

令和 4 年 6 月 25 日現在

機関番号：10104

研究種目：基盤研究(C)（一般）

研究期間：2019～2021

課題番号：19K01582

研究課題名（和文）超高次元データ解析のためのモデル平均法

研究課題名（英文）Model Averaging for Ultra-High Dimensional Data: Theory, Methods, and Applications

研究代表者

劉 慶豊 (LIU, QINGFENG)

小樽商科大学・商学部・教授

研究者番号：60378958

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：ビッグデータの時代が迎えられて、大規模データに適したデータ解析の方法が求められていた。超高次元データ解析の分野ではモデル（共変量）選択のミスペシファイケーションの課題と計算量と計算スピードの課題の解決が重要なテーマであった。本研究は超高次元データ解析の精度を高めるためのモデル平均法の開発した。さらに、モデル平均法の計算スピードを上げるための方法も提案した。また、より広範囲に適用できるモデル平均法の開発を行い、ファイナンス分野にお応用できる方法を提案した。すべての研究成果は5本の論文として纏められて、既に国際学術雑誌に掲載された。

研究成果の学術的意義や社会的意義

本研究の成果は、モデル平均法の分野では国際的に最先端の課題のチャレンジで、当該分野の発展に大きく貢献していると言える。ビッグデータの時代において、大規模データに適したデータ解析の方法を提供した。学術と社会の発展に技術の面で寄与した。

研究成果の概要（英文）：We developed several novel model averaging methods, and provided a solution to the high computational cost issue of model averaging. The achievements include model averaging method of OLS and GLS estimators, model averaging method of nonlinear model, model averaging method for GARCH-type models and model averaging method for Ultra-High Dimensional Data.

研究分野：経済統計学

キーワード：モデル平均 モデル選択 非線形モデル 超高次元データ

1. 研究開始当初の背景

研究開始当初では、データ収集技術の発展に伴い、想像を超える大規模なデータが蓄積されている状況であった。ビッグデータの時代が迎えられて、大規模データに適したデータ解析の方法が求められていた。超高次元データ解析の分野ではモデル(共変量)選択のミスペシファイケーションの課題と計算量と計算スピードの課題の解決が重要なテーマであった。当時、モデル平均法は推定の精度を高めて、リスクを大きく下げることができると同時にモデルのミスペシファイケーションの問題点のある程度克服できるため注目されている。しかし、データが高次元になるにつれ計算量が膨大になり、特に共変量の次元 P がサンプルサイズ n を大きく超える超高次元の状況では、実用的なモデル平均法とそれに関する統計理論が十分に確立されていなかった。

2. 研究の目的

本研究は共変量の次元 P がサンプルサイズ n を大きく超える超高次元の状況でも、実用的な超高次元データ解析の精度を高めるためのモデル平均法の開発を目的とする。さらに、そのモデル平均法に関する統計理論を構築する。具体的には超高次元の状況で大部分の共変量の係数がゼロであると仮定し、スパース推定を用いて次元を削減してから選ばれた共変量にモデル平均法を適用する二段階推定法(Post-LASSO Model Averaging)を構築する。さらに、モデル平均法の計算スピードを上げることも本研究の目的である。また、より広範囲に適用できるモデル平均法の開発も本研究のもう一つの目的である。

3. 研究の方法

本研究ではいくつかのモデル平均法を構築して、それぞれの方法の統計学的な性質を解析的に調べた。さらに、シミュレーション実験で各方法の有限標本の性質を確認した。その上、いくつかの方法に関して経済データに適用し、実証研究を行なった。

