

令和 6 年 6 月 13 日現在

機関番号：32689

研究種目：基盤研究(C)（一般）

研究期間：2019～2023

課題番号：19K12128

研究課題名（和文）スパースモデリングとベイズ決定理論に基づいた因果推論手法の構築

研究課題名（英文）Development of Causal Inference Methods Based on Sparse Modeling and Bayesian Decision Theory

研究代表者

堀井 俊佑（Horii, Shunsuke）

早稲田大学・データ科学センター・准教授

研究者番号：00552150

交付決定額（研究期間全体）：（直接経費） 1,900,000円

研究成果の概要（和文）：本研究は、データ分析における誤った知見を避けるため、統計的因果推論を統計的決定理論の枠組みで最適化することを目的とした。まず、ベイズ決定理論に基づく因果効果の推定法を提案し、シミュレーションおよび実データを用いた実験でその有効性を示した。また、MCMC法や変分ベイズ法を用いた効率的な近似アルゴリズムを構築し、計算量の問題を解決した。これにより、小規模データでも高精度の推定が可能となった。さらに、操作変数を利用した因果推論や条件付き平均処置効果の推定法も提案し、従来手法に対する優位性を実証した。研究成果はAISTATSやAAAIなどのトップ会議に採択され、統計学・人工知能分野に貢献した。

研究成果の学術的意義や社会的意義

本研究の学術的意義は、統計的因果推論を統計的決定理論の枠組みで最適化し、従来の方法よりも精度の高い因果効果の推定を可能にした点にある。これにより、因果推論の理論体系が深化し、多様なデータ分析における新たな知見の創出が期待される。社会的意義としては、データに基づく意思決定の精度向上が挙げられる。特に医療や経済学などの分野で、因果関係の正確な把握に基づく政策や治療法の最適化が可能となり、公共の福祉や経済の発展に寄与することが期待される。

研究成果の概要（英文）：This study aimed to optimize statistical causal inference within the framework of statistical decision theory to avoid erroneous insights in data analysis. First, a Bayesian decision theory-based method for estimating causal effects was proposed and its effectiveness was demonstrated through simulations and experiments with real data. Additionally, efficient approximation algorithms using MCMC and variational Bayesian methods were developed to address computational complexity issues, enabling high-precision estimation even with small-scale data. Furthermore, methods for causal inference using instrumental variables and for estimating conditional average treatment effects were proposed, demonstrating superiority over conventional methods. The research outcomes were presented at top conferences such as AISTATS and AAAI, contributing to the fields of statistics and artificial intelligence.

研究分野：統計的因果推論

キーワード：統計的因果推論 構造的因果モデル 潜在反応モデル 統計的決定理論 ベイズ統計学 スパースモデリング

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

計測技術・ネットワーク技術の向上に伴うデータ収集コストの低下や、多変量解析・機械学習を用いたデータ分析ソフトウェアの開発に伴い、大量のデータを分析することが一般的になってきていた。しかし、データを分析する際に、分析の目的やデータの背後にある構造を意識しないで分析を行うことで、誤った知見や無意味な分析をしてしまうことも多いという問題があった。何らかの行動をした場合の結果を統計的に分析する方法として**統計的因果推論**の研究が注目を浴びていた。統計的因果推論には様々なアプローチがあるが、中でも**因果ダイアグラム・構造方程式モデル**に基づいた因果推論は統計学や人工知能分野で活発に研究が行われている。

構造的因果モデルに基づいた因果推論は、因果ダイアグラムという変数間の因果関係をグラフで表現したものに基いて因果効果を推定する。そのため、因果効果の推定は、1. 専門知識に基づいて因果ダイアグラムを構築するか、データに基づいて因果ダイアグラムを推定する、2. データから因果ダイアグラムを規定する確率分布を推定する、3. 因果効果を計算する、という手順で計算されるのが一般的であった。しかし、因果効果の推定を**統計的決定理論**の枠組みで定式化すると、この手順で推定することは必ずしも最適とはならないということが研究代表者の研究によって示されていた。このように統計的決定理論の枠組みで統計的因果推論を扱っている研究は、研究代表者が行っているのみであった。その主な理由として、

1. 因果ダイアグラムの推定と因果効果の推定は分けて良いと考えられている
2. 最適な推定法により推定した場合の恩恵の大きさが認知されていない

ということが挙げられた。

2. 研究の目的

本研究の主な問いは以下の2つであった。

・統計的因果推論の様々なモデルにおいてベイズ決定理論に基づくアプローチをとることの有効性はどの程度か？

・統計的因果推論の様々なモデルにおいて、スパースモデリングの考え方やMCMC法・変分ベイズ法などを応用することで、ベイズ最適な決定法を効率的に近似計算することは可能か？

本研究の目的は、これら2つの問いに対して、統計的因果推論、機械学習、最適化理論、ベイズ決定理論、スパースモデリングの分野の知見を融合したアプローチによって、肯定的な回答を与えることであった。

