

平成 30 年 6 月 23 日現在

機関番号：13801

研究種目：若手研究(B)

研究期間：2014～2017

課題番号：26730023

研究課題名(和文) 高次元医用データのための関数データに基づく統計モデルの開発

研究課題名(英文) The development of statistical models based on function data for the high-dimensional medical data

研究代表者

荒木 由布子 (Araki, Yuko)

静岡大学・情報学部・准教授

研究者番号：80403913

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：本研究では、超高次元データの情報を落とさず次元縮小し情報抽出するための統計モデル開発に取り組んだ。医用画像データから疾病発症を早期診断するための判別モデルを構築した。さらに脳検診データと身体活動度、臨床的評価項目の関連性分析のためのモデルを構築し変数間の機序を明らかにした。また、健康長寿社会の実現を目指した全国長期大規模追跡調査データの一部にスパース性を仮定した生存時間解析法を適用し、予測精度を考慮しつつ予後因子の特定に取り組んだ。以上の研究成果は国内・国際学会で発表した。関数データ構造方程式モデルの開発にも取り組んだが、推定値の性質に関して検討の必要があり未完成である。

研究成果の概要(英文)：In this research, we have developed statistical models to extract useful information from super-high dimensional data without loss of information by creating some new dimension reduction techniques. We constructed a classification model which detects onset of some disease on early stage based on MRI data. Further, we elucidated a mechanism between brain structure, levels of activity and clinical endpoints. In addition, the survival model with sparse constraints were applied to the long term and large sample size follow up data aimed at realizing health longevity society in Japan. All the results described above were presented at both domestic and international conferences. We had further worked out for constructing functional structural equation modeling(SEM). However, since there were some issues which need further consideration regarding to characteristics of estimator, the SEM modeling was still developing.

研究分野：統計科学

キーワード：関数データ解析 高次元データ 判別モデル 情報量規準 非線形モデル 多変量データ解析 スパース推定 構造方程式モデル

1. 研究開始当初の背景

本研究では、近年の著しい科学技術の発達に伴い複雑多様化したデータの中でも特に、医用画像データに代表される、複雑な構造を持つ高次元医用データに注目する。例えばアルツハイマー病診断のため撮影される脳磁気共鳴(MRI)データや PET、CT などの医療画像データは複雑な構造を持つ高次元データで、脳 MRI データは各個体が約 100 万個のボクセルと呼ばれる立方体の集まりから脳構造を捉える。このような高次元医用データは、観測の時間的・経済的コストの問題から標本数が次元に比べて圧倒的に少ない場合が多く、通常の統計モデルではモデルのパラメータ推定において安定した汎用性の高い推定値を得ることが困難となる。高次元小標本問題は近年の統計学の課題であり、特に臨床研究ではその扱いの困難さから、部分データや単純な記述統計量が解析に用いられ、大幅に情報量が削減されたデータに従来の統計モデルが適用されているのが実態である。

2. 研究の目的

そこで研究代表者は、高次元医用データから有効に情報を抽出するための統計モデルを提唱する。まず、高次元データの情報を落とさずに次元縮小することで、ロジスティック回帰モデルやグラフィカルモデルなどの統計モデルの適用を可能とする。高次元データの次元縮小において近年主流となっているのは、スパース推定 (Tibshirani 1996, Zou 2006) である。しかしこの方法が有効に機能する次元は数百程度であり、本研究で想定する医用データは遥かに高次元であるため直接適用できない。事前に変数を絞り込む方法も提唱されているが (SIS, Fan and Lv 2008)、画像の場合には現実的な解釈が困難である。

研究代表者はこれまで基底展開を用いた関数データ解析と呼ばれる方法に基づく一般化線形モデル (Araki et al. 2009a) や判別モデルを構築し、それぞれのモデル選択基準を導出してきた。提案したモデルはマイクロアレイで観測された周期性を持つ継時的なイースト菌の遺伝子発現データの解析へ適用し、モデルの有用性を検証した (Araki et al. 2009b)。H22-H24 若手研究 (B)、「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」(研究代表者)の研究では、関数データ解析の一部である基底展開法のモデル選択に関する正則化パラメータの選択を効率的に情報量規準に基づき実行する方法を提案し、胃がん転移データの解析へ適用した (Araki and Hattori 2013)。また、高次元脳 MRI データと年齢などの患者背景因子から、アルツハイマー病に関する脳の委縮の部位を特定し、疾病を早期発見するための方

