

令和 6 年 6 月 7 日現在

機関番号：12605

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K04527

研究課題名（和文）多重共線性の評価指標統一に向けた最良特徴選択手法の構築と回帰モデルの統計的解析

研究課題名（英文）Integration of indices for multicollinearity detection

研究代表者

宮代 隆平（Miyashiro, Ryuhei）

東京農工大学・工学（系）研究科（研究院）・准教授

研究者番号：50376860

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）：研究期間に得られた成果として、まず正準相関分析における特徴選択について、整数最適化を用いた特徴選択アルゴリズムを構築した。構築したアルゴリズムは、非凸非線形な整数最適化問題の求解を含み、汎用の最適化ソルバーでは解くのが困難な問題である。この整数最適化問題に対して新しく分枝限定法を実装し、既存のソルバーより100倍程度高速に計算が可能になったことを確認した。また、サポートベクターマシンなどに現れる高次元空間におけるクラス間の重心間距離最大化問題は、指数関数を含む非凸非凹離散最適化問題であり最適化が非常に困難であったが、この問題に対し線形整数最適化としての定式化を構築することに成功した。

研究成果の学術的意義や社会的意義

現象を観察して得られたデータから重回帰分析で回帰モデルを作成する際に、無関係な特徴を削除して必要な特徴だけを回帰モデルに組み入れることを特徴選択と呼ぶ。特徴選択は変数選択とも言われ、古くから統計学における課題であったが、近年のデータサイエンスの流行に伴い改めて重要性が指摘されている。本研究では、特徴選択における二つの重要な問題（正準相関分析における特徴選択問題、高次元空間におけるクラス間の重心間距離最大化問題）に対して、新しい数理モデル化を提案した。これらの問題は非線形性から、従来のソフトウェアでは解くのが困難だったが、本研究の提案手法により高速に最適な特徴選択が行えるようになった。

研究成果の概要（英文）：In this study, we first developed a feature selection algorithm using integer optimization for the feature selection problem in canonical correlation analysis. The feature selection problem contains the process of solving nonconvex and nonlinear integer optimization problems, which are difficult to handle by an off-the-shelf optimization solver. We also implemented a new branch-and-bound method for this integer optimization problem and confirmed that the algorithm is about 100 times faster than existing solvers. Next, we proposed a formulation for the problem of maximizing the distance between centers between two classes in high-dimensional spaces, which appears in support vector machines and other applications. The maximization problem has a nonconvex, nonconcave and nonlinear objective function, and is extremely difficult to optimize for a general purpose solver. For this maximization problem, we developed an integer linear optimization formulation.

研究分野：OR

キーワード：OR 統計学 特徴選択 変数選択 整数最適化

## 様式 C - 19、F - 19 - 1 (共通)

### 1. 研究開始当初の背景

現象を観察して得られたデータから重回帰分析で回帰モデル(回帰式)を作成する際に、無関係な特徴(説明変数)を削除して必要な特徴だけを回帰モデルに組み入れることを特徴選択と呼ぶ。特徴選択は変数選択/素性選択とも言われ、古くから統計学における重要な課題であったが、近年のデータサイエンスの流行に伴い改めて脚光を浴びている。

特徴選択の際に問題となるのが、多重共線性である。多重共線性が大きい回帰モデルは、その原因となっている特徴を取り除くことが必要であるため、多重共線性の判定は重要度が高い手続きである。回帰モデルの多重共線性の大小を判定するには、主に以下の三種類の指標がある。

- ・回帰モデルに含まれる各2つの特徴間の相関係数
- ・回帰モデルに含まれる特徴について、その相関係数行列の条件数
- ・回帰モデルに含まれる特徴について、その相関係数行列から計算される分散拡大要因(VIF)

ただしこれまで、情報工学周辺の分野では条件数が、統計学関連の分野ではVIFが主に用いられており、両指標の関係はほとんど議論されてきていなかった。それにより、回帰モデルによっては条件数では多重共線性があると判断されるがVIFではされないなど、多重共線性の評価指標に関する不整合が起きていた。

### 2. 研究の目的

本研究では当初、前項のような背景から、多重共線性に関する評価指標の統一を目的としていた。この目的は新しい指標の導出という意味だけには限定しておらず、例えば条件数とVIFを融合した多重共線性の評価アルゴリズムの開発なども含んでいた。

しかしながら、特徴選択における多重共線性の問題点について研究を進めていくうちに、非凸非線形な目的関数を持つ特徴選択問題に対する新しいアルゴリズムの構築を目指す方向に研究方針を転換することとした。当初掲げていた研究目的は、主に線形の回帰分析における特徴選択および付随する多重共線性を扱うものであったが、線形回帰分析における特徴選択問題より解くのが困難な、目的関数に非線形性を含むいくつかの特徴選択問題に対して、先行研究より優れている数理モデル化(定式化)の開発に目途がたったためである。

新しい研究目的として、目的関数に非線形性を含むような、従来の手法では最適化が困難な特徴選択問題に対し、より高精度な特徴選択アルゴリズムを開発することを狙いとする。特に、申請者の得意とする、数理最適化の技術を用いた特徴選択アルゴリズムの開発を目指す。数理最適化はモデル化により計算時間が大幅に異なるという特性があり、いかに計算時間が短くなるモデル化を開発できるかが研究の主要なターゲットとなる。