4. 研究成果

研究期間中で以下の研究成果を出した。

- 異なる推定法をベースにしたモデル平均法を開発した。それまでの頻度論のモデル平均法は殆ど同じ推定法による推定量の平均であった。本研究は最小二乗法による推定量と一般化最小二乗法による推定量を一緒に平均化する方法を提案した。さらに、その方法は推定のリスクを小さくする性質を持っていることを示した。この研究成果はモデル平均法の研究の新しい可能性を示した。
- 非線形モデルに適用できる平均法を開発した。このモデル平均法は機械学習のモデルを含めた殆どの非線形モデルの定式に適用できることで適用範囲が極めて広い。この研究は開発した非線形モデルの平均法の統計的性質である、最適性と平均のためのウエイトの収束性を一定の条件のもとで示した。さらに、シミュレーション実験で既存方法より優れたパフォーマンスを持っていることを確認した。実証研究として、賃金関数の推定を行った。実証研究におけるモデル平均法の結果によって、賃金関数の非線形性の重要度が明らかになった。この研究成果の論文は既に数多くの研究論文に引用されている。
- ポラティリティモデルの平均法を開発した。この研究成果はモデル平均法の応用を計量ファイナンスの分野へ広げた。本研究はKullback-Leibler divergenceをリスク関数として、それを小さくすることを目指して、モデル平均のウエイトを選ぶための情報量基準を提案した。開発した方法の応用として、主な株式市場の指数のポラティリティを推定し予測した。既存方法より優れていることを示した。この研究成果の論文も既に数多くの研究論文に引用されている。
- モデル平均法のLASSO表現を提案し、そのウエイトのスパース性を示した上で、モデル平均法を高次元データに適用するための高速アルゴリズムを開発した。今までのモデル平均法の計算スピードの問題の解決に道を開いた。この研究は新しい高速なアルゴリズムとしてcoordinate-wise descent algorithmを提案した。この新しいアルゴリズムによる計算は既存の方法より、共変量の数が膨大な場合、計算スピードが数十倍速くなる。
- 超高次元データに適用できるモデル平均の手法としてLASSOなどの手法のsolution pathを利用したモデル平均法を開発した。超高次元データ解析の分野でのモデル(共変量)選択のミスペシファイケーションの課題に新しい解決法を提供した。具体的には次元を削減するために、まずLASSOなどの高次元データ分析の方法で推定を行い、その推定結果である異なるチューニングパラメーターによるソリューションパスをモデル平均に利用する。そうすることで、超高次元データに伴う計算量の問題を解決しながら、推定のリスクをモデル平均法で小さく抑えることができた。この研究で開発したモデル平均法の応用として、アメリカの犯罪率の分析を行なった。
- 欠損値のあるデータのモデル平均法を開発した。この研究成果はデータに欠損値が存在

する場合において、一般化線形モデルのためのモデル平均法を提案した。モデル平均のウエイトはこの研究で提案した情報量基準を利用して、Kullback-Leibler divergence が小さくなるように選ばれる。

以上の成果は、モデル平均法の分野では国際的に最先端の課題のチャレンジで、当該分野の発展に大きく貢献していると言える。以上の成果は既に6本の論文にまとめて、国際学術誌に掲載した。

さらに、最終年度でモデル平均に関する更なる知見とそれまでの研究成果の発展を求めて、最終年度でモデル平均法の考え方を機械学習の分野へ拡張して、アンサンブル学習の開発に取り込んでいる。異なる種類の手法の長所を発揮させるための機械学習のモデルを結合させるMachine Collaborationの研究を展開してきた。その初期的な成果を論文にまとめた。Machine Collaborationは既存の平行型(Staking)やシクエンシャル型(Boosting)のアンサンブル学習の手法と異なって、情報の伝達は周回するため、学習システム全体の最適化が実現できる。複数の機械学習の手法を総合利用するところはモデル平均の発想の拡張となっている。

