Vol.2021-QS-4 No.12
2021/10/15

ブラックボックス最適化による行列の非可逆圧縮

門脇正史[a]　安倍満[b]

概要：ベイズ最適化に代表されるブラックボックス最適化は、限られたデータから背後の関数をモデル化し、その関数の最適値を与える入力を求める手法である。近年、入力が二値や整数である組合せ最適化問題のブラックボックス最適化に関してイジングソルバを用いた研究が複数提案されている。我々は、この手法を整数基底分解法と呼ばれる行列の非可逆圧縮に適用した。整数基底分解法は機械学習の学習済みモデルをエッジコンピューティングで実行するための汎用的な技術である。我々の提案はある種の混合整数非線形計画法をイジングソルバで解くことを可能とする。

キーワード： ブラックボックス最適化, 整数基底分解法, 混合整数非線形計画法, イジングソルバ, 量子アニーリング

Lossy compression of matrices by black-box optimization

TADASHI KADOWAKI[a]　MITSURU AMBAI[b]

Abstract: Black-box optimization, such as Bayesian optimization, is a method to find the input values to minimize or maximize a black-box function by a limited amount of data acquired sequentially. Recently, several studies using Ising solvers have been proposed for black-box optimization of combinatorial optimization problems with binary or integer inputs. We apply this method to the lossy compression of matrices, called integer decomposition. Integer decomposition method is a general-purpose technique for running trained models of machine learning on edge computing. Our proposal makes it possible to solve certain types of mixed integer non-linear programming with Ising solvers.

Keywords: black-box optimization, integer decomposition, mixed integer non-linear programming, MINLP, Ising solver, quantum annealing

1. はじめに

ある評価値を最大化もしくは最小化することは、研究上の興味だけでなく産業や社会においても重要な問題である。このような問題は最適化問題と呼ばれ、最適化対象である説明変数が連続値、離散値、その両方の問題がある。評価値は目的関数(目的変数)であり、それが微分可能と不可能の違いで適用する手法に大きな違いがある。特に、離散変数の場合を組み合わせ最適化問題と呼ぶ。目的関数が陽に与えられる場合には定式化ができ、その中でも目的関数や制約条件が連続変数の線形結合の場合に線形計画問題といい 1940 年代から研究されている。

目的関数が陽に与えられていない場合には、実験やシミュレーション等によって取得したデータを基に最適化を行う。このような場合には実験計画法や、ブラックボックス最適化が用いられる。前者はあらかじめデータ点を決めて測定するのに対し、後者はデータから目的関数を推定し、推定精度の向上(探索)と最適解の取得(活用)のためにデータ点を逐次追加する。さらに、説明変数が離散変数であるブラックボックス最適化では、推定した目的関数を用いて次のデータ点を決定する繰り返しにおいて、何度も組み合わせ最適化問題を解く必要がある。

ブラックボックス関数の最適化問題と、整数最適化問題の組み合わせは、非常に計算コストの高い問題であるが、素材設計や学習モデル設計など、整数変数やカテゴリカル変数をもつ目的関数の最適化は多く存在する。本報告では、整数基底分解法[1]と呼ぶ機械学習モデルを圧縮する手法が、混合整数非線形計画問題としての定式化から離散変数のブラックボックス最適化問題に変形できることを示し、近年開発された最適化手法を適用することで従来手法よりも良い解が得られることを示す。また、シミュレーテッドアニーリング(SA)[2]と量子アニーリング(QA)[3]を用いた際の性能の比較を行う。

