは独立、3) 被覆割合はマルコフ確率場に従う という仮定のもとで被覆割合を推定する手法を考察した。局所事後密度を最大にするカテゴリ比率を導出し、ICM 法によって画像全体のカテゴリ分解カテゴリ分解を可能とした。教師データが全くない場合 (un-supervised) や、少ない場合 (semi-supervised) を考察し、論文を

発表した。

(2) 企業との共同研究によって時系列データの目的変数が閾値を越えるか否かの判別問題を考察した。これは回帰モデルとしてとらえたほうが良い予測となることがわかり、時系列回帰モデルの変数選択について議論した。また特許出願を予定している。

(3) 研究分担者や連携研究者はそれぞれ高次元データ (関数データ) の統計解析 (判別分析, 主成分分析, ブースティング, 環境科学への応用等) についてそれぞれ研究を推進し、多くの研究成果を発表している。

3. 現在までの達成度

② おおむね順調に進展している。

本研究グループはすでにパターン認識やブースティング、および関連する統計的問題を取り扱った論文を数多く発表した。また関連する工学的問題についても考察を始め、結果を出しつつある。

4. 今後の研究の推進方策

(1) 超高次元問題へ適応できる手法への改良を行っていて、現在発表準備中である。高速な最適化、母数の決定則の導出、が現在の課題となっている。

(2) 時間とともに母数に変化する時系列モデルの推定法、およびその工学的応用について議論していく。

(3) 研究分担者や連携研究者と情報交換を密にし、高次元データの解析手法、および実

データへの応用研究を推進する.

5. 代表的な研究成果

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 44 件)

- ① Uchida, M. (2010). Contrast-based information criterion for ergodic diffusion processes from discrete observations, *Annals of the Institute of Statistical Mathematics*, **62**(1), 161-187.
- ② Tanaka, S. and Nishii, R. (2009). Non-linear regression models to identify functional forms of deforestation in East Asia. *IEEE Transactions on Geoscience and Remote Sensing*, **47**(8), 2617-2626.
- ③ Kawakita, M. and Eguchi, S. (2008). Boosting method for local learning in statistical pattern recognition. *Neural Computation*, **20**(11), 2792-2838.

[学会発表] (計 59 件)

- ① Nishii, R., Sawamura, T. and Ozaki, T. Semi-supervised contextual unmixing of geospatial data. IASC 2008, Pacifico Yokohama, Japan (招待講演), 2008. 12. 07.

[図書] (計 6 件)

- ① Konishi, S. and Kitagawa, G. (2009). *Information Criteria and Statistical Modeling*. Springer. (273 頁)