

令和 3 年 5 月 13 日現在

機関番号：12102

研究種目：若手研究(B)

研究期間：2017～2020

課題番号：17K12983

研究課題名(和文) 混合整数最適化を用いた制約付き変数選択による高精度パラメータ推定

研究課題名(英文) High-accuracy parameter estimation using constrained variable selection based on mixed-integer optimization

研究代表者

高野 祐一 (Takano, Yuichi)

筑波大学・システム情報系・准教授

研究者番号：40602959

交付決定額(研究期間全体)：(直接経費) 2,000,000円

研究成果の概要(和文)：回帰分析や判別分析における変数選択のための厳密解法として、混合整数最適化が近年注目を集めている。一方で混合整数最適化による変数選択では、選択された説明変数に多重共線性が残ってしまう場合が多く、データに内在する事前知識も活用されていないという欠点がある。そこで本研究では「混合整数最適化」に「多重共線性の除去」と「構造的正則化(事前情報を活用したモデル構築)」を組み合わせ、高精度のパラメータ推定が可能な各種の制約付き変数選択手法を提案した。人工データと実データを用いた数値実験を実施し、提案手法の有効性を検証した。

研究成果の学術的意義や社会的意義

回帰分析や判別分析における変数選択の先行研究では、発見的解法が用いられる場合が多く、パラメータ推定の精度に着目した変数選択の厳密解法は先駆的かつ実用上重要な研究だと言える。多重共線性を除去する制約条件は扱いが難しく、有効な求解アルゴリズムを考案することは最適化理論の観点からも意義がある。本研究の目的である高精度パラメータ推定は、データ分析の信頼性向上に直結し、多くの企業や行政機関の意思決定に寄与することが期待される。

研究成果の概要(英文)：Mixed-integer optimization has been attracting attention in recent years as an exact solution for variable selection in regression and discriminant analyses. On the other hand, variable selection based on mixed-integer optimization has the disadvantage that multicollinearity often remains in the set of selected explanatory variables, and the prior knowledge inherent in the data is not utilized. Thus, in this study we combined "mixed-integer optimization" with "removal of multicollinearity" and "structured regularization (model construction using prior information)" to propose constrained variable selection methods that enable high-accuracy parameter estimation. Numerical experiments using synthetic and actual data were conducted to verify the effectiveness of the proposed methods.

研究分野：数理最適化

キーワード：数理最適化 機械学習 アルゴリズム 計算機統計

1. 研究開始当初の背景

回帰分析や判別分析において、候補となる変数の中から有用な説明変数の集合を選び出すことは変数選択(他に特徴選択・部分集合選択・素性選択など)と呼ばれ、統計学の分野では最も基本的かつ重要な問題として知られている。変数選択のための厳密解法として、混合整数最適化が近年注目を集めている。混合整数最適化による変数選択は1970年代から提案されていたが、当時は小規模の問題しか扱うことができなかった。しかし現在では、計算機や求解アルゴリズムの性能が飛躍的に向上し、実用規模の問題に対して混合整数最適化を適用するための環境が整いつつある。一方で混合整数最適化による変数選択には、以下のような欠点があると考えられる:

- 選択された説明変数の中には共線形の強い変数集合が残ってしまう場合が多い。このような問題は多重共線性と呼ばれ、パラメータの推定精度を大きく低下させる。
- 説明変数間の構造情報のような、データに内在する事前知識が活用されていない。このような事前情報を活用したモデル構築は構造的正則化と呼ばれ、機械学習の分野で有効性が主張されている。

以上を考慮して本研究では、「混合整数最適化」に「多重共線性の除去」と「構造的正則化」を組み合わせ、高精度のパラメータ推定が可能な制約付き変数選択手法を確立することを目的とする。