目的関数に非線形性を含むような特徴選択問題は、これまでは主にヒューリスティックなアルゴリズムによって解が求められており、高速に解が得られる反面、その解の質という意味では保証が得られないことがほとんどであった。数理最適化の技術を用いたアルゴリズムの開発に成功した場合には、定めた評価規準の下での最適な特徴量が得られることが期待できるため、新たな研究目的は研究課題として十分に適切と判断した。

### 3. 研究の方法

非線形性を含むような特徴選択問題は、目的関数が取り扱い困難な形をしていることが多く、

従来の「素直な」モデル化ではそもそも数理最適化ソルバーで扱えないケースがほとんどである。このような場合の研究方針として、以下の三点があげられる。

- (1) 厳密な最適化にかえて、高性能なヒューリスティック・アルゴリズムの開発を行う
- (2) 問題の特性を解析し、ソルバーを用いない自前の分枝限定法を実装する
- (3) 数理モデル化を工夫し、非線形性を緩和したもしくは除去した定式化を行う

方法(1)は現実的に有用ではあるが、これまでの先行研究と方向性が同じため、より優れた成果を生み出すのは難しいと判断した。したがって本研究では方法(2)もしくは方法(3)によるアルゴリズムの開発を主に考える。方法(2)については、どのような特徴選択問題を扱うかによって分枝限定法の性質が異なってくるため、一概に「このような分枝限定法を組む」と前もって具体的な計画を立てることは難しい。方法(3)については、やはりどの特徴選択問題を扱うかによりモデル化の方針が異なってはくるものの、基本的には目的関数の非線形性を(問題の性質を変えないようにしながら)緩和または除去したモデル化を行い、解きやすい数理最適化問題の形としての定式化を目指すこととなる。

#### 4. 研究成果

研究期間における主な成果は以下の二点である。

##### ・正準相関分析におけるスパース特徴選択に対する分枝限定法の開発

本課題に対しては、まず正準相関分析におけるスパース特徴選択問題を整数最適化問題として定式化した。この整数最適化問題は、一般のソルバーでは解きにくい形をしているため、一般化固有値問題との関連を導出し、それを繰り返し解くことにより効率的な上下界を算出した。また、アルゴリズムの最適性についても証明を行った。計算機実験の結果、従来の数理最適化ソルバーを用いた場合より約 100 倍程度高速に解が得られ、また既存手法として頻繁に用いられている L1 正則化法によるヒューリスティクスより質の良い解が得られることを確認した。

##### ・非線形カーネル SVM における特徴選択に対する混合整数線形最適化モデルの開発

非線形カーネル SVM によるクラス分類では、分類の前処理としてガウスクーネル関数に基づく高次元特徴空間における 2 クラス間距離(DBTC)を用いた特徴選択を行う。DBTC を最大化する特徴選択問題を考えると、目的関数として最大化される DBTC は非線形・非凸・非凹である。このような非線形関数を線形化することは一般には非常に困難であるが、本課題では DBTC を最大化する特徴選択問題に関して、混合整数線形最適化モデル(定式化)を提案することに成功した。提案モデルは通常の混合整数線形最適化問題であるため、既存の最適化ソルバーを用いて解くことができる。計算機実験の結果として、提案手法は既存手法である線形 SVM に基づく特徴選択や RFE と呼ばれるヒューリスティクスよりもモデルの予測性能で同等か上回った。特に、データのインスタンス数が比較的小さな場合は優位性が高かった。

上記の二点の特徴選択問題は、いずれも現実的に有用なだけでなく、これまでは経験的な最適化による特徴選択しか提案されていなかった問題である。本研究ではこれらに対して厳密解を与える手法を開発できたという点に成果の意義がある。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2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 R. Tamura, Y. Takano, R. Miyashiro	4. 巻 E107-A
2. 論文標題 Mixed-integer linear optimization formulations for feature subset selection in kernel SVM classification	5. 発行年 2024年
3. 雑誌名 IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences	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） 10.1587/transfun.2023EAP1043	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 A. Watanabe, R. Tamura, Y. Takano, R. Miyashiro	4. 巻 217
2. 論文標題 Branch-and-bound algorithm for optimal sparse canonical correlation analysis	5. 発行年 2023年
3. 雑誌名 Expert Systems with Applications	6. 最初と最後の頁 119530
掲載論文のDOI（デジタルオブジェクト識別子） 10.1616/j.eswa.2023.119530	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3件（うち招待講演 0件/うち国際学会 1件）

1. 発表者名 田村隆太, 高野祐一, 宮代隆平
2. 発表標題 混合整数最適化によるカーネルSVMの変数選択
3. 学会等名 数理最適化：モデル, 理論, アルゴリズム
4. 発表年 2022年

1. 発表者名 田村隆太, 高野祐一, 宮代隆平
2. 発表標題 混合整数最適化によるカーネルSVMの変数選択
3. 学会等名 日本オペレーションズ・リサーチ学会2023年春季研究発表会
4. 発表年 2023年

1. 発表者名 Y. Takano, A. Watanabe, Ryuta Tamura, R. Miyashiro
2. 発表標題 Branch-and-bound algorithm for optimal sparse canonical correlation analysis
3. 学会等名 INFORMS Annual Meeting 2021 (国際学会)
4. 発表年 2021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