

令和 6 年 5 月 3 日現在

機関番号：14201

研究種目：基盤研究(C)（一般）

研究期間：2019～2023

課題番号：19K11858

研究課題名（和文）関数データに基づく経時測定データ解析法の発展

研究課題名（英文）Development of the analysis of longitudinal data via functional data analysis

研究代表者

松井 秀俊（Matsui, Hidetoshi）

滋賀大学・データサイエンス学系・教授

研究者番号：90633305

交付決定額（研究期間全体）：（直接経費） 2,900,000円

研究成果の概要（和文）：観測個体それぞれが時間の経過とともに繰り返し計測されたデータに対して、各観測個体を関数として処理し分析を行う統計的モデリング手法を発展させる研究を行った。特に、回帰や分類、空間データの予測など、さまざまな分析目的に応じて、データに内在する特徴を明らかにする方法を提案した。提案した方法それぞれについて、数値実験を通じてそれらの有効性を検証した。また、実際のデータに対して、さまざまな視点からデータの価値を引き出すための分析を行った。

研究成果の学術的意義や社会的意義

関数データに基づく分類問題に対しては、関数データを効率的かつ精度よく多群に分類する方法を構築できた。また、空間情報を伴う関数データを用いて未観測地点の関数を予測する問題では、予測に用いられるデータの地点を選択する方法を導出した。このような地点の選択ができる方法はこれまでに提案されておらず、学術的意義は高いと考えられる。さらに、長期栽培される農作物を分析するための統計モデルとその推定法を提案することで、これまでは農家が肌感覚で捉えていた環境要因や季節と収量との関係を定量化することができた。この結果は、実際に栽培を行う農家に対してフィードバックを与えられるという点で意義のあるものと考えられる。

研究成果の概要（英文）：We conducted research to develop statistical models for analyzing repeatedly measured data for each individual. Specifically, we treated the data for each individual as a smooth function and then considered methods for analyzing functional data. We have proposed methods to reveal features inherent in the data for various purposes, such as regression, classification, and prediction of spatial data. For each of the proposed methods, we verified their effectiveness through numerical experiments. In addition, we applied the proposed method to the analysis of real data to extract the value of the data from various perspectives.

研究分野：データサイエンス

キーワード：統計的モデリング 関数データ解析 モデル選択 空間データ解析

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

近年の計測機器の発展と普及に伴い、時間の経過に伴い繰り返し測定される形式のデータである経時測定データが収集される場面が増加している。経時測定データの分析にあたっては、主に次の2点の性質から、しばしば古典的な多変量解析手法を適用することが困難となる。

- ・ 観測時点数の増加によりデータの次元が増大する
- ・ 観測の欠損や観測時点または時点数の不一致が生じる

経時測定データの分析のためのアプローチの1つである**関数データ解析**は、離散時点で観測されたデータを関数化処理し、得られた関数集合をデータとして扱う解析方法である。これにより、上記2点の問題点を解消できる。関数データ解析に基づく統計的モデリング手法の開発を通じて、変数間の関係の経時的な変動を定量的に明らかにし、なおかつ精度の良い予測を行うためのモデルを構築すること、そして、大規模かつ複雑な構造を持つデータから解釈の容易な情報を集約して抽出するための推定方法を確立することが求められている。

2. 研究の目的

関数データに基づくより発展的な統計的モデリング手法を開発するにあたり、次の3点のテーマを目的とした。

- (1) 関数データに基づく多群分類において、変数選択に加えて、各変数がどの群の分類に寄与しているかを選択する**決定境界選択**を行うための方法を導出する。
- (2) データの関数化における最適なモデル選択を、関数データに基づく分析の枠組みで行うための方法を確立する。
- (3) 関数データの説明変数とスカラー目的変数との関係を表す関数回帰モデルの一つで、外政変数による影響を考慮に入れた**変化係数関数回帰モデル**の発展を通して、気象条件と農作物の収量との関係を明らかにする。

本研究を遂行することで、従来の方法では見出せなかった、経時測定データを含んだデータ間の複雑な関係や特徴を浮かび上がらせる。

3. 研究の方法

関数データに基づく回帰、分類モデルを既存のモデルを拡張する形で提案する。導出されたモデルに対して、主として正則化法に基づく推定法および、交差検証法や情報量規準によるモデル評価法を導出する。特に、正則化法としては ridge 型を用いて安定した推定量を得る方法や、lasso 型を用いて重要な情報を縮約して抽出する方法について検討する。