法を提唱した。上記のこれまでの研究成果から、関数データ解析とよばれる基底展開法を含む統計モデルが高次元医用データを扱うのに非常に有用である事に気が付いた。そこで本研究では、これまでの研究成果を次の様に発展させることとした。

(ア) グラフィカルモデルで各脳部位の関連性の構造を捉え、ロジスティックモデルでそれらと疾病の関係を捉える。

(イ) スパース性を仮定した正則化多変量解析と基底展開を組み合わせた、合成基底関数展開による次元縮小を提案する。

(ウ) 統計グラフィックの有効活用でデータと解析結果の解釈を容易にする。

超高次元小標本問題を伴う医用データから有効に情報を抽出する統計手法の研究は、世界でも近年開始されてばかりである。高次元データ解析の主流となっているスパース推定に関する理論的・実践的研究は数多いが、その多くは次元が数百までの限界があり、目的変数と説明変数を直接的に関連付けるモデルである。一方、本研究で開発するモデルは超高次元となる医用画像データを絞り込みなしで扱えると同時に汎化能力が高く、更に変数間の関連性構造を捉えた上で疾病との関係をモデル化する点も、共変量と疾病の因果関係が脳の構造を仲介したものであるか調べる点も、独創的であると同時に、学術的に高い意味を持つと考えられた。

3. 研究の方法

(ア) について：(イ) で構築する合成基底関数を用いて、高次元データを変数に持つロジスティックモデルと構造方程式モデルを開発する事に努めた。ここで、第 1 点目として、これまでのロジスティックモデルを用いた研究では説明変数が直接目的変数へ影響を与えるという仮定を用いていたが、例えば脳画像データでは、各脳部位がお互いに影響し合っていると仮定するほうが自然である。このため、グラフィカルモデルを用いてその構造を捉え、それらと疾病との関係を表すロジスティックモデルを開発しようと考えた。実際には、ロジスティックモデルの未知パラメータ推定にスパース推定を取り入れることで直接的な影響を捉えることが可能である事がわかり、ロジスティックモデルへ尽力した。それに伴い、間接的効果や共変量間の構造を捉えるための構造方程式モデルの構築は、推定量の性質の保証に検討の余地があり、研究期間内では完成しなかった。

第 2 点目として、パラメータの推定にスパースな制約をつけることで、更なる次元縮小とパラメータ推定を同時に行うことを可能とする。スパースな制約の量を定める正則化パラメータの選択法を情報量規準の観点か

らも考慮したが、推定量の非連続性などから、想定していた情報量規準を導出することは困難であったため、想定外の方法で正則化パラメータを考慮した。

(イ)について：スパース性を仮定した正則化多変量解析と基底展開を組み合わせた、合成基底関数展開による次元縮小を提案した。Araki et al. 2013 では通常の主成分分析 (PCA) と基底展開を組み合わせてロジスティック判別に用いたが、PCA は全変数の線形結合で一つの主成分を構築するため、判別に影響のある変数 (ボクセル) を特定するにはボクセルごとの信頼区間を構築する必要があり、計算量が膨大であった。そこで、スパースな制約をつけた PCA (SPCA, Shen and Huang, 2008) を次の様に発展させて用いる事を考えた。(1) Shen and Huang, 2008 らの SPCA は対象としているデータの次元が医用画像データなどの超高次元ではないため、直接適用しても機能しない。そこで、SPCA に基底展開を組み合わせた合成基底関数を用いることを提案する。この方法により、(ア)における目的変数に影響のある部位をパラメータ推定と同時に特定する方法を考えた。(2) SPCA のスパース性の制約の量を決める正則化パラメータの選択問題に情報量規準の観点で取り組み、汎用性の高いモデル構築を目指した。(3) スパース推定により 0 となる係数がパラメータ推定と同時に特定されるが、パラメータの推定値の推論を構築することで、医学データ解析のエビデンスを与えた。

(ウ)について：高次元データは、データ構造の把握も解析結果の解釈も困難である。これを医師など統計家以外の研究者にもわかりやすくするため、データ構造の把握と解析結果の可視化を可能とするための統計グラフィックスの有効活用と開発を目指したが、開発途中である。

4. 研究成果

高次元医用データから有益な情報を効率的に抽出するための統計モデルの開発を目的とした当該研究の研究成果を以下にまとめる。

本研究では、高次元データの情報量を落とさずに次元縮小しつつ、ロジスティック回帰モデル、グラフィカルモデルといった統計モデルの適用での情報抽出のためのモデル開発に取り組んだ。各被験者の観測に約 100 万個のボクセルが含まれる sMRI データは医用画像データの中でも特に高次元であることで知られるが、この超高次元データから認知症発症を早期発見するための判別モデルを構築した。

