

令和 5 年 6 月 10 日現在

機関番号：17701

研究種目：若手研究

研究期間：2018～2022

課題番号：18K18011

研究課題名（和文）ノンパラメトリック回帰手法の展開研究

研究課題名（英文）Developments of nonparametric regression

研究代表者

吉田 拓真（Yoshida, Takuma）

鹿児島大学・理工学域理学系・准教授

研究者番号：80707141

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：本研究課題は5年間の研究期間で、6編の学術論文が国際ジャーナルから掲載された。また、関連した話題で9件の国内会議で、2件の国際会議で成果を報告した。得られたすべての成果は研究計画通り、ノンパラメトリック回帰を極値分位点、高次元データ、機械学習理論の応用のいずれかに関するもので、当テーマの深化的発展に貢献できたと言える。

研究成果の学術的意義や社会的意義

開発されたノンパラメトリック回帰手法により、近年話題となっている大規模データの解析や、データの最大値をより正確に予測する方法を提案できた。このような研究はただちに実用科学に応用できるものではないが、データ解析においては、数ある統計手法の共通点、相違点からそれぞれの問題について適切な手法を用いることが重要であり、手法が豊富に存在すること自体が重要である。今回の研究はその一端を担うものである。

研究成果の概要（英文）：During the five-year period of this research project, six academic papers were published in international journals. In addition, the results were reported in 9 national conferences and 2 international conferences on related topics. All the results obtained were related to nonparametric regression with extreme quantiles, high-dimensional data, or applications of machine learning theory, as planned, and contributed to the deepening and development of our research theme.

研究分野：統計科学

キーワード：ノンパラメトリック平滑化 高次元データ 次元削減 極値統計学

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

1. 研究開始当初の背景

ノンパラメトリック回帰の研究は1990年代に黎明期を迎え、現代まで時代のニーズと共に発展してきた。その大きな流れの中で未開発な分野・応用はまだ多く、それらをひとつひとつ調べ、手法と理論の枠組みを拡張していく必要性があった。特に、極値分位点回帰への応用、パラメトリックモデルとノンパラメトリック法の融合モデリング、機械学習ベースの手法への応用は拡張の余地があり、統計的でも重要な応用先を持つテーマであった。

極値分位点回帰については、自然災害のリスク予測などに有用であることが知られている。しかし、回帰の枠組みで考えられている研究は少なく、また、漸近収束レートの導出は未だに理論整備がなされていなかった。統計学の問題は数学的な問題として現れることも多いため、どのような条件で最適な性質を有するのかを調べることは手法構築の上で重要である。また理論から得られる信頼区間の構成も応用上は役立つ場合も多いため、統計理論とさらなる推測は極値分位点回帰の深化に必要な課題であった。

パラメトリックモデルとノンパラメトリック法の融合した研究は基本的な方法はすでに得られているが、スパースモデリングなど最新の統計手法と組み合わせた新規手法は研究例が少ない。特に、ノンパラメトリック補完法については、従来は統計的検定などが主流であったが、推定ベースの方法は未だ未開発であった。しかし最新手法がもたらす利益を活かし、より高性能な推定ベースの手法構築が可能となれば計算効率や解釈性についても言及することができる。

機械学習法を統計学的に評価する研究はまさに現代の統計科学の主流のひとつであった。特にブースティングやランダムフォレストのようなアルゴリズムベースの手法と理論研究はその重要性を増していた。それらの手法は弱い推定量(学習器)を重ねることで強力な推定量を構成するものであるが、その重ね方(係数の推定)が複雑である。一方で強い学習器をデータ依存でない係数を付与して重ねる方法は開発が進んでいなかった。このような方法は係数がデータ依存でないので統計理論を構成しやすく、理論的には非漸近理論を構成でき、従来の漸近理論によらない別角度から理論的な考察を与えることができる。

2. 研究の目的

本研究課題の目的は、複数の変量間の関数関係を解析する回帰分析において、線形などの統計モデルでは表現しきれない複雑な構造を持つデータに対する有用な方法であるノンパラメトリック法の発展と体系の整備である。展開の方向性として、具体的に以下の4つのテーマに取り組むことであった：

- (A) 極値分位点回帰の漸近理論の構築
- (B) 極値分位点回帰の信頼区間の構成
- (C) ノンパラメトリック補完法の発展
- (D) 逐次推定法へのスパース性の導入