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 6件/うちオープンアクセス 2件）

1. 著者名 Feng Yang, Liu Qingfeng, Yao Qingsong, Zhao Guoqing	4. 巻 Online
2. 論文標題 Model Averaging for Nonlinear Regression Models	5. 発行年 2021年
3. 雑誌名 Journal of Business & Economic Statistics	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） 10.1080/07350015.2020.1870477	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 該当する
1. 著者名 Feng Yang, Liu Qingfeng	4. 巻 9
2. 論文標題 Nested model averaging on solution path for high dimensional linear regression	5. 発行年 2020年
3. 雑誌名 Stat	6. 最初と最後の頁 1~13
掲載論文のDOI（デジタルオブジェクト識別子） 10.1002/sta4.317	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 該当する
1. 著者名 Qingfeng Liu, Miaomiao Zheng	4. 巻 11
2. 論文標題 Model Averaging for Generalized Linear Model with Covariates that are Missing completely at Random	5. 発行年 2020年
3. 雑誌名 数理経済研究	6. 最初と最後の頁 25-40
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 オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Liu Qingfeng, Vasnev Andrey L.	4. 巻 7
2. 論文標題 A Combination Method for Averaging OLS and GLS Estimators	5. 発行年 2019年
3. 雑誌名 Econometrics	6. 最初と最後の頁 38~38
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/econometrics7030038	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 該当する

1. 著者名 Feng Yang, Liu Qingfeng, Okui Ryo	4. 巻 187
2. 論文標題 On the sparsity of Mallows model averaging estimator	5. 発行年 2020年
3. 雑誌名 Economics Letters	6. 最初と最後の頁 108916 ~ 108916
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.econlet.2019.108916	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 該当する

1. 著者名 Liu Qingfeng, Yao Qingsong, Zhao Guoqing	4. 巻 Online First
2. 論文標題 Model averaging estimation for conditional volatility models with an application to stock market volatility forecast	5. 発行年 2020年
3. 雑誌名 Journal of Forecasting	6. 最初と最後の頁 1-23
掲載論文のDOI (デジタルオブジェクト識別子) 10.1002/for.2659	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 該当する

[学会発表] 計8件 (うち招待講演 7件 / うち国際学会 5件)

1. 発表者名 劉慶豊
2. 発表標題 Machine Collaboration
3. 学会等名 日本経済学会2021秋季大会
4. 発表年 2021年

1. 発表者名 劉慶豊
2. 発表標題 On the sparsity of Mallows model averaging estimator
3. 学会等名 4th International Conference on Econometrics and Statistics (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 劉慶豐
2. 発表標題 Machine Collaboration
3. 学会等名 Applications of Data Science in Social Science, TRANSDISCIPLINARY ECONOMETRICS & DATA SCIENCE SEMINAR (招待講演)
4. 発表年 2022年

1. 発表者名 劉慶豐
2. 発表標題 Machine Collaboration
3. 学会等名 科研費シンポジウム多様な分野における統計科学に関する理論と方法論の革新的展開 (招待講演)
4. 発表年 2021年

1. 発表者名 Qingfeng Liu
2. 発表標題 Model averaging estimation for conditional volatility models with an application to stock market volatility forecast
3. 学会等名 The Ninth International Symposium of Quantitative Economics, August 26, 2020, (Online, Changchun, China) (招待講演) (国際学会)
4. 発表年 2020年

1. 発表者名 Qingfeng Liu
2. 発表標題 A Machine Learning Perspective on Model Averaging
3. 学会等名 ACIEK(Winter)- IMIP 2020 (招待講演) (国際学会)
4. 発表年 2020年

1. 発表者名 劉慶豐
2. 発表標題 On the Sparsity of Mallows Model Averaging Estimator
3. 学会等名 The 2019 International Symposium of Quantitative Economics (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 劉慶豐
2. 発表標題 Model Averaging for Nonlinear Regression Models
3. 学会等名 The 2019 ICOSA China Conference (招待講演) (国際学会)
4. 発表年 2019年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

<p>Current Working Papers https://qingfeng-liu.github.io/wp.html Selected Publications https://qingfeng-liu.github.io/publication.html</p>
--

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1件

<p>国際研究集会 TRANSDISCIPLINARY ECONOMETRICS & DATA SCIENCE SEMINAR</p>	<p>開催年 2021年～2022年</p>
--	-------------------------------------

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
---------	---------