提案した手法の有効性および特性を、数値実験を通して検証する。さらに、実際のデータ分析を通して、提案手法の適用事例を示すとともに、データに内在する複雑な関係性を明らかにする。

4. 研究成果

- (1) 関数データに基づくロジスティック回帰モデルにおいて、スパース推定法の一つである sparse group lasso を用いて推定することで、関数データの変数選択と決定境界の選択を同時に行う方法を提案した。提案手法を、人工データおよび実データ(遺伝子発現データ)の解析に適用することで、有効性を検証した。
- (2) 関数データとして与えられた説明変数と、多変量のベクトルからなる目的変数との関係を

表す多変量関数回帰モデルに対して、スパース推定に基づき変数選択を行う方法を提案した。関数データは基底関数展開で表されると仮定し、その係数に対して標準化 group lasso 制約を課すことで、多変量の目的変数に寄与する説明変数を選択できる。

- (3) 説明変数と目的変数が共に時間の関数データであるとき、両変数の時間の依存関係を考慮に入れてモデル化した historical functional linear model を、説明変数が多変量の形へ拡張し、さらにこのモデルをスパース正則化に基づき推定することで、回帰係数の推定と変数選択を同時に行う方法を提案した。
- (4) 関数データに対して、部分空間法を用いて多群分類を行う方法を提案した。関数データを各群で低次元の部分空間に射影し、さらに予測したい対象の関数データをこれらの部分空間へ射影したとき、その類似度が最大となる群へデータを分類する手順を導出した。実データへの適用例として、特定の交差点を通過する交通オブジェクトの軌跡や速度の経時データから、交通オブジェクトの種類を分類する課題に取り組み、各オブジェクトの類似性について考察した。
- (5) 多変量データに対する 2 群分類問題において、一方の群において一部のデータのみバイアスを伴ってラベル付けされており残りはラベルが未観測という状況下において、そのバイアスを補正する損失関数を構築することで分類を行う方法を提案した。提案方法は Positive and biased Negative (PbN) 分類法と名付けた。数値実験やベンチマークデータの分析を通して、既存手法よりも精度よく分類ができることを示した。
- (6) 空間情報を伴う経時測定データを空間関数データとして扱い、関数クリギングを利用して未観測地点における関数を予測し、さらにそのうえで空間関数データをクラスタリングする方法を提案した。観測地点のデータと未観測地点の予測値をまとめて、x-means 法に基づくクラスタリングを適用することでクラスタリングと最適なクラスター数の選択を同時に行った。提案した手法を世界の気象のデータ分析に適用し、気温と降水量から未観測地点のクラスターへの割当を行った。
- (7) 関数クリギングに対してスパース推定を適用することで、未観測地点の予測に寄与する地点を選択する方法を提案した。関数クリギングでは未観測地点の関数を観測地点の関数データの線形結合で表すが、その係数の推定にスパース正則化を導入することで、未観測地点の関数の予測に用いる観測地点を選択できる。本研究では、スパース正則化に基づく係数の推定アルゴリズムを導出し、数値実験や実データ分析を通してその有効性を検証した。
- (8) 農作物の収穫量は、栽培期間中の気温や日射量といった環境要因に影響を受けると考えられている。特に、トマトのように 1 年のうち長期にわたり収穫される作物の収量は、収穫日から遡った環境要因の経時変化だけでなく、収穫時期(季節)にも依存していると考えられている。本研究では、この関係性を捉える統計モデルとして変化係数関数回帰モデルに注目し、このモデルをさまざまな方向へ発展させた。

変化係数関数回帰モデルは、関数データの時点を固定した下では線形モデルとなり、柔軟性に欠ける。本研究では、変化係数関数回帰モデルを加法モデルの枠組みへ拡張した変化係数関数加法モデルを提案し、その推定法および、推定されたモデルの評価法を導出した。提案した方法は予測精度の向上だけでなく、説明変数がどの時点でどのように目的変数に寄与しているかを定量化できる。

変化係数関数回帰モデルにおいて、説明変数に対応する関数データが複数与えられたとき、これらをスパース推定に基づき変数選択する方法を提案した。正則化項として group lasso 型の制約を用い、モデルに含まれるパラメータを推定するアルゴリズムを