2. 整数基底分解法

整数基底分解法は、機械学習によって生成される重み行列を圧縮するため開発された手法である。サポートベクトルマシンやニューラルネットワークなどにより生成される重み行列は、学習のタスクの難易度に依存して規模が大きくなる。この重み行列を用いた学習済みモデルはデータサイズ大きく演算回数が多いため、エッジコンピューティン

a （株）デンソー
DENSO CORPORATION
b （株）デンソーアイティーラボラトリ
DENSO IT Laboratory

グで利用する際のハードルとなる。整数基底分解法を用いて重み行列を整数基底とその係数に分解することで、整数基底はメモリ効率の高いビットデータとなり、係数は行列サイズが縮小された浮動小数データとなる。非可逆圧縮であり、圧縮率に依存して学習済みモデルの再現性は低下するため、目的に応じて圧縮率は設定される。

重み行列の分解は、以下の最適化問題となる。

$$\underset{\substack{M\in\{-1,1\}^{NK} \\ C\in\mathbb{R}^{KD}}}{\operatorname{argmin}} |W-MC|^2 \tag{1}$$

ここで、W は N×D の重み行列、M は N×K の二値行列、C は K×D の係数行列である。メモリ効率の高い M を無視すると、圧縮率は K/N となる。この問題は、整数(二値)と実数を同時に最適化する必要があり、また非線形な L2 ノルムを計算するため、混合整数非線形計画問題である。

安倍と佐藤は、この最適化問題を図 1 のように 1 次近似とその残差として逐次的に分解した行列を構成することを提案している。この手法は高速に実行できる半面、局所解の探索のため、より良い解や厳密解の探索を省略している。我々は、式 1 の混合整数非線形計画問題を非線形整数計画問題に変換し、その評価関数の最適解をブラックボックス最適化により探索することで、従来の手法よりも良い解が得られることを示す。なお、混合整数非線形計画問題と非線形整数計画問題はどちらも NP 困難な問題である。

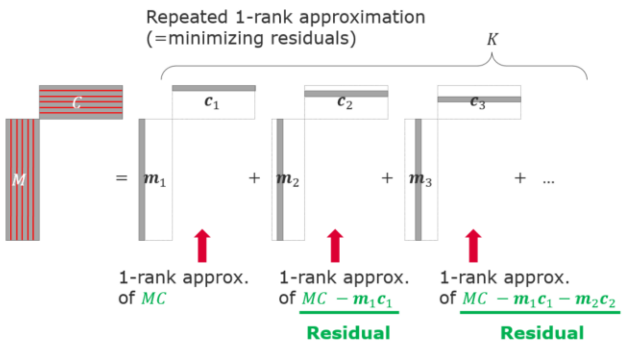

図 1 整数基底分解法の近似解法

3. ブラックボックス最適化

ブラックボックス最適化として最も成功している手法に、ベイズ最適化がある。ベイズ最適化は、連続変数の最適化手法であり、与えられたデータの入出力関係をガウス過程によりモデル化する。ここで得られるモデルは確率分布のため、入力に対して分布が出力として得られる。この分布から、獲得関数を定義することができる。獲得関数は次の測定データ点を決定するための関数であり、この関数が最大となる点を次の候補点として選ぶ。獲得関数の定義は複数あるが、例えば Upper Confidence Bound(UCB)の場合、期待値と分散の線形結合である。期待値が高く分散が大きな点は、良い解候補であるだけでなく、データが少なく不確実性が高いため、次のデータ点として望ましい条件となる。この処理を繰り返すことで、過去のデータを利用した効率的な探索ができる。UCB は、決定的アルゴリズムだが、確率的アルゴリズムを用いることもできる。例えば、Thompson Sampling は確率的に候補を採択する。

ベイズ最適化は逐次的に候補点(説明変数)を決定と測定データ(目的変数)の取得を行うことでデータは蓄積され、そこから生成されるモデルがより精緻になる。このデータ取得の戦略を決めるのが獲得関数である。一般に探索初期においては解空間を広く探索し、探索後期にはデータの蓄積に合わせてより良い解がありそうな領域を重点的に探索するように設計されている。この特長により、グリッドサーチよりも少ない測定回数でよりよい解が得られる。また、実験計画法と異なり一部のデータに欠損があった場合にも最適化を行える。