研究代表者はすでに本研究開始前に構造的因果モデルに基づく因果推論アプローチにベイズ決定理論の考え方を組み込むという試みに取り組んでおり、新たなフレームワークができつつあった。他の様々な統計的因果推論の方法においても同様のアプローチをとることで、統計的因果推論の理論体系全体に広がりができることが期待されていた。

3. 研究の方法

本研究では、上記の2つの問いに対して、3つの研究課題を設定して研究を行った。

1. 様々な統計的因果推論モデルに対するベイズ最適な推定法の性質の解明

既存の統計的因果推論モデルにおける強い仮定を、より広いクラスの事前分布を導入することでモデルの拡張を行い、そのもとでベイズ最適な決定を導出する。また、ベイズ最適な決定と従来手法との違いを、シミュレーション実験、実データを用いた実験を通して調べる。

2. ベイズ最適な推定法の理論的性質の解明

古典的な統計的推測の問題や、統計的学習理論の問題の一部では、ベイズ最適な推定法の理論的性質も解明されている。統計的因果推論におけるベイズ最適な因果効果推定法の理論的性質を明らかにすることで、本研究の有効性を示す。

3. MCMC法や変分ベイズ法を用いた効率的な近似アルゴリズムの構築

構造的因果モデルに基づく因果推論では、ベイズ最適な推定量を計算する際の計算量的困難性が問題となった。研究代表者は、スパース性を誘導する事前分布と変分ベイズ法を利用することで近似計算アルゴリズムを構築した。他の統計的因果推論モデルにおいても同様の問題が発生すると考えられるため、1.で得られたベイズ最適な決定法に対する近似計算アルゴリズムを構築する。

4. 研究成果

本研究の予備研究として、真の介入効果と推定介入効果の間の距離をカルバック・ライブラー距離で計る場合のベイズ最適な推定量を導出していたが、その研究の拡張として、平均介入効果の推定問題を扱い、二乗誤差損失を考えた場合のベイズ最適な推定量を導出した。一般的に介入効果を推定する場合、まず変数間の関係性を表す因果ダイアグラムを推定し、推定された因果ダイアグラムのもとで介入効果を推定するという二段階のアプローチがとられるが、本研究で提案した推定法は、各因果ダイアグラムのもとで推定した介入効果の推定量を、因果ダイアグラムの事後確率で期待値をとるというものになる。これにより、サンプルサイズが小さい場合でも平

均的に推定精度の良い推定が可能となることを示した。ただし、この手法は変数の数が増えると、候補となる因果ダイアグラムの数が増加することで、計算量的に計算が困難になるという問題があった。この問題を解決するために、ベイズ的スパースモデリングの知見を活用し、変分ベイズ法に基づく効率的近似アルゴリズムを構築し、その有効性を実験により検証した。特に実データを用いた実験により、提案手法が幅広いデータに対して従来の推定法よりも良い推定精度を示すことを確認した。これらの結果をまとめた論文は、統計学・人工知能分野のトップ会議の1つである AISTATS に採択された。

また、操作変数を利用した因果推論の研究を行った。操作変数とは、処置変数と相関を持ち、目的変数とは処置変数を通して以外では相関を持たない変数であり、操作変数を利用した因果推論は経済学の分野で盛んに研究されている。操作変数は除外制約という制約を満たしている必要があるが、この制約はデータからのみでは検証することが出来ない。本研究では、除外制約を満たさない可能性が数多くあるという状況を考え、そのような場合でも効率的に因果推論が可能な手法を構築した。この研究成果は統計学の国際会議 CMStatistics で発表した。

部分線形モデルとよばれるモデルを拡張したモデルにおいて、条件付き平均処置効果とよばれる量をベイズ的方法で推定するアルゴリズムを提案し、その性能を理論的・実験的に解析を行った。統計的因果推論の分野では近年、因果効果の異質性に注目した研究が多く行われているが、本研究もその一種である。関連する手法として、Double/Debiased Machine Learning とよばれる手法が広く用いられているが、実験によりこの手法と比較して条件付き平均処置効果を高い精度で推定できることを示した。本研究の研究成果は人工知能分野のトップ国際会議 AAAI に採択された。

また、部分線形モデルにおいて、因果効果推定を目的とした能動学習の手法を提案し、その性能を実験的に解析した。例えば、ある新薬の効果を明らかにしたい場合、できるだけ知見者の数を少ない状況で、効果を高い精度で推定したい。本研究は、このように少ないサンプルサイズで処置変数の目的変数に与える因果効果を高精度で推定したい問題で有効である。この研究成果については統計学の国際会議 CMStatistics で発表を行った。

5. 主な発表論文等

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

1. 著者名 野村亮, 堀井俊佑	4. 巻 vol. 120, no. 320, IT2020-75
2. 論文標題 ノイズを含むデータからの仮説検定問題に対する情報スペクトルのアプローチ	5. 発行年 2021年
3. 雑誌名 信学技法	6. 最初と最後の頁 69-74
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 無
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20件（うち招待講演 0件／うち国際学会 11件）