導出した。提案手法をトマトの収量データの分析に適用し、収量に関連する環境要因を取捨選択した。

関数データに対する回帰モデルでは、係数関数の定義域の一部を0と推定することで説明変数に対応する関数データについてどの時点の情報が目的変数に寄与しているかを明らかにできる。この方法は説明変数のドメイン選択とよばれている。本研究では、変化係数関数回帰モデルに対してドメイン選択を行う方法を、スパース正則化項を拡張し導出することで提案した。提案手法をトマトの収量データの分析に適用することで、何日前までの気温が収穫量に寄与しているかを、任意の季節において定量化した。

以上の成果のうち、(1)から(6)については査読付き論文誌に採択済みである。また、(7)および(8) ~ の研究については、プレプリントで成果を公開している。

5. 主な発表論文等

〔雑誌論文〕 計9件（うち査読付論文 9件/うち国際共著 0件/うちオープンアクセス 5件）

1. 著者名 Watanabe Shotaro, Matsui Hidetoshi	4. 巻 35
2. 論文標題 Classification From Positive and Biased Negative Data With Skewed Labeled Posterior Probability	5. 発行年 2023年
3. 雑誌名 Neural Computation	6. 最初と最後の頁 977-994
掲載論文のDOI（デジタルオブジェクト識別子） 10.1162/neco_a_01580	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Esaki, T., Horinouchi, T., Natsume-Kitatani, Y., Nojima, Y., and Matsui, H.	4. 巻 20
2. 論文標題 Estimation of relationships between gene expression and chemical substructure: adapting canonical correlation analysis for small sample data by gathered features using consensus clustering	5. 発行年 2020年
3. 雑誌名 Chem-Bio Informatics Journal	6. 最初と最後の頁 58-61
掲載論文のDOI（デジタルオブジェクト識別子） 10.1273/cbij.20.58	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Inoue, K., Takahagi, K., Kouzai, Y., Koda, S., Shimizu, M., Uehara-Yamaguchi, Y., Nakayama, R., Kita, T., Onda, Y., Nomura, T., Matsui, H., Nagaki, K., Nishii, R., and Mochida, K.	4. 巻 2
2. 論文標題 Parental Legacy and Regulatory Novelty in Brachypodium Diurnal Transcriptomes Accompanying their Polyploidy.	5. 発行年 2020年
3. 雑誌名 NAR Genomics and Bioinformatics	6. 最初と最後の頁 1-14
掲載論文のDOI（デジタルオブジェクト識別子） 10.1093/nargab/lqaa067	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Matsui, H.	4. 巻 53
2. 論文標題 Variable selection for historical functional linear model.	5. 発行年 2021年
3. 雑誌名 Bulletin of informatics and cybernetics	6. 最初と最後の頁 1-19
掲載論文のDOI（デジタルオブジェクト識別子） 10.5109/4151124	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Hidetoshi Matsui, Yuta Umezu	4. 巻 3
2. 論文標題 Variable selection in multivariate linear models for functional data via sparse regularization	5. 発行年 2020年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 453-467
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s42081-019-00055-x	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 松井秀俊	4. 巻 67
2. 論文標題 関数データに基づく統計的モデリング	5. 発行年 2019年
3. 雑誌名 統計数理	6. 最初と最後の頁 73-96
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Fukuda Tatsuya, Matsui Hidetoshi, Takada Hiroya, Misumi Toshihiro, Konishi Sadanori	4. 巻 -
2. 論文標題 Multivariate functional subspace classification for high-dimensional longitudinal data	5. 発行年 2023年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s42081-023-00226-x	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 新井仁智, 松井秀俊, 三角俊裕, 小西貞則	4. 巻 -
2. 論文標題 空間経時測定データに対する多変量関数クラスタリング	5. 発行年 2024年
3. 雑誌名 応用統計学	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計30件（うち招待講演 6件 / うち国際学会 8件）

1. 発表者名 Matsui, H.
2. 発表標題 Truncated estimation for functional linear model and its application to agricultural data
3. 学会等名 5th International Conference on Econometrics and Statistics (国際学会)
4. 発表年 2022年

1. 発表者名 Zeng, Y., Shimizu, S., Matsui, H. and Sun, F.
2. 発表標題 Causal discovery for linear mixed data
3. 学会等名 1st Conference on Causal Learning and Reasoning (国際学会)
4. 発表年 2022年

