

科学研究費助成事業 研究成果報告書

平成 28 年 5 月 30 日現在

機関番号：32641

研究種目：挑戦的萌芽研究

研究期間：2013～2015

課題番号：25610035

研究課題名(和文) 汎関数に基づく推測理論とL1型正則化モデリングの研究

研究課題名(英文) Inference via functional theory and L1 regularization modeling

研究代表者

小西 貞則 (Konishi, Sadanori)

中央大学・理工学部・教授

研究者番号：40090550

交付決定額(研究期間全体)：(直接経費) 2,800,000円

研究成果の概要(和文)：科学の様々な分野，産業界において獲得されるデータは，大規模化，高次元化し，これに伴い新たな解析手法の開発研究が急速に進展してきた．本研究では，モデリングの研究開発と関連する理論研究に取り組み，次のような研究成果を挙げた．(1) L1 ノルム制約を課した正則化法をベイズ的視点で捉え，ベイズ型スパースモデリングを提案した．(2) 非線形構造をモデル化した時変型回帰係数モデルや関数混合効果モデルなど，複雑かつダイナミックな様相を呈する非線形現象の解析に有効な手法を提案した．(3) 回帰モデルなどに対してスパースモデリングの理論的研究を推進し，汎関数理論に基づくモデル評価基準について研究した．

研究成果の概要(英文)：In various fields of science and industry a huge amount of data with complex structure and/or high-dimensional data have been accumulating. The effective use of these data sets requires new modeling strategies in order to perform extraction of useful information and knowledge discovery. Through this research we have investigated the problem of analyzing such datasets, and proposed various statistical modeling strategies: (1) Various regularization methods with L1 norm penalty have been proposed for effective regression modeling from a Bayesian point of view. (2) For analyzing data with complicated structure or substantial longitudinal heterogeneity between subjects, we introduced a varying coefficient modeling and a nonlinear functional mixed modeling through the nonlinear regression approach. (3) We developed a general framework for constructing model selection criteria in the context of functional statistics.

研究分野：統計科学

キーワード：スパースモデリング ベイズモデリング 非線形モデリング 混合効果モデリング モデル評価基準
汎関数理論

1. 研究開始当初の背景

電子化された計測・測定技術の進歩は、生命科学におけるゲノムデータ、医用工学における高精細画像処理データなど、次元数が数千次元にもおよぶ極めて次元の高いデータの獲得と蓄積を可能としてきた。このような超高次元データに基づく回帰モデル、識別・判別モデルなどの構築に当たっては、従来手法によるモデルの推定、適切な変数の選択は困難であり、従って予測に有効なモデルを構築できないなど、数々の問題点があった。このような状況の中で回帰モデリング、特に、線形回帰モデルの推定と変数選択に新たな方向性を示したのが lasso (Tibshirani, 1996, JRSS B) であり、その特徴はモデルの推定と変数選択を同時に実行できる点にあった。このため、高次元データに基づく有効なモデル推定法として注目を集め、様々な L_1 型正則化線形回帰モデリングの研究が急速に進展しつつあった。

しかしながら、 L_1 型正則化線形回帰モデリングについては、次のような問題点が浮かび上がってきた。(1) 提案手法は、あたかも真のモデルは想定した線形回帰モデルの部分モデルであるとしているが、実際には線形性やデータの発生機構は真の現象構造の近似にすぎない。(2) L_1 型正則化線形回帰モデリングの最大のメリットは、データ数に比して次元の高いデータによるモデル化を可能とする点にあるが、その理論研究と実際問題への適用上の有効性の研究は十分でない。(3) モデリングに本質的な正則化項の影響を制御する調整(減衰)パラメータの選択に必要なモデル評価基準の導出と関連する理論研究が未解決問題として残っている。

そこで、これまでの研究成果を発展させて、データ数に比して次元の高いデー

タ集合に基づくモデリングの研究に理論的・実際の両側面から取り組み、新たな解析法の提唱を目指すことを計画した。

2. 研究の目的

本研究では、データ数に比して極めて次元の高いデータ集合から、背後に潜む有益な情報やパターンを高効率に抽出・処理するための新たなモデリングの理論・方法論の開発研究を目的とした。このため、超高次元データ解析に一つの方向性を提示し、国際的に研究が進展しつつある L_1 タイプのノルムを損失関数に付与する正則化法について、様々な現象解析に対処するため、新たなスパースモデリングの開発と理論・方法論の研究推進と、生命科学、システム工学、生物工学、地球環境科学など諸科学への応用研究を目的とした。また、モデリングで最も重要かつ難しい問題である推定したモデルの評価ステップの研究が極めて少ない点に着目し、適切なモデルを構築するには、 L_1 型正則化法の根幹をなす正則化項の影響を制御する正則化パラメータと呼ばれる値を適切に決める必要があり、モデル評価基準の研究を目的の一つとした。スパースモデリングの研究成果は、回帰分析はもとより識別・判別、パターン認識、次元圧縮等への研究の展開を目指した。