2. 研究の目的

(1) 多重共線性を除去する制約付き変数選択

共線形の強い変数集合を取り除き、パラメータの推定精度を高めるために、多重共線性を除去する制約条件を混合整数最適化モデルに導入する。

(2) 構造的正則化に基づく制約付き変数選択

データに内在する事前知識を活用して有用な説明変数を選択し、パラメータの推定精度を高めるために、構造的正則化の制約条件を混合整数最適化モデルに導入する。

(3) 数値実験による有効性の検証

数値実験を実施し、提案手法の有効性を検証する。

3. 研究の方法

(1) 多重共線性を除去する制約付き変数選択

多重共線性を検出するための指標として「相関係数行列の条件数」と「分散拡大要因(VIF)」を利用し、これらの指標の値を基準値以下とする制約条件の下で、最良の説明変数集合を選択する混合整数最適化モデルを開発し、有効な求解アルゴリズムを提案する。

(2) 構造的正則化に基づく制約付き変数選択

構造的正則化の例として、関連する変数をまとめて選択・削除するグループ正則化や、変数間の優先関係に従って変数を選択する階層的な正則化、隣接する変数のパラメータ推定値を類似させる連結正則化などがある。このような説明変数間の構造情報に基づく制約条件の下で、最良の説明変数集合を選択する混合整数最適化モデルを開発し、有効な求解アルゴリズムを提案する。

(3) 数値実験による有効性の検証

パラメータの真値が既知の人工データを作成し、提案手法を適用してパラメータ推定精度を検証する。また、申請者の研究グループが所有するスキャンパネル(購買履歴)データに提案手法を適用し、既存手法との分析結果の差異や妥当性を考察する。

4. 研究成果

(1) 条件数制約付き変数選択の研究

本研究では、線形回帰モデルから多重共線性を排除するための方法を提案した。この方法では、選択した変数の相関行列の条件数の上限制約の下で、説明変数の最適な部分集合を選択する。最初に条件数の制約を近似するために、混合整数凸2次最適化問題に妥当不等式を繰り返し追加する切除平面法を開発した。また、条件数の制約の下で最良の変数集合を選択するために、混合整数半正定値最適化問題による定式化を考案した。数値実験を実施し、切除平面法は変数選択のための既存の局所探索法よりも、多くの場合に良質の変数集合を選択することを検証した。さらに、候補変数の数が少ない場合には、混合整数半正定値最適化問題を用いて変数選択問題の厳密求解が可能であることを検証した。

(2) VIF制約付き変数選択の研究

本研究では、線形回帰モデルから多重共線性を排除するために、多重共線性の検出に使用される最も一般的な指標である分散拡大要因(VIF)を使用して、説明変数の有用な部分集合を選択する方法を提案した。混合整数最適化による解法の既存研究では、VIFに基づいて多重共線性を排除するための計算上扱いやすい定式化は提案されていない。本研究では、選択された説明変数の

VIFの上限制約の下で、説明変数の最良の部分集合を選択するための混合整数凸2次最適化問題による2種類の定式化を提案した。これらの定式化は、VIFの2種類の等価な定義に基づく。数値実験を実施し、既存の局所探索法や(混合整数最適化に基づく)切除平面法と比較することにより、提案手法の有効性を検証した。

(3) 構造制約付き変数選択の研究

本研究では、スキャンパネルデータに基づいて消費者の店舗選択行動を分析するための店舗選択モデルを対象とする。この店舗選択モデルにより、消費者や製品の属性が消費者の店舗選択だけでなく購入数量に与える影響も評価できる。さらに混合整数最適化を利用して、店舗選択モデルを構築するための説明変数の最適な部分集合を選択する。本研究では、製品種類の階層構造を使用して変数選択の信頼性と計算効率を向上させる階層変数選択のための、2種類の混合整数最適化モデルを考案した。また実際のスキャンパネルデータを用いた数値実験を通じて、混合整数最適化モデルの有効性を評価した。この実験では、消費者がコンビニエンスストア、ドラッグストア、食料品スーパーの中から店舗を選択することを想定している。数値実験を実施し、提案手法が変数選択の一般的な方法であるステップワイズ法とL1正則化法と比較して、いくつかの利点があることを検証した。さらに分析結果から、コンビニエンスストアはギフトカードとごみ処理券を求めて選択され、ドラッグストアはドラッグストアに固有の製品を求めて選択され、食料品スーパーは家族をもつ女性により健康食品を求めて選択されることが分かった。

