

機関番号：32641

研究種目：基盤研究(B)

研究期間：2009～2013

課題番号：21300106

研究課題名(和文) 高次非線形モデリングの統合的研究

研究課題名(英文) Nonlinear modeling based on high-dimensional data

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) 12,500,000円、(間接経費) 3,750,000円

研究成果の概要(和文)：計測・測定技術の高度な進展は、諸科学・産業界で大規模・高次元データの獲得と蓄積を促進し、現象解明に有用な情報を抽出するには従来手法は有効に機能せず、新たな解析手法の研究が希求されるようになった。本研究では、現象解明と予測・制御に不可欠な現象のモデル化とデータ解析手法の開発研究に取り組み、次のような研究成果を挙げた。(1) L1型正則化法を理論的・数值的に研究し、汎化能力の高い回帰モデリング、識別・判別法を提唱した。(2) モデリングの過程で重要なモデルの評価・選択問題に取り組み、新たなモデル評価基準を導出した。(3) ベイズ型モデリングについて研究を推進し、汎用性の高いモデリングを提唱した。

研究成果の概要(英文)：The development of electronic measurement and instrumentation technologies enables us to accumulate a huge amount of data with complex structure and/or high-dimensional data. Through this research project we have investigated the problem of analyzing such datasets, and proposed various statistical modeling strategies including nonlinear regression modeling via L1 regularization, model evaluation and selection criteria, and Bayesian statistical modeling. Various regularization methods with L1 norm penalty have been investigated both in theoretical and numerical aspects. The problem of evaluating statistical models is a crucial issue in model building process. We proposed various types of model evaluation and selection criteria from both an information-theoretic point of view and a Bayesian approach. The proposed techniques may be used in the extraction of useful information and patterns from data with complex structure in the life sciences, systems engineering, and other fields.

研究分野：情報学

科研費の分科・細目：情報学基礎・統計科学

キーワード：非線形モデリング スパースモデリング 高次モデル評価基準 正則化法 ベイズモデリング

1. 研究開始当初の背景

計算機システムの高度な発展と電子化された計測・測定技術の進歩とが相俟って、諸科学、産業界のあらゆる分野で日々大量かつ多様なデータが獲得、蓄積され、データベースとして組織化されつつある。例えば、現象過程や動作過程の連続的な計測・測定データ、生命科学における遺伝子やタンパク質を特徴付けるデータ、高精細な画像データ、形や立体などの構造体の特徴付ける大規模3Dデータなど、極めて次元の高い、またある意味で無限次元のデータの獲得と蓄積を可能としてきた。このようなデータ数に比して極めて次元の高いデータの解析に当たっては、従来の解析手法は有効に機能せず、新たなデータ解析技術の開発と理論・方法論の研究の必要性が強く認識されるようになった。特に、ノイズを伴って確率的に変動する高次元データに基づく回帰モデル、識別・判別モデル、潜在変数モデル、次元圧縮モデルなどのさまざまな線形・非線形モデリングを通じた複雑な現象の解明と予測・制御、そしてデータの背後に潜む法則性の探求は、諸科学、産業界にとって極めて重要な課題といえる。

現在、国際的に多くの研究者が多様なアプローチによって線形・非線形データ解析手法の開発研究に取り組み、科学技術、産業技術への応用研究が積み重ねられている。しかしながら、高次元のデータに基づくモデリングの理論・方法論の研究に関しては、それが極めて重要な問題にも拘わらず解決すべき課題が多数残されていることを認識した。これは、これまでの研究成果が、従来のデータ解析手法の枠組みでのモデル化、非線形化には成功し、現象解明に大きな役割を果たしているものの、解析上、計算上のさまざまな問題点もフィードバックされていること、また、計測・測定技術の高度な発展が、新たにより複雑かつ高次の種々多様な形式のデータを生み出し、その中から有益な情報やパターン抽出を可能とするモデリング手法を諸科学から希求されるようになってきたことによる。

2. 研究の目的

本研究は、統計科学、情報科学、計算機科学の知識融合と数学的成果の有効利用によって、複雑な構造を内包する高次

元データの背後にある現象の理解と予測・制御、さらには新たな知識発見のための基礎的な役割を担う非線形現象のモデル化の理論・方法論の研究に統合的に取り組むことを目的とした。特に、複雑かつ高次の現象分析に有効に機能する汎化能力の高いモデリング手法の開発研究を目指して、以下のような研究を重点目標として取り組んだ。