1. 発表者名 中島誉也, 佐藤俊太郎, 松井秀俊, 水野篤
2. 発表標題 非線形混合効果モデルに基づく関数データを用いた急性期循環器疾患における急性期離脱時期の予測
3. 学会等名 日本臨床疫学会第5回年次学術大会
4. 発表年 2022年

1. 発表者名 松井秀俊
2. 発表標題 空間関数データに対するスパース推定
3. 学会等名 2022年度統計関連学会連合大会
4. 発表年 2022年

1. 発表者名 福井悠介, 平塚真之輔, 井口修巳, 吉村雅寛, 永谷幸裕, 西原明日香, 松井秀俊
2. 発表標題 MRI検査内容における分散表現取得方法の見当
3. 学会等名 第4回日本メディカルAI学会学術集会
4. 発表年 2022年

1. 発表者名 土井博文, 伊藤ゆり, 松井秀俊, 佐浦隆一
2. 発表標題 関数データ解析を用いた長距離走に伴う疲労発生の識別検知
3. 学会等名 日本計算機統計学会第36回大会
4. 発表年 2022年

1. 発表者名 松井秀俊
2. 発表標題 変化係数関数切断線形モデルの推定
3. 学会等名 2021年度統計関連学会連合大会
4. 発表年 2021年

1. 発表者名 渡邊翔太郎, 松井秀俊
2. 発表標題 Classification from Positive and Biased Negative Data with Skewed Labeled Posterior Probability
3. 学会等名 第24回情報論的学習理論ワークショップ (IBIS2021)
4. 発表年 2021年

1. 発表者名 松井秀俊
2. 発表標題 関数データに基づく統計的モデリングと農業データ分析への応用
3. 学会等名 大阪大学数理・データ科学セミナー
4. 発表年 2021年

1. 発表者名 松井秀俊
2. 発表標題 関数回帰モデルに対するドメイン選択と収量データ分析への応用
3. 学会等名 九州大学統計データサイエンス研究集会
4. 発表年 2021年

1. 発表者名 Matsui, H.
2. 発表標題 Truncated estimation for varying-coefficient functional linear models
3. 学会等名 14th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2021)
4. 発表年 2021年

1. 発表者名 Matsui, H.
2. 発表標題 Truncated estimation for varying-coefficient functional linear models and its application to agricultural data
3. 学会等名 The 11th Conference of the Asian Regional Section of the IASC (IASC-ARS 2022)
4. 発表年 2022年

1. 発表者名 Matsui, H.
2. 発表標題 Variable selection in varying-coefficient functional linear models.
3. 学会等名 13th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2020) (招待講演) (国際学会)
4. 発表年 2020年

1. 発表者名 杉山聖貴, 松井秀俊
2. 発表標題 個人向けローン利用顧客のデフォルト判別のための転移学習
3. 学会等名 第23回情報論的学習理論ワークショップ (IBIS2020)
4. 発表年 2020年

1. 発表者名 松井秀俊
2. 発表標題 変化係数関数加法モデルと農業データ分析への応用
3. 学会等名 2020年度統計関連学会連合大会
4. 発表年 2020年

1. 発表者名 Matsui, H
2. 発表標題 Varying-coefficient functional additive models
3. 学会等名 12th International Conference of the ERCIM WG on Computational and Methodological Statistics (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 Arai, N., Misumi, T., Matsui, H., Maesono, Y. and Konishi, S.
2. 発表標題 Automatic multivariate functional clustering for spatial longitudinal data
3. 学会等名 12th International Conference of the ERCIM WG on Computational and Methodological Statistics (国際学会)
4. 発表年 2019年

1. 発表者名 Misumi, T. and Matsui, H.
2. 発表標題 Joint modeling and estimation for multivariate longitudinal data and binary outcome
3. 学会等名 12th International Conference of the ERCIM WG on Computational and Methodological Statistics (国際学会)
4. 発表年 2019年

1. 発表者名 Matsui, H
2. 発表標題 Regularized interaction models for function-on-function regression
3. 学会等名 3rd International Conference on Econometrics and Statistics (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 西内俊策, 松井秀俊.
2. 発表標題 栽培記録から見出された水稻の環境応答変動性の解析
3. 学会等名 日本育種学会第137回講演会
4. 発表年 2020年

1. 発表者名 新井仁智, 三角俊裕, 松井秀俊, 小西貞則, 前園宜彦
2. 発表標題 Automatic multivariate functional clustering for spatial longitudinal data
3. 学会等名 情報・統計科学シンポジウム
4. 発表年 2019年