連続変数に対してはベイズ最適化を用いることができるが、二値変数など離散変数に対しては、そのまま用いることができない。連続変数では、獲得関数の微分から最適な点を求めることができるが、離散変数では組合せ最適化問題となり一般的に不可能(NP 困難)なためである。近年、この組合せ最適化をイジングソルバで解く手法が提案されている。Factorization Machine with Quantum Annealing (FMQA)[4] と、Bayesian Optimization of Combinatorial Structures (BOCS)[5]である。FMQA は決定論アルゴリズムであり、BOCS は確率的アルゴリズムである。

FMQA はデータの入出力関係を Factorization Machine (FM)でモデル化する。FM は離散値の入力を学習することを想定したモデルであり、入力に対して任意の多項式でモデル化できるが、通常は二次形式で表現される。二次形式で表現されたモデルは、Quadratic Unconstrained Binary Optimization (QUBO)と呼び、そのままイジングソルバで最適解あるいは準最適解を求めることができる。したがってこの解を次のデータ点にすることで、ブラックボックス最適化を実施できる。

BOCS も任意の多項式でモデル化することができるが、ここでも通常は二次形式が用いられる。データから QUBO の確率分布モデルを構築する。Thompson Sampling では次のデータ点を決定する際に確率分布を用いたが、BOCS では、QUBO の確率分布からイジングソルバで最適化する QUBO のサンプリングを行う。探索初期で用いる QUBO はほとんどランダムにサンプリングされるため、得られる候補点は解空間を広く分布し、探索後期になると QUBO の確率分布が収束し、最適解の周辺の探索を可能とする。QUBO を用いたブラックボックス最適化は、コスト関数の構造が複雑であっても解の周辺においては二次近似が有効であることが前提であり、初期探索から解周辺にたどり着くための獲得関数による探索戦略が重要である。

4. 提案手法

整数基底分解法では、整数基底の最適化と、係数の最適化を交互に行っている。係数の決定は線形回帰問題であり、以下の式で計算できる。

$$C = (M^T M)^{-1} M^T W \qquad (2)$$

係数行列 C を陽に表すことができるので、式 1 の係数行列 C と整数基底行列 M に関する最適化問題は以下のように M に関する非線形整数計画法に帰着できる

$$\underset{\substack{M \in \{-1,1\}^{NK} \\ C \in \mathbb{R}^{KD}}}{\operatorname{argmin}} |W - MC|^2 = \underset{M \in \{-1,1\}^{NK}}{\operatorname{argmin}} f(M) \qquad (3)$$

ここで、目的関数は

$$f(M) = |W - M(M^T M)^{-1} M^T W|^2$$

である。目的関数は陽に書くことができるが、二次や有限次の多項式で表すことができないため、このままイジングソルバで最適化することはできない。

ブラックボックス最適化を利用することで、この目的関数そのものを直接用いず、入出力データの取得を繰り返しながら入出力関係を QUBO で近似し、その QUBO を解くことで最適解を求めることができる。つまり、QUBO 表現ができない目的関数をブラックボックス関数として扱うことで最適化を行う。

ブラックボックス最適化には BOCS と FMQA を用い、いくつかの条件で実験を行った。BOCS では Baptista と Poloczek が用いた馬蹄事前分布(以後 vBOCS)に加え、ガウス事前分布(以後 gBOCS)による計算も行った。FMQA では、ハイパーパラメータを、論文で提案している k=8(以後 FMQA08)と、QUBO として十分な表現力のある k=12(以後 FMQA12)で計算した。

5. 結果

図 2 に 8×100 の重み行列を、8×3 の(二値)基底行列と 3×100 の係数行列に分解する例を示す。重み行列は、畳み込みニューラルネットワーク VGG16 の学習済みデータから抽出したデータを使用した。N=8×3 の二値変数の最適化問題に対して、初期データは N=24、繰返し回数を $2N^2$=1152 とし、ランダムサーチは 100 回、その他は 25 回の実験を行い、各実験でそれまでに得られた最もコストの低い解の平均と標準誤差を求めた。イジングソルバはシミュレーテッドアニーリングを用いた。図中の赤い点線は既存手法、灰色の破線と点線は厳密な最適解と、最適解の次に低いコストの解である。