(1) 高次元データに基づく複雑現象解明のためのモデリングは、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となり、現象の構造を有効に捉えるモデルの推定が極めて難しくなる。この問題に対して、データからの情報に加えて、当該分野の蓄積された知識を事前情報としてベイズアプローチによってモデルに同化した推定法を確立する。このため、正則化法、ベイズアプローチ、制約付最適化法の相互関係を理論的に再構築して、かつ事前情報を適切に捉える柔軟な発想のもとでの確率分布モデルを導入し、計算アルゴリズムの有効利用による新しいモデリング手法の提唱を目的とした。

(2) モデル推定へのアプローチとして、大規模モデルの推定とモデル選択に関して一つの方法を提示していると考えられる LASSO (Least Absolute Shrinkage and Selection Operator) によるモデルの推定法の研究推進と理論的拡張に、数学的成果と計算アルゴリズムの融合研究によって取り組み、新たな非線形モデリングの理論・方法論の提唱を目的とした。

(3) 識別・判別問題の研究は、学習データとよばれる既にどのクラスに属するか判明しているラベルデータに基づいて線形・非線形判別法を構成する研究が主として行われてきた。これに対して、本研究では、極めて次元の高い小数のラベルデータに、一見判別には無情報と思われるクラスの所属が未知のラベルなしデータを融合して判別モデルを構成し、予測性能を向上させる研究に取り組むことを目的とした。

3. 研究の方法

高次元データに基づく複雑現象解明のためのモデリングの研究は、現象を近似する適切なモデルの想定、想定したモデルのデータに立脚した推定、そして推定したモデルの予測の観点からの評価と選

択を一体化させることが必要である。このため、現象の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、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チによってモデルに同化した有効な推定法を確立する方法に取り組んだ。さらに、モデリングの過程において重要な役割を果たす正則化パラメータの適切な選択法について研究した。

データ数に比して変数の次元が極めて高次元の場合のモデル化と変数選択に有効な正則化法である **LASSO** と呼ばれるモデルの推定法の研究推進と基本的な考え方を拡張し、線形・非線形モデリングに有効に機能する手法提唱を目指して研究に取り組んだ。これは、すべての変数の組み合わせに対してモデルの評価・選択が難しい場合にも適用でき、安定したモデルの推定と変数選択を同時に行うことができる手法であり、大規模モデルの推定とさらに評価・選択に関して一つの方法を提示しており、同時に次元圧縮の側面からの研究を推進した。本研究では、数学的成果を有効に取り入れ、計算アルゴリズムの利用によって適用上の問題点を克服し、汎用性の高いモデリング手法の開発を目指して重点的に研究を行った。特に、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属を導入し、解析的・数値的に有効な推定法を研究した。

4. 研究成果

ベイズモデリング: 自然現象や社会現象の結果は、さまざまな要因によって生じ、結果とそれに影響を与える要因を結びつけるモデルの構築は、現象の解明と理解に大きな役割を果たす。特に、生命科学、医用工学、システム工学などの分野では、結果を生じさせる極めて多数の要因が候補として挙げられ、その中から現象予測に有効に機能する要因を、観測・測定されたデータからどのようにして探索するかが重要な課題である。本研究では、この問題に対して、当該分野の先験的な知見をモデルの中に融合させて、より有効な解析を行う方法について研究した。すなわち、ベイズ理論に基づいて先験的な知見を確率分布で表現してモデルに融合し、モデルの推定に拡張型正則化法を用

いて行う方法に取り組み、新たなモデル化のための計算アルゴリズムを提唱した。 **L_1 ノルム正則化法:** 複雑かつ多様な大規模データの分析に有効に機能する回帰モデリングの提唱を目指して研究を推進し、数理的アプローチに計算アルゴリズムを融合した方法によって、 L_1 タイプのノルム制約を課す正則化法によって推定したモデルの評価を可能とする新たなモデル評価基準を提唱した。さらに、 L_1 タイプのノルムを損失関数に課した正則化法について、その理論構造を統一的に捉え、回帰、識別・判別、パターン認識、次元圧縮への適用研究を行った。