このモデルへの取り組みは、まず滑り出しをスムーズにする目的から、Advances in Data Analysis and Classification 誌に掲載

された Araki et. al. (2013) のモデルを発展させて、スパースな制約を課した非線形ロジスティック判別モデルを構築した。このモデルでは、スパース性を仮定した正則化多変量解析法と基底展開法を組み合わせることで、情報量の損失をおさえた次元縮小が可能になるというアイデアを加えたモデル拡張を行った。次元縮小のアイデアは基底関数、正則化項、モデル選択法それぞれ複数ある手法の組み合わせでどれが最も予測誤差を小さくするか数値実験を行い吟味した。また提案モデルの予測の性能を、感度特異度の観点から数値実験で他の方法と比較検討し、その有用性を検証した。また、提案法をオープンデータに適用し、認知症発症を予測するのに有用であることが示唆された。本結果は国際会議で発表を行い、論文投稿準備中である。

脳検診データと身体活動度、記憶障害、臨床的評価項目の関連性を分析するためのモデルにおいては、健診データへ適用し、変数間の機序を明らかにした。さらに、健康長寿社会の実現を目指した全国大規模追跡調査データから有益な情報を抽出し今後の政策に役立てるため、パイロットスタディとして 10 年間の追跡調査の一部のデータにスパース性を仮定した生存時間解析法を適用し、予測精度を検討するとともに予後因子の特定に取り組んだ。この結果はシンポジウムと国際疫学会で発表し、論文投稿準備中である。

さらに、超高次元共変量を含む媒介分析を、上記の次元縮小の工夫を行い関数化した関数データをモデルに含めた関数データ構造方程式モデルの開発に取り組んだ。このモデルにより、地点や時点の変化に伴い変動する変数の目的変数への直接効果、間接効果を捉える可能性が示唆されたが、推定値の一意性の問題があり、研究期間内には完成しなかった。今後の展望として、本研究期間中のモデリングでは、最適なモデルを選ぶためのモデル選択規準の導出が未完成であったため、今後はこの点を十分に検討し開発を進めていく必要がある。また、時間や空間の変動を有する情報からイベント発生までの時間を予測するための統計モデルの開発が、疾病や放射線量暴露の影響を分析する際に必要とされている。これまで培ってきた次元縮小の技術を元に生存時間解析モデルの構築が可能であるため、取り組んでいく予定である。さらに、本研究期間内では未完成である関数データ構造方程式モデルの開発を引き続き行う。

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10件)

