

平成 21 年 5 月 14 日現在

研究種目： 基盤研究(B)
研究期間： 2005 ~ 2008
課題番号： 17300089
研究課題名（和文） 超高次元データに基づく非線形多変量解析手法の開発と応用
研究課題名（英文） Nonlinear multivariate analysis based on high-dimensional data and its application
研究代表者
小西 貞則 (KONISHI SADANORI)
九州大学・大学院数理学研究院・教授
研究者番号： 40090550

研究成果の概要： 高次元データに基づく複雑現象解明を目的として、非線形多変量解析の理論・方法論の開発研究と応用研究に取り組み、現象の構造を有効に捉える非線形回帰モデリング、ロジスティックモデルに基づく非線形識別・判別法、リスクを定量的に評価・予測するリスク予測モデルを提唱した。また、観測・測定されたデータを関数化処理し、処理した関数化データ集合に基づく解析手法について研究し、回帰、識別・判別、次元圧縮に関して、関数データ解析手法を提唱することができた。開発した解析手法は、生命科学、システム工学、地球環境科学などの問題解決に応用した。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2005 年度	3,000,000	0	3,000,000
2006 年度	2,900,000	0	2,900,000
2007 年度	3,200,000	960,000	4,160,000
2008 年度	3,700,000	1,110,000	4,810,000
年度			
総計	12,800,000	2,070,000	14,870,000

研究分野： 統計科学

科研費の分科・細目： 情報学・統計科学

キーワード： 非線形モデリング，関数データ解析，機械学習，ベイズ理論，カーネル法，計算統計，確率過程，共分散構造分析

1. 研究開始当初の背景

計算機の高度な発展と計測・測定技術の進展が相俟って、現象過程の連続的な計測・測定データ、ゲノムデータ、時空間データなど、複雑かつ大規模なデータが科学の諸分野で観測され、データベースとして組織化される

ようになってきた。これにともない、不確実性を有する自然現象・社会現象の解明とその本質の探求のためのより高度な数理の展開と効率的な情報処理の必要性が強く認識されるようになった。例えば、生命科学における遺伝子やタンパク質を特徴付ける高次元

データに基づく分析, また経時的に高頻度で計測された動作過程のデータは, ある意味で無限次元のデータを分析対象とすることになる. このような超高次元のデータの分析に対しては, 従来の手法は有効に機能せず, 新たな手法開発と数理の研究が強く望まれるようになっていた.

2. 研究の目的

複雑な非線形構造を内包する超高次元のデータの分析に対しては, 従来の統計的データ解析手法は有効に機能しない場合が多く, 非線形回帰モデリング, 識別・判別法, 次元圧縮法, 分類法などの非線形多変量解析手法の開発と有効に機能する新しいモデル評価基準の開発は極めて重要であることを認識した. このため, 非線形現象解明を目的として, 統計科学, 情報科学等を融合させることによって, 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分析手法の開発と理論・方法論の研究を目的として研究を推進した. 開発したデータ解析手法は, 生命科学, システム工学, 環境科学, 地球科学などの分野の情報抽出の技術として応用し, 諸科学の問題解決に取り組んだ.

3. 研究の方法

(1) 複雑な非線形現象の解明を目的として, 新しい非線形回帰モデリングの研究に取り組んだ. 特に, モデルの構築過程においてデータの確率構造を有効に取り入れ, 正則化法, ベイズアプローチによってモデルを推定する方法を研究した. このため, モデルの複雑さの程度を調整するパラメータの選択をモデル選択の問題として捉え, 新たなモデル評価基準の導出を目指した.

(2) 超高次元データの次元圧縮とデータの可視化の方法について研究を行った. このた

め, 基底展開法による非線形モデルのデータへの当てはめを行い, その推定係数を個体あるいは対象の特徴ベクトルとして対応させてモデリングを試み, 情報の損失をできる限り少なくするという意味において適切な基底関数の個数の選択法について研究した.

(3) 機械学習の分野で, 極めて次元の高いデータに基づく識別法として研究が進められているサポートベクターマシーンと動径基底関数ネットワークモデルとの関係について, その理論構造と手法の特徴について研究し, 予測性能の優れた識別・判別法の提唱に向けて研究を行った.

(4) ウェブレットを複雑な現象解析に適用する問題に取り組み, ノイズを含み確率的に変動するデータに基づく統計的モデリングの研究を推進した.