モデルの自由度と計算アルゴリズム: 現象予測のためのモデリングは、将来の現象をいかにして精度良く予測するかが重要な問題となる。このためには、現象予測に有効でないと思われる変数を除去して、現象の構造を適切に近似する予測モデルを構築することが必然となる。この変数選択とよばれる問題は、モデルの自由度という概念が鍵を握ることが知られている。本研究では、構築したモデルの自由度を捉える問題に取り組み、モデルの自由度とは何か、またそれはどのようにして捉えればよいかを、理論的・数値的両側面にわたって研究し、新たな計算アルゴリズムを提唱した。

予測情報量規準: 複雑かつ高次の現象分析に有効に機能する汎化能力の高い非線形回帰モデリングの研究に取り組み、正則化最尤法によるモデルの推定について、様々なタイプの正則化項を課した目的関数の性質と特徴を、理論的・数値的両側面から研究した。その結果、非線形基底関数の決め方に対して L_1 型正則化法を適用したモデリングを提案した。また、正則化最尤法によるモデリングに本質的な役割を果たす正則化パラメータの選択を目的として、ベイズ型予測分布モデルの評価を可能とするモデル評価基準を導出した。これによって、複雑な非線形構造を内包するデータに基づくモデリングであるベイズ型非線形回帰モデリングを提案した。

変化点解析: 複雑な現象の構造を捉えるとき、真の構造が滑らかに変化していると考えられる場合には、スプラインなどに基づく非線形回帰モデルの想定とモデル評価基準の適用によって、現象の構造

を捉えるモデルの構築が可能である。しかし、データを発生している現象の真の構造が何らかの要因によって急激に変化しているような場合、観測されるデータに従来の方法で非線形回帰モデルを当てはめても適切なモデルは推定できない。このような問題を克服するために、新たな非線形回帰モデルの開発研究に取り組み、急激な変動を内在する非線形構造を適切に捉えることのできるモデリング手法を提案した。さらにこの研究を発展させ、現象構造が急激に変化する変化点を捉えるための不連続な基底関数と、滑らかな構造を捉えるための連続な基底関数を混合した非線形モデリング手法の開発研究に取り組み、複数の変化点をもつ非線形構造を適切に捉える非線形モデリングについて研究し、地球環境データ、気象データなどへの応用研究を行った。

ロバスト回帰: 高次元データに基づく回帰モデリングの研究を行い、異常値の混入を検知するロバスト損失関数に L_1 タイプの正則化項を付与した新たな解析法を提唱した。特に、モデリングの過程で本質的な異常値検出と制約の程度を制御する調整パラメータの選択基準を導出し、その有効性を数値実験、実データの解析を通して立証した。

半教師付識別判別法: 識別・判別法の構成は、学習データと呼ばれる既にどのクラスに属するか判明しているラベルデータに基づいて行われる。しかしながら、例えば、生命科学やテキスト分類などの分野では、ラベル付けに専門的知識を要する、あるいは極めて高コストとなるなどの理由により、極めて次元の高い小数のラベルデータとクラスの所属が未知のラベルなしデータが混在しているデータ集合を解析の対象とする場合がしばしば見られる。このような問題に対して、これまで研究を重ねてきた基底展開法に基づく非線形ロジスティック判別と新たな正則化最尤法を融合した識別・判別法を提唱した。

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 28 件 ; すべて査読有り)

[1] Park, H., Sakaori, F. and Konishi, S. (2014). Robust sparse regression

modeling and tuning parameter selection via the efficient bootstrap information criteria. *Journal of Statistical Computation and Simulation*, 84-7, 1596 - 1607

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

[3] Park, H., Sakaori, F. and Konishi, S. (2012). Selection of tuning parameters in robust sparse regression modeling. *Proceedings of COMPSTAT 2012*, pp.713-723.

[4] Kawano, S. and Konishi, S. (2012). Semi-supervised logistic discrimination for functional data, *Bulletin of Informatics and Cybernetics*. 44, 1-15.

[5] Kim, D., Kawano, S. and Konishi, S. (2012). Predictive information criteria for Bayesian nonlinear regression models. *Bulletin of Informatics and Cybernetics*, 44, 17-28.

[6] Hirose, K. and Konishi, S. (2012). Variable selection via the weighted group lasso for factor analysis models. *Canadian Journal of Statistics*, 40, 345-361.

[7] Kawano, S., Misumi, T. and Konishi, S. (2012). Semi-supervised logistic discrimination via graph-based regularization. *Neural Processing Letters*, 36(3), 203-216.

