

Branch-and-Bound Algorithms for Variable Selection and Shortest Vector Problem

木村, 圭児

<https://hdl.handle.net/2324/2236041>

出版情報 : Kyushu University, 2018, 博士 (機能数理学) , 課程博士
バージョン :
権利関係 :

氏 名 : 木村 圭児

論 文 名 : Branch-and-Bound Algorithms for Variable Selection and Shortest
Vector Problem
(変数選択と最短ベクトル問題のための分枝限定法)

区 分 : 甲

論 文 内 容 の 要 旨

【概要】近年、混合整数非線形計画問題を解くソフトウェアは著しく進歩している。特に、最先端の最適化ソフトウェアは様々な最適化問題を扱えることができ、研究者や実務家に幅広く使用される。しかしながら、最先端の最適化ソフトウェアを使ったとしても、混合整数非線形計画問題の計算コストは、混合整数線形計画問題の場合より高いのが現状である。計算コストが高い理由の一つとして、汎用的な最適化ソフトウェアの多くは、問題特有の性質や関数の構造を必ずしも有効活用できていないことが考えられる。一方で、特定の問題のために設計されたソフトウェアは、汎用的な最先端ソフトウェアよりも良い性能を発揮することがある。本研究では、(i) **変数選択**と(ii) **最短ベクトル問題**と呼ばれる二つの問題に対して、求解アルゴリズムとして効率的な分枝限定法をそれぞれ提案する。分枝限定法は、整数性を持つ最適化問題を解くためのアルゴリズムであり、以下のような様々な構成要素から形成される: 緩和, 分枝の仕方, ヒューリスティクスや前処理など。提案した分枝限定法は、計算時間を改善するために、それぞれの構成要素に問題が持つ性質や関数の構造を有効活用する。先に述べた問題に関する数値実験を示し、提案した分枝限定法は、最先端の最適化ソフトウェアと比べ、どの程度高速に解けるか実験的に示す。

(i) 変数選択

【背景と研究内容】 統計学のモデル推定において、情報量規準などの評価関数を用いて変数選択を行うことがある。この問題に対して、ステップワイズ法と呼ばれる手法が広く利用されている。ステップワイズ法は、計算は高速であるが、評価関数値が最も良い値であるモデルを提供するとは限らない。一方、モデルの候補を全て評価するのは現実的ではない。本研究は、 ℓ_0 -正則化項付きの評価関数が与えられたとき、変数選択を混合整数非線形計画問題として定式化し、その問題を効率良く解く分枝限定法を提案する。その結果、評価関数値が最良であるモデルを求めることができる。

【定式化と応用先】 一般に、変数選択に用いられる評価関数は以下の二つの項で形成される: (1) データとモデルの適合度を表す項, (2) モデルに用いる変数を抑えるための罰則項。これらの項を用いた最小化問題として定式化する。本研究において、定式化された問題は、線形回帰とロジスティック回帰における AIC 最小化 (AIC: Akaike Information Criterion) に適用可能であることを示す。

【効率良く解くための工夫】 定式化された問題に対して、効率良く解く分枝限定法を提案する。分枝限定法の構成要素に、問題が持つ性質や関数の構造を有効活用する。開発した構成要素は、SCIP (Solving Constraint Integer Programs) を用いて実装する。SCIP は、分枝限定法を実現するため

のフレームワークを提供していて、分枝限定法の細かな制御を行うプログラムを組み込むことができる。効率良く解くための工夫を一部紹介する。

- 緩和問題：部分問題の最小値の下界値を効率良く計算するための緩和問題を提案
- 分枝変数の選び方：変数選択の解の傾向を基にした **Most frequent branching** を提案
- ヒューリスティクス：部分問題に対するステップワイズ法ベースのヒューリスティクスを実装

【数値実験】 ここでは、ロジスティック回帰における AIC 最小化に関する数値実験を紹介する。下図は、最先端の最適化ソフトウェアである **CPLEX** を用いる先行研究[1]との比較である。「>5000」は、時間制限 5000 秒内に解くことができなかったことを意味する。提案した分枝限定法は、先行研究[1]の手法に比べ、高速に解くことができていることが分かる。

[1] T. Sato, et al.: Feature subset selection for logistic regression via mixed integer optimization. Computational Optimization and Applications, 64 (2016), 865-880.

データ名	データ数	変数の候補	計算時間(秒)	
			提案手法	先行研究[1]
breast-P	194	34	25.8	112.4
biodeg	1055	42	221.5	>5000
spectf	267	45	432.5	515.7
musk	6598	166	>5000	>5000

(ii)最短ベクトル問題

【背景と研究内容】 近年、次世代の暗号の一つとして格子暗号が注目されている。格子暗号の安全性を支えている最短ベクトル問題 (Shortest Vector Problem: SVP) は、与えられた整数格子に対して、原点に最も近い格子点を求める問題であり、NP 困難であることが知られている。SVP に対する既存手法として、格子縮約を繰り返すアルゴリズムなど数多く提案されている。本研究では、SVP に対して効率的な分枝限定法を提案し、どの程度の規模まで解くことができるか実験的に検証する。

【効率良く解くための工夫】 格子暗号に関する諸性質や既存手法は系統立てて研究されている。一方、数理最適化の分野において、分枝限定法に対する高速化のテクニックは近年盛んに研究されている。これらを結び付けることで、アルゴリズムの高速化が期待できる。本研究において提案した効率良く解くための分枝限定法の工夫を一部紹介する。

- 格子縮約+前処理：SVP の既存手法である格子縮約によって得られる解を用いて、定式化された問題の変数の上下制限約を求めて、探索領域を小さくする。
- グラムシュミットの直交化+緩和：グラムシュミットの直交化を用いて、分枝限定法における部分問題の最小値の下界値を求める。

【数値実験】 提案した分枝限定法と最先端の最適化ソフトウェアである **CPLEX** の性能を比較するために、本研究で SVP に対する整数二次計画法を提案し、以下のアプローチを比較する。

- BaB: 提案した分枝限定法であり、C++により実装
- BIQO: SVP に対して 01 変数を用いて整数二次計画問題に定式化し、CPLEX を用いて解く
- IQP: SVP をいくつかの整数二次計画問題に分割し、CPLEX を用いて解く

下図の「n」は SVP の格子の次元を表し、「seed」はその次元の問題に対する乱数生成のためのシードである。時間制限 86400 秒 (= 1 日) 内に解けなかった場合は、出力されていない。提案した分枝限定法は、最先端の最適化ソフトウェアである CPLEX を用いる二つのアプローチよりも、高

速に解けていることが分かる.