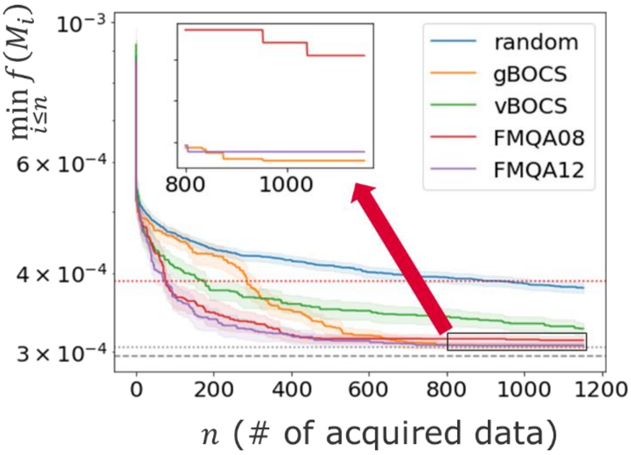

図 2 各アルゴリズムのデータ取得数と解精度

最も早く従来手法よりも良い解を出すのは FMQA12 であり、FMQA08、vBOCS、gBOCS の順で、約 50 から 300 回の繰り返し回数で到達する。一方でランダムサンプリングは約 1000 回を必要とする。終了時に最も低いコストの解を得たのは gBOCS であり、FMQA12、FMQA08 と続く。

この実験を 10 個の重み行列で行い、解の精度と実行時間を表 1 にまとめた。解の精度に関しては、厳密解の出現回数と、終了時のコストが全手法中で最も低くなった回数、実行時間は Intel Xeon processor (Cascade Lake)でのプロセス時間(秒)用いた。表より、解精度に関しては、gBOCS 及び FMQA08 の性能が良く、実行時間に関しては、ランダムサーチと gBOCS が良いことが分かる。

表 1 各アルゴリズ解精度と実行時間

	Random	gBOCS	vBOCS	FMQA08	FMQA12
厳密解	3	62	6	58	53
コスト	0	7	0	2	1
実行時間	0.35	51	24394	3719	3714

次に、gBOCS の QUBO を解くイジングソルバとして、シミュレーテッドアニーリングと量子アニーリングの比較を行った。図 3 は、図 2 と同じ問題を SA(図中 gBOCS)と QA(同 gBOCSqa)で比較した結果である。用いたソルバの違いによる結果の差異は小さいことが分かる。これは、Sherrington-Kirkpatrick モデルの最適化を扱った先行研究[6]が示した SA と QA で差は認められなかった結果と整合する。

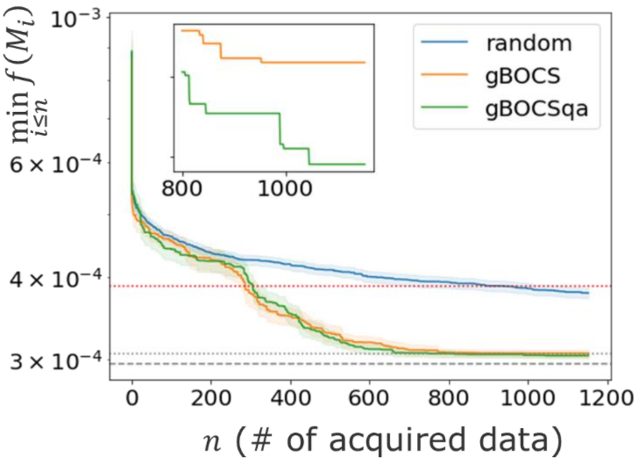

図 3 各ソルバでの解精度(N=8×3)

