

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 5 月 29 日現在

機関番号：37104

研究種目：若手研究(B)

研究期間：2010～2012

課題番号：22700298

研究課題名（和文） 医学研究における関数化データに基づく統計モデルの構築と適用

研究課題名（英文） Functional data analysis for medical data and its application

研究代表者

荒木 由布子 (ARAKI YUKO)

久留米大学・バイオ統計センター・助教

研究者番号：80403913

研究成果の概要（和文）： 各個体に観測点が多数あり高次元で、かつ複雑な構造を有する医学データから有用な情報を引き出すことを目的とし、統計モデルの理論・方法論の開発研究と応用研究に取り組んだ。医用画像や脳波、形状データなどの高次元データには関数データ解析の枠組みから次元縮小法を提唱し、さらに判別モデルとその評価法を提唱した。また、非線形モデルにおいて情報量規準から効率的にモデルを評価する手法を提唱した。提唱したモデルは生命科学における問題解決に適用しその有用性を示した。

研究成果の概要（英文）： To extract valuable information from medical data which has a lots of observational points result in high dimensional data with complex structure, we developed statistical methods and applied them to the analysis of medical data. We proposed functional data analytic approach for dimension reduction of high dimensional data such as medical imaging data, brain waves or shape data. We further proposed classification model for high dimensional data and its evaluation methods. We also developed an efficient nonlinear model selection method based on information criterion. The proposed methods were applied to life science problems.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2010年度	900,000	270,000	1,170,000
2011年度	800,000	240,000	1,040,000
2012年度	700,000	210,000	910,000
年度			
年度			
総計	2,400,000	720,000	3,120,000

研究分野：総合領域

科研費の分科・細目：情報学・統計科学

キーワード：医薬生物、ゲノム統計解析、関数化データ解析

1. 研究開始当初の背景

近年の著しい科学技術の発展に伴い観測・測定機器も発展し、複雑多様な医学データが観測される時代となった。しかしこれらの多様化した最先端の医学データから、治療法を初めとする医療の発展に役立つ情報を有効に抽出するための統計手法の開発は進

んでいなかった。本研究では多種多様化した医学データの中でも特に、各個体が多地点/多時点において観測され生成された変数を含む高次元データに注目した。

多点観測データの例として、例えば、児童へ言語的刺激を与えた時の前頭葉の働きを捉えるための観測は、各児童の頭部に 24 か

所のセンサーをおき、各センサーにおける数千時点での脳波の観測からデータが構成される。冷却装置により体温の変化を調査するための体感温度データは、24時間にわたり毎分数回の頻度で観測される。また、MRIで観測される人間の3次元脳画像データは、ボクセルと呼ばれる最小単位が各個体につき約100万個から形成される。

これらは莫大な情報を含む数十～数百万点からなる高次元データであるが、一般的にはデータの用量を減らし扱いを簡便にするため、データの一部だけを取り出して少数のスカラ値で要約するのが慣習的な方法で、データの持つ情報が大幅に失われていた。

2. 研究の目的

このような背景の中で、本研究では観測点が多数ある高次元医学データ全体から有益な情報を抽出するための統計モデルを理論・適用の両面から研究開発し、実際の医学研究における問題解決へ役立てることを目的とした。

これまでに研究代表者は関数データに基づく一般化線形モデルや関数データを判別するための関数データ判別モデルを構築し、構築したモデルを評価するためのモデル評価規準を導出してきた(Araki and Konishi 2004, 2006, Araki *et al.* 2009(a), (b)).

しかし実際の医学データには特有の性質があり、データから情報を正しく引き出すためにはそのデータの性質に適した統計モデルが必要である。例えば、共変量には2値変数やカテゴリカル変数、または連続変数など様々なタイプの変数が混在する。また、データからの情報を的確に捉えられない統計モデルからの結論は、医療の現場に誤った見解を与えることになってしまうため、統計モデルを適用した際にモデル評価を行う必要がある。

3. 研究の方法

- (1) 本研究では、共変量の一つとして関数データ解析の枠組みで高次元データを捉え、その他の離散型や連続型の変数も併せて共変量に取り込み、さらにモデルを評価・選択するための規準を求めた。
- (2) 共変量と疾病発症との非線形な関係構造を捉えるとき、非線形判別モデルは柔軟に判別関数を構成することが可能であり、推定量は正則化パラメータに依存する。情報量規準による正則化パラメータの選択法が非常に有用である事が知られているが(Konishi and Kitagawa 2008), 複数の共変量がある場合には効率的な選択法が必要となる。本研究では最適な正則化

パラメータを情報量規準により選択するときの効率的な方法を提案し、提案法により胃がん患者の肝転移発現と共変量との非線形な関連性を捉える。

4. 研究成果

(1) 多点観測データの1つにMRIにより観測される約百万次元の3次元脳画像がある。この超高次元データから局所脳萎縮を客観的に評価することでアルツハイマー病の早期発見が可能となる。従来は解析可能なデータ量とするため、関心領域を特定した評価が行われてきた(ROI)。本研究では脳画像全体を用いた超高次元データを関数データ解析の観点から、基底展開法と主成分分析法を組み合わせて次元縮小を行い、非線形ロジスティック判別モデルにより疾病の判別関数を構築した。この際、医学データでは個体差を考慮する必要があるため、2値、連続型、離散型と様々なタイプの共変量を含む判別モデルを構築した。また、最適なモデルを構築するための正則化パラメータの選択法をいくつかの規準を数値実験により比較検討したうえで定め、実データへ適用して構築したモデルの有用性を示した。