- [1]. Kennosuke Tsuda, Takeo Mukai, Sachiko Iwata, Jun Shibasaki, Takuya Tokuhisa, Tomoaki Ioroi, Hiroyuki Sano, Nanae Yutaka, Akihito Takahashi, Akihito Takeuchi, Toshiki Takenouchi, Yuko Araki, Hisanori Sobajima, Masanori Tamura, Shigeharu Hosono, Makoto Nabetani, Osuke Iwata. Therapeutic hypothermia for neonatal encephalopathy: a report from the first 3 years of the Baby Cooling Registry of Japan. Scientific Reports 7/39508 - (2017年) [査読] 有 DOI:10. 1038/srep39508
- [2]. Hiroshi Yao, Yuko Araki, Yuki Takashima, Akira Uchino, Takefumi Yuzuriha, Manabu Hashimoto. Chronic Kidney Disease and Subclinical Brain Infarction Increase the Risk of Vascular Cognitive Impairment: The Sefuri Study Journal of Stroke and Cerebrovascular Diseases 26/2 420-424 (2017年) [査読] 有 DOI:10.1016/j.jstrokecerebrovasdis. 2016.10.002
- [3] Manabu Hashimoto, Yuko Araki, Yuki Takashima, Kohjiro Nagami, Akira Uchino, Takefumi Yuzuriha, Hiroshi Yao. Hippocampal atrophy and memory dysfunction associated with physical inactivity in community-dwelling elderly subjects: The Sefuri study. Brain and Behavior 2016 29;7(2)/e00620 [査読] 有 DOI: 10. 1002/brb3. 620
- [4]. Tanaka Shoichiro, Iwata Sachiko, Kinoshita Masahiro, Tsuda Kennosuke, Sakai Sayaka, Saikusa Mamoru, Shindo Ryota, Harada Eimei, Okada Junichiro, Hisano Tadashi, Kanda Hiroshi, Maeno Yusuki, Araki Yuko, Ushijima Kazuo, Sakamoto Teruo, Yamashita Yushiro and Iwata Osuke. Use of Normothermic Default Humidifier Settings Causes Excessive Humidification of Respiratory Gases during Therapeutic Hypothermia. Therapeutic Hypothermia and Temperature Management 6/4 180-188 (2016年) [査読] 有 DOI: 10. 1089/ther. 2016. 0005
- [5]. Masahiro Kinoshita, Sachiko Iwata, Hisayoshi Okamura, Mamoru Saikusa, Naoko Hara, Chihoko Urata, Yuko Araki & Osuke Iwata. Paradoxical diurnal cortisol changes in neonates suggesting preservation of foetal adrenal rhythms. Scientific Reports 6/35553 - (2016年) [査読] 有 DOI:10. 1038/srep35553
- [6]. Sachiko Iwata, Reiji Katayama, Masahiro Kinoshita, Mamoru Saikusa, Yuko Araki, Sachio Takashima, Toshi Abe & Osuke Iwata. Region-specific growth restriction of brain following preterm birth. Scientific Reports 6/33995 - (2016年) [査読] 有 DOI: 10. 1038/srep33995
- [7]. Tsuyoshi Kurata, Sachiko Iwata, Kennosuke Tsuda, Masahiro Kinoshita, Mamoru Saikusa, Naoko Hara, Motoki Oda, Etsuko Ohmae, Yuko Araki, Takashi Sugioka, Sachio Takashima & Osuke Iwata. Physiological and pathological clinical conditions and light scattering in brain. Scientific Reports 6/31354 - (2016年) [査読] 有 DOI: 10. 1038/srep31354
- [8]. Yao H, Takashima Y, Araki Y, Uchino A, Yuzuriha T, Hashimoto M. Leisure-Time Physical Inactivity Associated with Vascular Depression or Apathy in Community-Dwelling Elderly Subjects: The Sefuri Study. J Stroke Cerebrovasc Dis. 11/ 2625-2631 (2015年) [査読] 有 DOI: 10. 1016/j. jstrokecerebrovasdis. 2015. 07. 018.
- [9]. Fujiyoshi N, Ushijima K, Kawano K, Fujiyoshi k, Yamaguchi T, Araki Y, Kakuma T, Watanabe S, Kaku T, Nishida T, Kamura T. Radiation effects on DNA content of cervical cancer cells: A rapid evaluation of radiation sensitivity by laser scanning cytometry. Molecular and clinical Oncology 3/ 51- 54 (2015年) [査読] 有 DOI: 10. 3892/mco. 2014. 413
- [10]. 大畑千佳, 濱田尚宏, 小野文武, 荒木由布子, 角間辰之, 橋本隆. 魚鱗鱗におけるリペアクトの臨床効果の検討. 薬理と治療 43/(2) 175- 180 (2015年) [査読] 有 DOI:なし
- [学会発表] (計23件)
- [1] Araki, Y. Functional path analysis with composite basis expansions. Organized invited session “Advanced methods in Biostatistics”, 10th International conference on Computational and Methodological Statistics 2017, 2017.
- [2] Araki, Y et. al. “Risk factors for the onset of long-term care need in community-dwelling older people in Japan: 10year follow-up study of participants in the JAGES project. ”. World Congress of Epidemiology 2017.
- [3] 荒木由布子. MRI データに基づくアルツハイマー病早期発見のための統計的モデリング, 生物統計・疫学公開講演会, 2017年.