3. 研究の方法

現象構造を近似する多数のパラメータによって特徴付けられた大規模モデルを、2乗誤差、対数尤度にモデルの複雑さを制御する L_1 型正則化項を付与して推定する正則化法の理論的研究を行うとともに、計算アルゴリズムを利用して数値的に実行する方法に取り組んだ。

データからの情報に加えて、蓄積された知識を事前情報としてベイズアプロー

ちによってモデルに同化させた有効な推定法について研究し、 L_1 ノルムタイプの正則化項に対して、ラプラス分布型のパラメータの絶対値に基づく制約を課した確率分布族を導入し、解析的・数値的に有効な推定法を研究した。また、大規模非線形回帰モデリングの研究をベイズアプローチによって取り組み、モデルの汎化能力に大きな影響を与える調整パラメータの選択に有効に機能するモデル評価基準の提唱を目指した。

さらに、ベイズアプローチに基づく予測分布モデルを評価するためのモデル評価基準の導出を、データを生成した真の確率分布と想定したモデルとの分離のもとで行い、これによって、種々の事前分布の探索を可能とする手法開発を目指した。同時に、実際問題への適用上の難点を克服するためにマルコフ連鎖モンテカルロ法などの計算アルゴリズムの開発研究を推進し、諸科学への応用研究を積み重ねた。

4. 研究成果

諸科学・産業界で蓄積されつつある超高次元データ集合の中から有益な情報やパターンを抽出し、現象解明や予測・制御に役立つ新たなモデリングの開発研究と関連する理論研究に取り組み、以下のような研究成果を挙げた。

ベイズスパースモデリング: 線形・非線形回帰モデルの回帰係数ベクトルに L_1 ノルム制約を課した正則化法をベイズ的視点で捉えたベイズスパースモデリングについて研究した。この手法の問題点は、推定したモデルの変動とスパース化に影響を与えるハイパーパラメータの選択にある。これらの問題に対処するため、ハイパーパラメータの選択に対して予測情報量規準を導出し、モデルの推定変動を抑えるために、ブートストラップ

法とモデル平均化法を融合したモデリング手法を提案した。

ベイズアプローチによるモデリングに組み込み、従来用いられていたラプラス分布に替わる新たな事前分布を提唱し、様々な L_1 型正則化法と融合した新しいベイズ型スパースモデリングを提案した。また、モデリングに本質的な損失関数と正則化項の関係を調整する正則化パラメータの選択を目的としたモデル評価基準について研究し、より汎用性の高いベイズ型予測モデル評価基準を提唱した。提案したモデリング手法は、生命科学における遺伝子構造データ、画像工学における文字認識などへ応用し、その有効性を立証した。

非線形回帰モデリング: 高次元データ集合に基づく非線形識別・判別法として、その柔軟性と汎化能力の高さから様々な分野の問題解決に適用されているのがサポートベクターマシン (support vector machine) である。本研究では、カーネル法に基づくサポートベクター識別・判別法、線形・非線形回帰モデリング、ロジスティック回帰モデリングに組み込み、多クラス非線形識別機への拡張法、汎化能力の向上に決定的な役割を果たすカーネル関数の選択、カーネル関数に含まれる調整パラメータのデータに基づく客観的な推定法について研究し、理論的・実際の側面から重要課題を提示した。

モデルの評価と選択: L_1 型正則化法によるモデリングに本質的な役割を果たすのが損失関数と正則化項の関係を調整する正則化パラメータである。この調整パラメータの役割を数値的に検証し、その決定法を目的としたモデル評価基準を、情報量およびベイズの観点から理論的・数値的に研究した。また、将来の現象予

測に有効に機能する評価基準構築に必要な推測理論の研究を行い、今後の研究課題を提示した。

混合効果モデリング: 多数の個体に関して経時的に観測・測定された欠測を含むデータ集合に基づいて、現象の結果に影響を与える様々な要因を結びつけてモデル化する回帰モデリングの研究に取り組み、基底展開法に基づいて非線形構造をモデル化した時変型回帰係数モデルや関数混合効果モデルなど、複雑かつダイナミックな様相を呈する非線形現象の解析に有効な新たなモデリング手法を提唱した。