5. 主な発表論文等

〔雑誌論文〕 計17件（うち査読付論文 14件 / うち国際共著 0件 / うちオープンアクセス 10件）

1. 著者名 高野祐一, 宮代隆平	4. 巻 50
2. 論文標題 混合整数最適化による線形回帰モデルの最良変数選択	5. 発行年 2021年
3. 雑誌名 日本統計学会誌	6. 最初と最後の頁 343-362
掲載論文のDOI (デジタルオブジェクト識別子) 10.11329/jjssj.50.343	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 Ryuta Tamura, Ken Kobayashi, Yuichi Takano, Ryuhei Miyashiro, Kazuhide Nakata, and Tomomi Matsui	4. 巻 73
2. 論文標題 Mixed integer quadratic optimization formulations for eliminating multicollinearity based on variance inflation factor	5. 発行年 2019年
3. 雑誌名 Journal of Global Optimization	6. 最初と最後の頁 431-446
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s10898-018-0713-3	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 Toshiki Sato, Yuichi Takano, and Takanobu Nakahara	4. 巻 13
2. 論文標題 Investigating consumers' store-choice behavior via hierarchical variable selection	5. 発行年 2019年
3. 雑誌名 Advances in Data Analysis and Classification	6. 最初と最後の頁 621-639
掲載論文のDOI (デジタルオブジェクト識別子) 10.1007/s11634-018-0327-0	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 田村隆太, 小林健, 高野祐一, 宮代隆平, 中田和秀, 松井知己	4. 巻 63
2. 論文標題 多重共線性を考慮した回帰式の変数選択問題の定式化	5. 発行年 2018年
3. 雑誌名 オペレーションズ・リサーチ: 経営の科学	6. 最初と最後の頁 128-133
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 オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 Ryuta Tamura, Ken Kobayashi, Yuichi Takano, Ryuhei Miyashiro, Kazuhide Nakata, Tomomi Matsui	4. 巻 60
2. 論文標題 Best subset selection for eliminating multicollinearity	5. 発行年 2017年
3. 雑誌名 Journal of the Operations Research Society of Japan	6. 最初と最後の頁 321-336
掲載論文のDOI (デジタルオブジェクト識別子) 10.15807/jorsj.60.321	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9件 (うち招待講演 4件 / うち国際学会 5件)

1. 発表者名 Yuichi Takano
2. 発表標題 Best subset selection for linear regression models via mixed-integer optimization
3. 学会等名 NUS-Tsukuba Joint-Online-Workshop on "Sustainable Management and Data Sciences" (招待講演) (国際学会)
4. 発表年 2021年

1. 発表者名 Yuichi Takano and Ryuhei Miyashiro
2. 発表標題 Best subset selection via cross-validation criterion
3. 学会等名 International Conference on Nonlinear Analysis and Convex Analysis -- International Conference on Optimization Techniques and Applications (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 高野祐一
2. 発表標題 多重共線性を考慮した最良部分集合選択
3. 学会等名 第29回RAMPシンポジウム (招待講演)
4. 発表年 2017年

1. 発表者名 Ryuta Tamura, Ken Kobayashi, Yuichi Takano, Ryuhei Miyashiro, Kazuhide Nakata, Tomomi Matsui
2. 発表標題 Mixed-integer quadratic optimization formulations for eliminating multicollinearity based on variance inflation factor
3. 学会等名 INFORMS 2017 Annual Meeting (国際学会)
4. 発表年 2017年

1. 発表者名 田村隆太, 小林健, 高野祐一, 宮代隆平, 中田和秀, 松井知己
2. 発表標題 多重共線性を除去するための最良部分集合選択
3. 学会等名 日本オペレーションズ・リサーチ学会 2017年春季研究発表会
4. 発表年 2017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