(2) 複雑な非線形関係を捉えるために有用な非線形一般化回帰モデルの推定量は正則化パラメータに依存するため、この選択は極めて重要である。一般化情報量規準や一般化ベイズ型情報量規準が有用である事が示されているが、複数の正則化パラメータを決定する場合、従来のグリッドサーチによる探索は計算負荷が膨大になる。そこで、複数の正則化パラメータを情報量規準に基づき効率的に選択する方法をチェインルールとニュートン法をもとに提案し、胃がん患者の肝転移の有無に関する実データの問題解決へ役立てた。

5. 主な発表論文等

(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8件)

- ① Araki, Y., Kawaguchi, A., Yamashita, F. Regularized logistic discrimination with basis expansions for the early detection of Alzheimer's disease based on three-dimensional MRI data. *Advances in Data Analysis and Classification*, 査読有, 7, 2013, 109-119.
DOI:10.1007/s11634-013-0127-5

Araki, Y. and Hattori, S. Efficient regularization parameter selection via information criteria. *Communications in Statistics, Simulation and Computation*, 査読有, 42(2), 2013, 280-293. DOI: 10.1080/03610918.2011.639969

- ② 荒木由布子, 川口淳. 高次元脳画像解析のための二段階正則化関数判別モデル. 応用統計学会年会講演予稿集, 査読無, 2012, 33-38.
- ③ 蔵本昭孝, 堀大蔵, 井上茂, 堀之内崇士, 品川貴章, 上妻友隆, 下村卓也, 河田高伸, 林龍之介, 嘉村敏治, 山下拓人, 荒木由布子, 角間辰之. 低置胎盤の術中出血量に影響を及ぼす因子についての検討. *日本周産期・新生児医学会雑誌*, 査読有, 47(4), 2011, 873-877.
- ④ 川口淳, 荒木由布子, 山下典夫. 関数データ解析の3次元脳画像解析への応用. *日本計算機統計学会第25回講演論文集*, 査読無, 2011, 85-88.
- ⑤ 荒木由布子, 川口淳, 山下典夫. 3次元脳画像によるアルツハイマー病早期発見のための関数データ解析. 応用統計学会年会講演予稿集, 査読無, 2011, 67-72.
- ⑥ Nagamitsu S, Araki Y., et al. (14人中2番目). Prefrontal brain function in children with anorexia nervosa: A near-infrared spectroscopy study. *Brain & Development*, 査読有, 33, 2011, 35-44. DOI:10.1016/j.braindev.2009.12.010.
- ⑦ Nagamitsu S, Yamasita F, Araki Y., et al. (11人中3番目). Characteristic prefrontal blood volume patterns when imaging body type, high-calorie food, and mother-child attachment in childhood anorexia nervosa: A near infrared spectroscopy study. *Brain & Development*, 査読有, 32, 2010, 162-167. DOI:10.1016/j.braindev.2009.01.002
- ⑧ Iizuka C, Yamashita Y, Nagamitsu S, Yamashita T, Araki Y, Ohya T, Hara M, Shibuya I, Kakuma T, Matsuishi T. Comparison of the strengths and difficulties questionnaire scores between children with high-functioning autism spectrum disorder (HFASD) and

attention-deficit/hyperactivity disorder (AD/HD). *Brain & Development*, 査読有, 32, 2010, 609-612. DOI: 10.1016/j.braindev.2009.09.009.

[学会発表] (計10件)

- ① 荒木由布子. 二段階正則化に基づく生存時間解析法と脳画像データへの適用. 2012年度統計関連学会連合大会, 2012年9月11日, 北海道大学.
- ② Araki Y. Two-Way Regularized Functional Classification for High Dimensional Brain MRI Data. 26th International Biometric Conference, 2012年8月30日, Kobe International Conference Center.
- ③ 荒木由布子. 正則化関数データモデリングとその適用. 第7回 Biostatistics Network 特別講演. 2012年8月8日, 統計数理研究所.
- ④ 荒木由布子. 高次元脳画像解析のための二段階正則化関数判別モデル. 応用統計学会2012年度年会, 2012年5月24日, 統計数理研究所.
- ⑤ Araki, Y. The analysis of high-dimensional brain image via sparse principal component functional regression. East Asia Regional Biometric Conference 2012. February 2-3, 2012, Seoul National University, Seoul, Korea.
- ⑥ Araki, Y. The early detection of Alzheimer's diseases based on MRI data via sparse functional logistic classification. Joint Meeting of the 2011 Taipei International Statistical Symposium and 7th Conference of the Asian Regional Section of the IASC. December 16-19, 2011, Academia Sinica, Taipei, Taiwan.
- ⑦ 荒木由布子. 認知症早期発見のための三次元脳画像に基づく関数データ解析, 2011年度統計関連学会連合大会, 2011年9月4-7日, 九州大学伊都キャンパス.
- ⑧ 荒木由布子. 三次元脳画像によるアルツハイマー病早期発見のための関数データ解析, 応用統計学会2011年度年会, 2011年6月3-4日, 大阪大学吹田キャンパス銀杏会館ホール.
- ⑨ 川口淳. 関数データ解析の三次元脳画像解析への応用. 日本計算機統計学会第25回大会, 2011年5月7-8日, 函館市亀田福祉センター.

- ⑩ Araki, Y. Regularized functional classification of the kinetic trace of developmental mouse retina. The 20th Australian Statistical Conference, December, 2010, Fremantle, Australia.

〔図書〕（計1件）

柳川堯, 荒木由布子, 近代科学社, バイオ統計の基礎, 2010, 234-254.

6. 研究組織

(1) 研究代表者

荒木 由布子 (ARAKI YUKO)

久留米大学・バイオ統計センター・助教

研究者番号：80403913