- [4] Takahashi T, Kamada T, Araki Y, Ozeki K, Okada E and Ojima T. Simulation of Optimum selection of pharmacies during a disaster in Hamamatsu City. 第22回静岡健康・長寿学術フォーラム, 2017年.
- [5] Nakamura Y, Yasuhira S, and Araki Y. Bayesian estimation of Birth rate. 第22回静岡健康・長寿学術フォーラム, 2017年.
- [6] Ohashi J, Nasu A, Tanabe Y, Yao H and Araki Y. Comparison of the effects on analysis results by imputation of missing values in the cross-sectional study of community-dwelling elderly people in Sefuri village. 第22回静岡健康・長寿学術フォーラム, 2017年.
- [7] 荒木由布子. MRIデータに基づくアルツハイマー病早期発見のための統計的モデリング, BIO tech 2017 アカデミックフォーラム, 2017年.
- [8] 八尾博史, 荒木由布子, 高橋由紀, 内野晃, et al. 「地域在住一般住民における認知機能評価のための多角的アプローチ」, 日本脳ドック学会総会, 2017年.
- [9] 八尾博史, 荒木由布子, 高島由紀, 内野晃, 杠岳文潜在性脳梗塞に対する保護因子としての出産数: 地域在住の一般住民における検討. 第42回日本脳卒中学会総会 2017年.
- [10] 荒木由布子, 岡田 栄作, 尾島 俊之, 近藤 克則. 大規模追跡調査に基づく要介護認定リスクモデルの提案. 科研費シンポジウム「複雑な生命現象を読み解くための大規模データ解析とモデリング」2016年.
- [11] 荒木由布子, 岡田 栄作, 近藤 克則, 尾島 俊之. 10年間追跡調査に基づく要介護認定リスクモデルの探索的検討. JAGES研究会. 2016年.
- [12] Mena, M. , Terada, A., Yao, H. , Araki, Y. Analysis of risk factors of silent brain infarction using the Bayesian Network: the Sefuri Study. 第21回静岡健康・長寿学術フォーラム. 2016年.
- [13] Shimizu, T., Ishikawa, Y., Yao, H. , Araki, Y. Propensity score analysis of causal effects of silent brain infarction on cognitive dysfunction: the Sefuri Study. 第21回静岡健康・長寿学術フォーラム. 2016年.
- [14] Kamada, T., Okuno, S., Araki, Y , Okada, E., Ojima, T. Regional comparison of risk factors related to the onset of certification of long-term care need in community-dwelling elderly people in Japan: Follow-up study of participants in the JAGES project. 第21回静岡健康・長寿学術フォーラム. 2016年.
- [15] 八尾博史, 目名まりあ, 荒木由布子. 潜在性脳梗塞における生物学的性差: 地域在住の一般住民における検討. 第39回日本高血圧学会総会. 2016年.
- [16] Araki, Y. and Kawaguchi, A. Sparse functional classification method with composite basis function for early detection of Alzheimer's disease based on brain MRI. The XXVIIIth International Biometric Conference. 2016年.
- [17] Yao H, Araki, Y., Takashima Y, Nogami K, Uchino A, Yuzuriha T, Hashimoto M. Physical inactivity associated with vascular depression or apathy leads to hippocampal atrophy and memory dysfunction in community-dwelling elderly subjects: The Sefuri study. International Stroke Conference 2016, 2016.
- [18] Eisaku Okada, Yuko Araki, Katsunori Kondo, Hiroshi Hirai, Toshiyuki Ojima. Elucidation of factors related to the onset of certification of long-term care need in community-dwelling elderly people in Japan: 10-year follow-up study of participants in the JAGES project. 第20回静岡健康・長寿学術フォーラム, 2016年.
- [19] Araki Y. Intrinsic feature extraction via functional data analysis. 8th International Conference of the ERCIM WG on Computational and Methodological Statistics. 2015.
- [20] 荒木由布子, 岡田栄作, 近藤克則, 尾島 俊之. 「10年間追跡調査に基づく高齢者の要介護認定リスク因子の探索的検討」. 科研費シンポジウム「生命科学データ解析の方法論と健康科学への応用」, 2015年.
- [21] 荒木由布子. Statistics modeling for high dimensional data and its application in Biostatistics ~Functional sparse logistic discrimination via composite basis expansions~, 第8回東海地区生物統計ネットワークセミナー, 2015年.
- [22] 荒木由布子, 高次元データの統計的モデリング, 2015年度 第1回情報学イブニングセミナー「生命, 脳, 音声から探る情報学」, 2015年.
- [23] Yuko Araki and Atsushi K. Two-step regularized Cox proportional hazard model for high-dimensional brain MRI data. 27th International Biometric Conference, 2014.
- [図書] (計1件)

柳川堯・荒木由布子. 近代科学社. 『バイオ統計シリーズ1 バイオ統計の基礎-医薬統計入門-』(258ページ)
第5版 2018年
Kindle 版 2016年
(2016年 シリーズ (第1巻～第6巻) で第9回日本統計学会出版賞受賞)

○出願状況 (計 0 件)

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況 (計 0 件)

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

[その他]

ホームページ等

6. 研究組織

(1) 研究代表者

荒木 由布子 (YUKO, Araki)
静岡大学・情報学部・准教授
研究者番号：80403913