上記4研究において、(A)と(B)は相互関連があり、(B)を達成するためには(A)の結果が必要となる。一方で(C)、(D)は(A)、(B)とは独立に推進する。上記4研究はすべて申請者のこれまでの研究を発展させた内容であるため、継続的なノンパラメトリック法の深化研究と捉えることができる。

3. 研究の方法

(A) 極値分位点は通常のノンパラメトリック分位点回帰と、極値分析特有の“外挿”と呼ばれる方法の融合によって構成される。本研究では回帰ではない1次元の極値分析の理論とノンパラメトリック回帰の理論を組み合わせることで目的を達成する。モデルの説明変数の連続性について注意深く議論し、弱い仮定で漸近的性質が成立するかどうかを確かめる。また、極値回帰の収束レートは極値ではない通常の分位点回帰のものとは比べ若干遅くなるはずであり、そこでどの程度の差があるのかを定量的に調べ、結果としてまとめたい。さらにこの研究は目的が達成されれば多次元モデルへの発展が可能となる。一定の成果が得られれば次々とモデルを進展させ、成果を充実・展開させていく。

(B) 極値分位点回帰においては各点の挙動が不安定なため、一様なバンド、つまり同時信頼区間の構成が重要となる。起点となる漸近的性質は研究(A)から得られているため、それを応用させる。信頼区間の構成にはVolume of Tube Formula(VTF)と呼ばれる方法の利用を考えている。本研究はVTFと極値理論の融合により同時信頼区間を構成する。その際に推定量の一樣収束の理論が必要となるが、それは蓄積されたノンパラメトリック回帰の漸近理論から克服できると予想している。また本手法はスプライン法を用いる。その理由はVTF構成において必要な関数の微分計算を可能とするためである。他の方法では基本的に数値微分が必要となり、端点で不安定な挙動をする可能性がある。大枠の研究推進の中の細かい部分で手法や理論の選定を行う必要があるが、その手法を選ぶ理由を明確にし、各段階で最適な方法を用いる根拠を付けて研究をすすめていく。

(C) パラメトリックモデルの誤差をノンパラメトリック法で補完するが、通常のノンパラメトリック法をそのまま用いると、補完の必要がない箇所も過剰に反応し、補完(過適合)してしまう。よって、lassoなどのスパース法をノンパラメトリック項に適用することにより、補完の必要の有無も確認する。それにより、推定量の解釈性の向上も期待できる。また提案する手法は局所的な補完を施すものであり、パラメトリックモデルでは捉えきれないスパイクを正確に補完できると考えている。そのため、適用可能な例を多くあげ、手法の汎用性を示したい。

(D) これまでに代表者が提案した逐次推定法に変数の特徴抽出を導入する。そのために、スパース法を導入し、不必要な変数は更新の度に除去していくアルゴリズムを構成していく。本手法は変数を1つずつモデルに組み込むため、スパース性の導入自体は容易であるが計算コストの軽減が課題である。この計算効率性を改善するためにモンテカルロ法を用いたアルゴリズムを開発する。この手法はランダムフォレストなど既存のアルゴリズムベースの推定法に通じるものがある。しかし、係数をデータから推定されるものでなく固定値を用いるため、それらとは明確な差別化がなされる。係数をデータから推定する場合は外れ値を含んでいる場合など想定外の影響を受けやすい。したがって、提案手法は良い意味でデータの特性にあまり依存せず、生きる場面も多い。そのようなケースを丁寧に紹介し、既存手法と比べた利点を明確にしていく。

4. 研究成果

- (A) この研究は順調に進捗し、関連研究で成果を3編の論文として Revstat-Statistical Journal, Communications in Statistics-Theory and Methods, Advances in Statistical Analysis, 発表された。
- (B) 本研究も(A)直後に取り組み成果を得た。成果は論文として、Extremes から発表された。
- (C) 本研究は一般的な問題ではなく、ある特定の条件下でうまく機能する構造を発見した。得られた成果は3編の論文として、Statistical Papers, Econometrics and Statistics, Canadian Journal of Statistics から発表された。
- (D) この研究はまず逐次推定の理論研究として1編の論文にまとめられ、Computational Statistics から発表された。特徴抽出は同論文内で探索的な方法として提案した。スパース性の導入については論文としてまとめられたが、ジャーナル採択には至っておらず、引き続き研究を進める。