[8] Matsui, H. and Konishi, S. (2011). Variable selection for functional regression models via the L_1 regularization. *Computational Statistics and Data Analysis*, 55, 3304-3310.

[9] Tateishi, S. and Konishi, S. (2011). Nonlinear regression modeling and detecting change points via the relevance vector machine. *Computational Statistics*, 26, 477 - 490.

[10] Matsui, H., Araki, T. and Konishi, S. (2011). Multiclass functional discriminant analysis and its application to gesture recognition. *Journal of Classification*, 28, 227 - 243.

[11] Kawano, S. and Konishi, S. (2011). Semi-supervised logistic discrimination via regularized Gaussian basis expansions. *Communications in Statistics-Theory and Methods*, 40, 2412-2423.

[12] Kitagawa, G. and Konishi, S. (2010). Bias and variance reduction techniques for bootstrap information criteria. *Annals of*

the Institute of Statistical Mathematics, 62, 209-234.

[13] Hirose, K., Kawano, S., Miike, D. and Konishi, S. (2010). Hyper-parameter selection in Bayesian structural equation models. *Bulletin of Informatics and Cybernetics*, 42, 55-70.

[14] Kayano, M., Dozono, K. and Konishi, S. (2010). Functional cluster analysis via orthonormalized Gaussian basis expansions and its application. *Journal of Classification*, 27, 211-230.

[15] Tateishi, S., Matui, H. and Konishi, S. (2010). Nonlinear regression modeling via the lasso-type regularization. *Journal of Statistical Planning and Inference*, 140, 1125 - 1134.

[16] 川野秀一, 廣瀬慧, 立石正平, 小西貞則 (2010). 回帰モデリングと L_1 型正則化法の最近の展開. *日本統計学会誌* 39 巻 2 号, 211 - 242.

[17] Pan, Q. and Nishii, R. (2010). Selection of ARX models estimated by the penalized weighted least squares method. *Bulletin of Informatics and Cybernetics*, 42, 35--43.

[18] Pan, Q., Nishii, R., Nakagawa, T. and Nakamoto, T. (2010). ARX models for time-varying systems estimated by recursive penalized weighted least squares method. *Journal of Math-for-Industry*, 2, 109--114.

[19] Maesono, Y. (2010). Edgeworth Expansion and Normalizing Transformation of Ratio Statistics and their Application. *Communications in Statistics-Theory and Methods*, Vol.39, pp.1344-1358.

[20] Araki, Y., Konishi, S., Kawano, S. and Matsui, H. (2009). Functional regression modeling via regularized Gaussian basis expansions. *Annals of the Institute of Statistical Mathematics*, 61-4, 811-833.

[21] Araki, Y., Konishi, S., Kawano, S. and Matsui, H. (2009). Functional logistic discrimination via regularized basis expansions. *Communications in Statistics - Theory & Methods*, 2944 - 2957.

[22] Kawano, S. and Konishi, S. (2009). Nonlinear logistic discrimination via regularized Gaussian basis expansions. *Communications in Statistics - Theory & Methods*, 38, 1414- 1425.

[23] Kayano, M. and Konishi, S. (2009). Functional principal component analysis via regularized Gaussian basis expansions and

its application to unbalanced data. *Journal of Statistical Planning and Inference*, 139, 2388 - 2398.

[24] Matsui, H., Kawano, S. and Konishi, S. (2009). Regularized functional regression modeling for functional response and predictors. *Journal of Math-for-Industry*, 1, 17 - 25.

[25] 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.

[26] Masuda, H. (2009). Joint estimation of discretely observed stable Levy processes with symmetric Levy density. *J. Japan Statist. Soc.* 39, No.1, 49--75.

[27] Masuda, H. and Morimoto, T. (2009). Empirical analysis on jump detection in high-frequency data. (in Japanese). *J. Japan Statist. Soc. (Japanese Issue)*. 39, No.1, 33--63.

[28] Masuda, H. (2009). Notes on estimating inverse-Gaussian and gamma subordinators under high-frequency sampling. *Ann. Inst. Statist. Math.* 61, No.1, 181--195.

[学会発表] (計 21 件)

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

[2] 土居正明, 小西貞則. Model selection for Cox model via sparse regularization. 2013 年度統計関連学会連合大会, 2013 年 9 月, 大阪大学.

[3] 松田和己, 川野秀一, 小西貞則. 変分ベイズ法による非線形回帰モデリング. 2013 年度統計関連学会連合大会, 2013 年 9 月, 大阪大学.