(5) 例えば, 人間の歩行やロボット制御における動作過程のデータは, 本来連続な時間データを離散化して観測していると考えることができる. そこで, 経時的に観測された離散時間データを平滑化処理して関数化し, 関数化データ集合に基づくさまざまな分析手法の開発研究に取り組んだ.

(6) 生命科学, 地球科学, システム工学などの分野において, 複雑な現象分析に携わる研究者との共同研究を通して, 開発した手法の適用を計るとともに, 起因した問題点をフィードバックさせて現象分析に有効に機能するデータ解析手法の開発に努めた.

(7) 開発したデータ解析手法は理論的・数値的に検証し, 手法の有効性, 特徴, 性質を明らかにすると共に, 実際問題への応用を通して有効性を検証した.

4. 研究成果

(1) 生命科学, 地球環境科学, システム工学等の分野でしばしば観測・測定される極めて

高次元のデータに基づく情報抽出に取り組み、この問題を、ベクトルデータを関数に置き換えて分析を実行する関数データ解析の枠組みで研究した。関数化処理のための非線形モデリングに基づく関数化と混合効果モデルに基づく関数化の研究を行い、関数化したデータ集合の中から有益な情報やパターンを効率的に抽出するための新たな統計的分析手法、特に、識別・判別、クラスタリング、次元圧縮に関していくつかの関数データ解析手法を提唱することができた。応用研究として、分光スペクトル解析、動作過程の機能異常解析、遺伝子構造データにおける周期性を有する遺伝子の特徴抽出、タンパク質の立体構造の分類と可視化、気候変動データの分類などの問題に取り組み、従来の手法では分析が難しかった問題に対して開発した手法を適用し、現象の解明に寄与することができた。

(2) 関数データ解析を適用するに当たって解決すべき重要な問題の一つは、ノイズを伴う多数の高次元ベクトルデータに含まれる情報を可能な限り失うことなくどのように関数データ化するかである。この問題に対して、分析結果に影響するベクトルデータの情報を高効率で含有する関数化の方法について、現象構造を近似するモデル、関数データ集合に基づくモデルの推定、適切なモデル選択の一連のプロセスを総合的に研究し、新たな分析手法を提唱することができた。

(3) 因子分析、共分散構造方程式モデリングについて研究し、バイズアプローチによって不適解問題を解消し、因子数を選択する方法を提唱した。この研究によって、実際問題へ因子分析を適用するに当たって、因子数の決定に一つの有効な方法を提供でき、従来の手法では捉えきれない要因抽出法が提案できた。また、特異モデルにおいて非正則性を扱

うために開発された局所錐母数化の概念を用いて、因子分析モデルに対する新しいモデル選択法を構築した。さらに、モデルの非線形化、混合効果モデルへの拡張研究を推進中である。

(4) リモートセンシング画像の土地被覆分類のために、マルコフ確率場を用いて複数の土地被覆が混在する画素を利用する手法や、分布間の距離に基づく確率場による手法を開発研究した。また統計的学習理論のアプローチにより、確率場に基づく手法と同等性能を高速度で実現する手法を開発した。

(5) 強い相関をもつ多重検定に対する保守的な検定の提案と、非正則な統計モデルに基づく林分成長分析方法を提案し、量的遺伝子座の検知問題やDNAシーケンスの中の変化領域検知問題など、ゲノムサイエンスの重要課題に前者の方法を適用し、現象解明に結びついた。

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 26 件)

① Ichikawa, M. and Konishi, S., Constructing second-order accurate confidence intervals for communalities in factor analysis, *British Journal of Mathematical and Statistical Psychology*, (査読有), **61**, 361 – 378, (2008).

② Araki, Y., Konishi, S., Kawano, S. and Matsui, H., Functional regression modeling via regularized Gaussian basis expansions, (査読有) *Annals of the Institute of Statistical Mathematics*, Springer Online 版, DOI (10.1007/s10463-007-0161-1), (2008).

③ Ando, T., Konishi, S. and Imoto, S., Nonlinear regression modeling via regularized radial basis function networks, (査読有) *Journal of Statistical Planning and Inference*, **138**, 3616 – 3633, (2008).

④ Maesono Y., Edgeworth Expansion and Normalizing Transformation of Ratio Statistics and their Application, (査読有)

Communications in Statistics-Theory and Methods, 掲載決定, (2008).