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 6件 / うち国際共著 0件 / うちオープンアクセス 1件）

| | |
|--|-------------------------|
| 1. 著者名 Takuma Yoshida | 4. 巻 49 |
| 2. 論文標題 Quantile function regression and variable selection for sparse models | 5. 発行年 2021年 |
| 3. 雑誌名 Canadian Journal of Statistics | 6. 最初と最後の頁 1196 1221 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/cjs.11616 | 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 Takuma Yoshida | 4. 巻 105 |
| 2. 論文標題 Additive models for extremal quantile regression with Pareto-type distributions | 5. 発行年 2021年 |
| 3. 雑誌名 ASTA Advances in Statistical Analysis | 6. 最初と最後の頁 103-135 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10182-020-00386-1 | 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 Takuma Yoshida | 4. 巻 - |
| 2. 論文標題 Nonparametric smoothing for extremal quantile regression with heavy tailed data | 5. 発行年 2020年 |
| 3. 雑誌名 REVSTAT-Statistical Journal- | 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 - |
| 1. 著者名 Takuma Yoshida | 4. 巻 23 |
| 2. 論文標題 Simultaneous confidence bands for extremal quantile regression with splines | 5. 発行年 2020年 |
| 3. 雑誌名 Extremes | 6. 最初と最後の頁 117-149 |
| 掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10687-019-00360-4 | 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 Takuma Yoshida | 4. 巻 - |
| 2. 論文標題 Extreme value inference for quantile regression with varying coefficients | 5. 発行年 2020年 |
| 3. 雑誌名 Communications in Statistics-Theory and Methods | 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) 10.1080/03610926.2019.1639752 | 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 Takuma Yoshida and Kanta Naito | 4. 巻 134 |
| 2. 論文標題 Regression with stagewise minimization on risk function | 5. 発行年 2019年 |
| 3. 雑誌名 Computational Statistics and Data Analysis | 6. 最初と最後の頁 123 ~ 143 |
| 掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.csda.2018.12.011 | 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 - |

〔学会発表〕 計8件 (うち招待講演 0件 / うち国際学会 1件)

| |
|--|
| 1. 発表者名 吉田拓真 |
| 2. 発表標題 極値統計学におけるセミパラメトリックモデリングについて |
| 3. 学会等名 広島大学金曜セミナー |
| 4. 発表年 2022年 |

| |
|--|
| 1. 発表者名 吉田拓真, 桃木光輝 |
| 2. 発表標題 極値統計学における予測モデリングについて |
| 3. 学会等名 科研費シンポジウム「データサイエンス・統計学における方法論と応用の新展開」 |
| 4. 発表年 2021年 |

| |
|---|
| 1. 発表者名 吉田拓真, 本間剛介 |
| 2. 発表標題 Single index models for conditional extremes |
| 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 Takuma Yoshida |
| 2. 発表標題 Simultaneous confidence bands for extremal quantile regression with splines |
| 3. 学会等名 11th International Conference on Extreme Value Analysis (国際学会) |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 吉田拓真 |
| 2. 発表標題 Nonparametric estimation of extreme value index for heavy tailed distributions |
| 3. 学会等名 第33回計算機統計学会シンポジウム |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 吉田拓真 |
| 2. 発表標題 ノンパラメトリックな極値統計解析法の漸近理論 |
| 3. 学会等名 科研費シンポジウム「大規模複雑データの理論と方法論の総合的研究」 |
| 4. 発表年 2019年 |

| |
|---|
| 1. 発表者名 Takuma Yoshida |
| 2. 発表標題 Extreme Inference for Nonparametric Quantile Regression with Heavy Tailed Data |
| 3. 学会等名 Joint Statistical Meeting 2018 |
| 4. 発表年 2018年 |

| |
|--|
| 1. 発表者名 Takuma Yoshida |
| 2. 発表標題 Extreme Value Inference for Nonparametric Quantile Regression |
| 3. 学会等名 シンポジウム「統計・機械学習の交わりと拡がり」 |
| 4. 発表年 2018年 |

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

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 |
|---------------------------|-----------------------|----|
|---------------------------|-----------------------|----|

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 |
|---------|---------|
|---------|---------|