1. 発表者名 松井秀俊, 持田恵一
2. 発表標題 変化係数関数回帰モデルによる作物収穫データの分析
3. 学会等名 2019年度統計関連学会連合大会
4. 発表年 2019年

1. 発表者名 松井秀俊, 山本倫生, 荒木由布子
2. 発表標題 関数回帰分析
3. 学会等名 日本行動計量学会 秋の行動計量セミナー (招待講演)
4. 発表年 2019年

1. 発表者名 松井秀俊
2. 発表標題 関数データにおける変数選択と時系列・空間データ解析
3. 学会等名 応用統計学会・計量生物学会チュートリアルセミナー (招待講演)
4. 発表年 2019年

1. 発表者名 Terada, Y., Matsui, H.
2. 発表標題 Dynamic prediction for variable-domain functional data
3. 学会等名 the 12th conference of the Asian Regional Section of the International Association for Statistical Computing (IASC-ARS 2023) (招待講演)(国際学会)
4. 発表年 2023年

1. 発表者名 土井博文, 松井秀俊, 西岡大輔, 伊藤ゆり, 佐浦隆一
2. 発表標題 関数データ解析を用いた異なるサンプリングレートのランニングデータの相互活用の検討
3. 学会等名 日本計算機統計学会第37回シンポジウム
4. 発表年 2023年

1. 発表者名 中島誉也, 佐藤俊太郎, 松井秀俊, 水野篤
2. 発表標題 循環器疾患における乳酸値変化を用いた循環安定の予測: 混合効果モデルに基づく関数データ解析の臨床的応用
3. 学会等名 2023年度統計関連学会連合大会
4. 発表年 2023年

1. 発表者名 松井秀俊, 寺田吉壱
2. 発表標題 可変ドメイン関数データに対する軌道の動的予測
3. 学会等名 2023年度統計関連学会連合大会
4. 発表年 2023年

1. 発表者名 松井秀俊
2. 発表標題 関数データに基づく農作物収穫量の分析
3. 学会等名 日本計算機統計学会第37回大会
4. 発表年 2023年

1. 発表者名 廣瀬慧, 増田弘毅, 松井秀俊
2. 発表標題 合計値回帰モデルのクラスタリングによる二重降下現象と電力需要量予測への応用
3. 学会等名 2023年度統計関連学会連合大会
4. 発表年 2023年

〔図書〕 計6件

1. 著者名 松井 秀俊	4. 発行年 2023年
2. 出版社 学術図書出版社	5. 総ページ数 256
3. 書名 多変量解析	

1. 著者名 D.P.Kroese、金森 敬文	4. 発行年 2022年
2. 出版社 東京化学同人	5. 総ページ数 416
3. 書名 データサイエンスと機械学習	

1. 著者名 笛田 薫、松井 秀俊	4. 発行年 2022年
2. 出版社 日経 B P	5. 総ページ数 240
3. 書名 Excelで学べるデータサイエンス入門講義	

1. 著者名 竹村 彰通、姫野 哲人、高田 聖治、和泉 志津恵、市川 治、梅津 高朗、北廣 和雄、齋藤 邦彦、佐藤 智和、白井 剛、田中 琢真、槇田 直木、松井 秀俊	4. 発行年 2021年
2. 出版社 学術図書出版社	5. 総ページ数 240
3. 書名 データサイエンス入門 第2版	

1. 著者名 松井秀俊，小泉和之（著），竹村彰通（編）	4. 発行年 2019年
2. 出版社 講談社サイエンティフィク	5. 総ページ数 224
3. 書名 統計モデルと推測（データサイエンス入門シリーズ）	

1. 著者名 滋賀大学，PwC Japan有限責任監査法人（編）	4. 発行年 2024年
2. 出版社 中央経済社	5. 総ページ数 200
3. 書名 「A I 監査」の基本と技術 データサイエンティストの活躍	

〔産業財産権〕

〔その他〕

滋賀大学 研究者情報総覧
https://researchers.shiga-u.ac.jp/html/100002604_ja.html
Hidetoshi Matsui
<https://sites.google.com/site/hidetoshimatsui/home>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究協 力者	三角 俊裕 (Misumi Toshihiro) (40817300)	国立研究開発法人国立がん研究センター (82606)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------