[4] Misumi, T. and Konishi, S., Mixed effects historical varying coefficient model for evaluating dose response in flexible dose trials. 2013 Joint Statistical Meeting, Montreal, 2013 年 8 月

[5] Hirose, K., Tateishi, S. and Konishi, S., Tuning parameter selection in sparse regression modeling. 2012 Joint Statistical Meetings, San Diego, USA, 2012 年 8 月.

[6] Hoshina, I. and Konishi, S., Algorithm for constructing model selection criteria in sparse regression modeling. 2012 Joint Statistical Meetings, San Diego, USA, 2012 年 8 月.

[7] Matsuda, K. and Konishi, S., Sparse Bayesian regression modeling via the relevance vector machine. 2012 Joint Statistical Meetings, San Diego, USA, 2012年8月.

[8] Matsuda, K. and Konishi, S. Nonlinear statistical modeling via the relevance vector machine and its application to change point analysis. ISBA 2012 World Meeting, 2012年6月 Kyoto.

[9] Hoshina, I. and Konishi, S. Algorithm for constructing sparse regression models via Bayesian lasso. ISBA 2012 World Meeting, 2012年6月 Kyoto.

[10] 保科架風, 小西貞則. Bayesian lassoとモデル選択. 2011年度統計関連学会連合大会. 2011年9月, 九州大学.

[11] 廣瀬慧, 立石正平, 小西貞則. L1型正則化法におけるチューニングパラメータの選択. 2011年度統計関連学会連合大会. 2011年9月, 九州大学.

[12] 立石正平, 小西貞則. Fused lassoに基づく非線形回帰モデリング. 2010年度統計関連学会連合大会. 2010年9月, 早稲田大学.

[13] Jan Dolinsky, 廣瀬慧, 小西貞則. Echo State Networks with Non-Monotonous Activation Functions. 2010年度統計関連学会連合大会. 2010年9月, 早稲田大学.

[14] Hirose, K. and Konishi, S., Variable selection via the lasso-type regularization for structural equation models. 2010 Joint Statistical Meetings, Vancouver Canada, 2010年8月.

[15] Tateishi, S. and Konishi, S., Nonlinear regression modeling and detecting change points via regularized basis expansions. 2010 Joint Statistical Meetings, Vancouver Canada, 2010年8月.

[16] 立石正平, 小西貞則. 関連ベクターマシンに基づく非線形回帰モデルと変化点探索. 2009統計関連学会連合大会. 2009年9月, 同志社大学.

[17] 川野秀一, 小西貞則. 関数データに基づく半教師あり識別・判別問題. 2009統計関連学会連合大会. 2009年9月, 同志社大学.

[18] 廣瀬慧, 小西貞則. Group Lassoに基づく因子分析モデリング. 2009統計関連学会連合大会. 2009年9月, 同志社大学.

[19] Kawano, S. and Konishi, S., Semi-supervised learning via regularized logistic discrimination. 2009 Joint Statistical

Meeting in Washington, DC, 2009年8月.

[20] Tateishi, S. and Konishi, S., Nonlinear regression modeling via Bayesian regularization with Lasso-type penalties, 2009 Joint Statistical Meeting in Washington, DC, 2009年8月.

[21] Hirose, K., Ichikawa, M. and Konishi, S., A choice of the number of factors and hyper-parameter selection in Bayesian factor analysis model. 2009 Joint Statistical Meeting in Washington, DC, 2009年8月.

〔図書〕 (計2件)

Konishi, S. (2014). Introduction to Multivariate Analysis: Linear and Nonlinear Modeling. CRC, Chapman & Hall, New York.

小西貞則 (2010). 多変量解析入門—線形から非線形へ—, 岩波書店.

〔産業財産権〕

○出願状況 (計0件)

○取得状況 (計0件)

〔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

西井 龍映 (NISHI RYUEI)
九州大学・MI研究所・教授
研究者番号: 40127684

前園 宜彦 (MAESONO YOSHIHIKO)
九州大学・大学院数理学研究院・教授
研究者番号: 30173701

(3) 連携研究者

二宮 嘉行 (NINOMIYA YOSHIYUKI)
九州大学・MI研究所・准教授
研究者番号: 50343330

増田 弘毅 (MASUDA HIROKI)
九州大学・MI研究所・准教授
研究者番号: 10380669