⑤ Ninomiya, Y. and Yoshimoto, A., Statistical method for detecting structural change in the growth process, (査読有) *Biometrics*, **64**, 46-53, (2008).

⑥ Masuda, H., Notes on estimating inverse-Gaussian and gamma subordinators under high-frequency sampling, (査読有) *Annals of the Institute of Statistical Mathematics*, 掲載決定, (2008)..

⑦ Masuda, H., On stability of diffusions with compound-Poisson jumps, (査読有) *Bulletin of Informatics and Cybernetics*, **40**, 61-74, (2008)..

⑧ Kawano, S. and Konishi, S., Nonlinear regression modeling via regularized Gaussian basis functions, (査読有) *Bulletin of Informatics and Cybernetics*, **39**, 83 – 96, (2007).

⑨ Kawaguchi, S. and R. Nishii, Hyperspectral image classification by Bootstrap AdaBoost with random decision stumps, (査読有) *IEEE Transactions on Geoscience and Remote Sensing*, **45**(11), 3845 –3851, (2007).

⑩ Maesono, Y. An Edgeworth expansion and a normalizing transformation for L-statistics, (査読有) *Bulletin of Informatics and Cybernetics*, **39**, 25-43, (2007).

⑪ 二宮嘉行・柳原宏和・吉本敦, 非正則な統計モデルに基づく林分成長分析, (査読有) 森林資源管理と数理モデル, **6**, 43—56, (2007).

⑫ Fujii, T. and Konishi, S., Nonlinear regression modeling via regularized wavelets and smoothing parameter selection, (査読有) *Journal of Multivariate Analysis*, **97**, 2023 – 2033, (2006).

⑬ 茅野光範・小西貞則・平川英樹・久原哲, 正則化基底展開法に基づく関数主成分分析とその応用 (査読有) 応用統計学, **35**(1), 1-16, (2006).

⑭ Nishii, R. and Eguchi, S., Image classification based on Markov random field models with Jeffreys divergence. (査読有) *Journal of Multivariate Analysis*, **97**(9), 1997-2008, (2006).

⑮ 川口修治, 山崎謙介, 西井龍映, ミクセルを考慮したマルコフ確率場に基づくリモートセンシング画像の教師なし土地被覆分類, (査読有) 日本リモートセンシング学会誌, **26**(2), 107-116, (2006).

⑯ 西井龍映, 江口真透, 統計的学習理論による多重分光画像の画素判別, (査読有) 計算機統計学, **19**(2), 151—164, (2006).

⑰ Nonaka, Y. and Konishi, S., Nonlinear regression modeling using regularized local likelihood method, (査読有) *Annals of the Institute of Statistical Mathematics*, **57**, 617 – 635, (2005).

⑱ 藤井 亨・小西 貞則, ウェーブレットによる関数データ判別, (査読有) 応用統計学, **34**-3, 151-169, (2005).

⑲ Nishii, R. and S. Eguchi, Supervised image classification by contextual AdaBoost based on posteriors in neighborhoods. (査読有) *IEEE Transactions on Geoscience and Remote Sensing*, **43**(11), 2547-2554, (2005).

⑳ Maesono, Y., Higher Order Comparisons of Asymptotic Confidence Intervals, (査読有) *Journal of Statistical Planning and Inference*, **133**, 359-379, (2005).

(21) Maesono, Y., Asymptotic representations of ratio statistics and their mean squared errors, (査読有) *Journal of the Japan Statistical Society*, **35**, 73-97, (2005).

(22) Ninomiya, Y., Information criterion for Gaussian change-point model, (査読有) *Statistics and Probability Letters* (査読有), **72**, 237—247, (2005).

(23) 柳原宏和・吉本敦・二宮嘉行, 多変量線形モデルによる林分成長要因探索のための変数選択, (査読有) 日本森林学会誌, **87**, 504—512, (2005).

(24) Masuda, H., Classical method of moments for partially and discretely observed ergodic models, (査読有) *Stat. Inference Stoch. Process*, **8**, 25-50, (2005).

(25) Masuda, H. and Yoshida, N., Asymptotic expansion for Barndorff-Nielsen and Shephard's stochastic volatility model, (査読有) *Stochastic Processes Appl.*, **115**, 1167-1186, (2005).