1. 発表者名 Shunsuke Horii, Yoichi Chikahara
2. 発表標題 Uncertainty Quantification in Heterogeneous Treatment Effect Estimation with Gaussian-Process-Based Partially Linear Model
3. 学会等名 38th AAAI Conference on Artificial Intelligence (AAAI-24) (国際学会)
4. 発表年 2024年

1. 発表者名 Shunsuke Horii
2. 発表標題 Bayesian sequential experimental design for Gaussian-process-based partially linear model
3. 学会等名 16th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2023) (国際学会)
4. 発表年 2023年

1. 発表者名 Shunsuke Horii
2. 発表標題 Bayesian Sequential Experimental Design for a Partially Linear Model with a Gaussian Process prior
3. 学会等名 NeurIPS 2022 workshop Gaussian Processes, Spatiotemporal Modeling, and Decision-making Systems (国際学会)
4. 発表年 2022年

1. 発表者名 堀井俊佑
2. 発表標題 Bayesian Sequential Experimental Design for a Partially Linear Model with a Gaussian Process Prior
3. 学会等名 第25回情報論的学習理論ワークショップ (IBIS2022)
4. 発表年 2022年

1. 発表者名 Shunsuke Horii
2. 発表標題 Bayesian Model Averaging for Causality Estimation and its Approximation based on Gaussian Scale Mixture Distributions
3. 学会等名 Proc. of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 Shunsuke Horii
2. 発表標題 Heterogeneous treatment effect estimation based on a partially linear model with a Gaussian process prior
3. 学会等名 14th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2021) (国際学会)
4. 発表年 2021年

1. 発表者名 堀井俊佑, 戸堂康之
2. 発表標題 企業間ネットワーク構造の変化が企業成長に及ぼす因果効果の異質性
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 堀井俊佑
2. 発表標題 ガウス過程事前分布を用いた部分線形モデルによる異質因果効果推定
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 堀井俊佑
2. 発表標題 ガウス過程事前分布を用いた部分線形モデルによる条件付き平均処置効果推定
3. 学会等名 第44回情報理論とその応用シンポジウム予稿集
4. 発表年 2021年

1. 発表者名 Shunsuke Horii
2. 発表標題 Bayes optimal estimator of the mean intervention effect and its approximation based on variational inference
3. 学会等名 13th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 Yuji Iikubo, Shunsuke Horii, Toshiyasu Matsushima
2. 発表標題 Bayesian approach for uplift linear regression
3. 学会等名 13th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 Shunsuke Horii
2. 発表標題 Improved Computation-Communication Trade-Off for Coded Distributed Computing using Linear Dependence of Intermediate Values
3. 学会等名 2020 International Symposium on Information Theory (ISIT2020) (国際学会)
4. 発表年 2020年

1. 発表者名 堀井俊佑
2. 発表標題 Bayesian Model Averaging による因果効果推定と変分ベイズ法に基づく近似アルゴリズム
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 堀井俊佑
2. 発表標題 スパース性を考慮した操作変数を用いた因果推論に対するベイズ的アプローチ
3. 学会等名 第42回情報理論とその応用シンポジウム
4. 発表年 2019年

1. 発表者名 井上一磨, 雲居玄道, 堀井俊佑, 須子統太, 後藤正幸
2. 発表標題 真の因果構造が未知の場合の因果効果の推定精度について
3. 学会等名 第42回情報理論とその応用シンポジウム
4. 発表年 2019年

1. 発表者名 須子統太, 安田豪毅, 堀井俊佑, 小林学
2. 発表標題 潜在クラスを含むラベルノイズモデルにおける分類アルゴリズム
3. 学会等名 第42回情報理論とその応用シンポジウム
4. 発表年 2019年

1. 発表者名 堀井俊佑
2. 発表標題 除外制約を満たさない操作変数のスパース性を考慮した因果推論に対するベイズ的アプローチ
3. 学会等名 第22回情報理論的学習理論ワークショップ
4. 発表年 2019年

1. 発表者名 Shunsuke Horii
2. 発表標題 A Bayesian approach for estimating the causal effects using sparse invalid instrumental variables
3. 学会等名 12th International Conference of the ERCIM WG on Computational and Methodological Statistics (CMStatistics 2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Yuji Iikubo, Shunsuke Horii, Toshiyasu Matsushima
2. 発表標題 Model Selection of Bayesian Hierarchical Mixture of Experts based on Variational Inference
3. 学会等名 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC2019) (国際学会)
4. 発表年 2019年

1. 発表者名 Shunsuke Horii
2. 発表標題 A Note on the Bayes Optimal Estimator of the Expected Intervention Effect
3. 学会等名 11th Asia-Europe Workshop on Concepts in Information Theory (AEW11) (国際学会)
4. 発表年 2019年

〔図書〕 計1件

1. 著者名 松嶋 敏泰、早稲田大学データ科学教育チーム	4. 発行年 2022年
2. 出版社 サイエンス社	5. 総ページ数 192
3. 書名 データ科学入門I	

〔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
---------	---------