また、図 4 では問題サイズを N=16×4 にした例を示す。図 3 と同様に、ソルバの違いによる結果の差異は小さい。

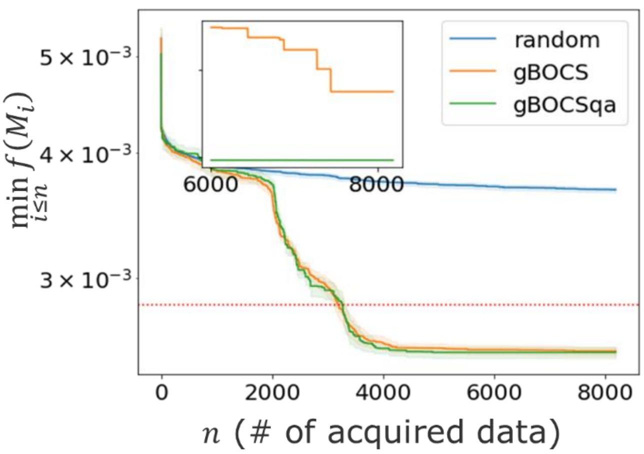

図 4 各ソルバでの解精度(N=16×4)

6. まとめ

ニューラルネットワークなどの大規模な重み行列をエッジコンピューティングで利用するために圧縮する技術である整数基底分解法をブラックボックス最適化により実行する手法を考案した。整数基底分解法は混合整数非線形計画問題を解くため、NP 困難問題である。我々は、この問題を非線形整数計画問題に変換することで、ブラックボックス最適化問題として定式化した。離散変数のブラックボックス最適化はブラックボックス関数を QUBO に近似してイジングソルバで解く手法(BOCS や FMQA)が提案されおり、これらの手法を用いて性能の評価を行った。

その結果、整数基底分解法においては、ガウス事前分布を用いた BOCS が実行時間と解精度の点で他の手法よりも優れていることが分かった。また、用いるイジングソルバは、SA と QA で解精度に大きな違いはないことが分かった。

BOCS と FMQA は確率的と決定的なアルゴリズムであり、問題に応じて向き不向きがあると考えられる。今回明らかにした解精度や実行時間の特徴は、新しいアルゴリズムを考案するヒントになるかもしれない。ブラックボックス関数の組合せ最適化手法にイジングソルバを用いる研究はまだ少なく、また、これまで検証されてきた問題は全て小規模な問題であり、より大きな問題でソルバの解精度がブラックボックス最適化の解精度にどの程度影響を与えるのか、SA と QA を用いたときに性能差があるのか、QA ソルバの特徴に合わせてスパースな QUBO で近似をした場合の性能はどうか、など取り組むべき課題は多い。また、馬蹄分布のような事前分布は計算に時間がかかり、QUBO の最適化よりもデータからのモデル化に時間がかかっている。量子アルゴリズムの利用も含めて、大規模化に耐えるモデル化手法の検討も必要である。

参考文献

[1] M. Ambai and I. Sato (2014). Computer Vision – ECCV 2014, 267–281.
[2] D. Sherrington and S. Kirkpatrick (1975). Physical Review Letters, 35(26), 1792–1796.
[3] T. Kadowaki and H. Nishimori (1998). Physical Review E, 58(5), 5355–5363.
[4] R. Baptista and M. Poloczek, in Proceedings of the 35th International Conference on Machine Learning, ed. J. Dy and A. Krause (2018) Proceedings of Machine Learning Research, Vol. 80, p. 462.
[5] K. Kitai, J. Guo, S. Ju, S. Tanaka, K. Tsuda, J. Shiomi, and R. Tamura (2020). Physical Review Research, 2(1), 013319.
[6] A. S. Koshikawa, M. Ohzeki, T. Kadowaki, and K. Tanaka, (2021). Journal of the Physical Society of Japan, 90(6), 064001.