(26) Masuda, H., Simple estimators for parametric Markovian trend of ergodic processes based on sampled data, (査読有) *J. Japan Statist. Soc.*, **35**, 147-170, (2005).

[学会発表] (計 14 件)

① 立石正平, 松井秀俊, 小西貞則, Lassoタイプの正則化法に基づく非線形回帰モデリング, 統計関連学会連合大会, 慶応大学, 2008年9月.

② 川野秀一, 三角俊裕, 小西貞則, 正則化法に基づく半教師あり識別・判別問題, 統計関連学会連合大会, 慶応大学, 2008年9月.

③ 廣瀬慧, 芹川義和, 小西貞則, 確率的カ

ーネル主成分分析とその応用, 統計関連学会
連合大会, 慶応大学, 2008年9月.

④ Konishi, S., Recent Development in
Nonlinear Statistical Modeling and Model
Selection, International Indian Statistical
Association Conference, Connecticut, 2008
年5月.

⑤ Konishi, S., Recent development in
statistical modeling and model selection,
The 21st Century COE program
Development of Dynamic Mathematics
with High Functionality, Fukuoka, 2007
年10月.

⑥ 松井秀俊, 荒木貴光, 小西貞則, 正則化
法に基づく関数回帰モデルの推定とその評
価, 統計関連学会連合大会, 神戸大学, 2007
年9月.

⑦ 茅野光範, 小西貞則, 不均一関数データ
に基づく正則化主成分分析とその応用, 統計
関連学会連合大会, 神戸大学, 2007年9月.

⑧ 川野秀一, 三角俊裕, 小西貞則, 半教師
あり学習法に基づく非線形判別, 統計関連学
会連合大会, 神戸大学, 2007年9月.

⑨ 廣瀬慧, 川野秀一, 小西貞則, 市川雅教,
ベイジアン因子分析モデルと因子数の選択,
統計関連学会連合大会, 神戸大学, 2007年9
月.

⑩ Matsui, H., Kawano, S., Kayano, M.
and Konishi, S., Regularized functional
regression modeling for functional
response and predictors, International
Statistical Insitute, Lisboa, Portugal,
2007年8月.

⑪ Kayano, M. and Konishi, S., Principal
component analysis for sparse functional
data via regularized basis expansions,
International Statistical Insitute, Lisboa,
Portugal, 2007年8月.

⑫ Konishi, S., Nonlinear statistical
modeling and its application to the
functional data analysis, Taipei
International Statistics Workshop, 2006年
12月.

⑬ Araki, Y. and Konishi, S., Functional
supervised and unsupervised classification
of gene expression data, COMPSTAT 2006,
17th Symposium of IASC-ERS, Italy, 2006
年8月.

⑭ Kayano, M. and Konishi, S., Functional
principal component analysis via
regularized basis expansions and its
application to protein structural data, 23rd
International Biometric Conference,
Montreal, Canada, 2006年7月.

[図書] (計3件)

① Konishi, S. and Kitagawa, G. (2008).
Information Criteria and Statistical Modeling,
Springer, New York, (273頁).

② 小西貞則・越智義道・大森裕浩 (2008). 計
算統計学の方法 -ブートストラップ, EM ア
ルゴリズム, MCMC-, 朝倉書店, (223頁).

③ 小西貞則・竹内純一 (2008). 統計的モデリ
ング/情報理論と学習理論, 講談社, (120頁).

[その他]

ホームページ等

[http://hyoka.ofc.kyushu-u.ac.jp/search/
details/K000534/index.html](http://hyoka.ofc.kyushu-u.ac.jp/search/details/K000534/index.html)

6. 研究組織

(1) 研究代表者

小西 貞則 (KONISHI SADANORI)
九州大学・大学院数理学研究院・教授
研究者番号: 40090550

(2) 研究分担者

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
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
百武 弘登 (HYAKUTAKE HIROTO)
九州大学・大学院数理学研究院・准教授
研究者番号: 70181120
二宮 嘉行 (NINOMIYA YOSHIYUKI)
九州大学・大学院数理学研究院・准教授
研究者番号: 70280526
増田 弘毅 (MASUDA HIROKI)
九州大学・大学院数理学研究院・助教
研究者番号: 10380669

(3) 連携研究者

内田 雅之 (UCHIDA MASAYUKI)
大阪大学・大学院基礎工学研究科・准教授
研究者番号: 70280526