複数の個体に対して経時的に観測・計測されたデータを関数化処理し、得られた関数データ集合に基づく非線形混合効果モデリングについて研究した。本研究では、平均効果項およびランダム効果項の基底関数の個数をそれぞれモデル評価基準によって選択し、さらに、推定した個体変動を表すランダム効果関数を対象として、自己組織化マップやウォード法に基づく関数データクラスタリングを行い、気候変動データなどの現象分析に応用して、有効性を立証した。

スパース主成分分析: これまで回帰モデリングを通して研究に取り組んできた L_1 型正則化法を、高次元データの次元圧縮とデータ構造の視覚化に有効な手法である主成分分析、カーネル主成分分析、部分空間法に適用し、理論的・数値的に研究を推進した。特に、応用上重要な多クラス識別・判別に有効な手法の一つである部分空間法について研究し、CLAFIC法と L_1 型正則化法を融合した識別・判別法を提唱し、画像解析、遺伝子構造データの分析に適用した。

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:すべて査読有り)

- [1] Misumi, T. and Konishi, S. (2016). Mixed effects historical varying coefficient model for evaluating dose-response in flexible-dose trials. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 65, 331-344.
- [2] Park, H. and Konishi, S. (2016). Robust logistic regression modeling via the elastic net-type regularization and tuning parameter selection. *Journal of Statistical Computation and Simulation*, 86 – 7, 1450 – 1461.
- [3] Park, H. and Konishi, S. (2016). Robust coordinate descent algorithm robust solution path for high dimensional sparse regression modeling. *Communications in Statistics - Simulation and Computation*, 45-1, 115-129.
- [4] Kawano, S., Hoshina, I., Shimamura, K. and Konishi, S. (2015). Predictive model selection criteria for Bayesian lasso regression. *Journal of the Japanese Society of Computational Statistics*. Vol. 28, 67–82.
- [5] 嶋村海人・川野秀一・小西貞則 (2015). モデル平均化法による Bayesian lasso 回帰モデリング. *応用統計学*, 44-3, 101 – 117.
- [6] Park, H. and Konishi, S. (2015). Principal component selection via adaptive regularization method and generalized information criterion. *Statistical Papers*, 1-14.
- [7] Park, H., Sakaori, F. and Konishi, S. (2014). Robust sparse regression and tuning parameter selection via the efficient bootstrap information criteria. *Journal of Statistical Computation and Simulation*, 84-7, 1596 - 1607

[8] Hirose, K., Tateishi, S. and Konishi, S. (2013). Tuning parameter selection in sparse regression modeling. Computational Statistics and Data Analysis, 59, 28-40.

【学会発表】(計8件)

[1] 植木優夫, 嶋村海人, 川野秀一, 小西貞則, 田宮元. 超高次元スパース回帰法によるゲノムデータ解析. 岡山大学 IPSR・九州大学 IMI・理研 CSRS シンポジウム, 2016年2月, 理化学研究所(横浜市).

[2] 西塚真太郎, 嶋村海人, 小西貞則. 一般化スパース主成分分析による部分空間法と画像解析への応用. 第20回情報・統計科学シンポジウム, 2015年12月, 九州大学.

[3] 嶋村海人, 植木優夫, 小西貞則. Bayesian generalized fused lasso modeling via NEG distribution. 2015年度統計関連学会連合大会, 2015年9月, 岡山大学.

[4] 嶋村海人, 植木優夫, 小西貞則. NEG事前分布によるベイズスパース回帰モデリング. 第9回日本統計学会春季集会, 2015年3月, 明治大学.

[5] 嶋村海人, 川野秀一, 小西貞則. モデル平均化法による Bayesian lasso 回帰モデリング. 日本計算機統計学会第28回シンポジウム, 2014年11月, 沖縄科学技術大学院大学.

[6] 中村永友, 土屋高宏, 小西貞則. 潜在変数を含む統計モデルにおけるブートストラップ分散減少法. 2014年度統計関連学会連合大会, 2014年9月, 東京大学.

[7] 保科架風, 小西貞則. Tuning parameter selection in elastic net regularization via the generalized Bayesian information criterion. 第8回日本統計学会春季集会, 2014年3月, 同

志社大学.

[8] Misumi, T. and Konishi, S. Mixed effects historical varying coefficient model for evaluating dose response in flexible dose trials. 2013 Joint Statistical Meeting. August, 2013, Quebec, Canada.

【図書】(計1件)

Konishi, S. (2014). Introduction to Multivariate Analysis: Linear and Nonlinear Modeling. CRC/Chapman & Hall, New York. (総ページ数: 312)

【産業財産権】

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

小西 貞則 (KONISHI SADANORI)
中央大学・理工学部・教授
研究者番号: 40090550

(2) 研究分担者 (0)

()

研究者番号:

(3) 連携研究者 (0)

()

研究者番号: