



令和2年8月17日

GPUの計算能力を最大限活用する組合せ最適化問題の新解法 ～1兆探索/秒を超えるアダプティブ・バルク・サーチ～

【本研究成果のポイント】

- ・ 二次無制約二値最適化（QUBO）問題に対し、複数のGPUを用いてコストが最小となる最適解を探索する新しい探索フレームワーク「アダプティブ・バルク・サーチ」を開発しました。
- ・ GPUのアーキテクチャの特性を活かし計算性能を最大限に引き出す解探索手法により、最新のNVIDIA GPUを4基搭載した計算サーバーで1秒間に1兆を超える解を探索する計算速度を達成しました。
- ・ 探索手法を適応的に変化させることが可能で、幅広い問題に有効です。
- ・ より大規模な計算機システムやスーパーコンピュータを用いることで台数に比例した計算速度のスピードアップが可能です。

【概要】

広島大学大学院先進理工系科学研究科の中野浩嗣教授らの研究チームは、株式会社NTT データと共同で、組合せ最適化問題の解を高速に探索する新しい計算方式「アダプティブ・バルク・サーチ」の開発に成功しました。その計算方式は、二次非制約二値最適化（QUBO）問題の解を複数のGPU（グラフィクス向けプロセッサ）を用いて効率よく並列に探索します。

巡回セールスマン問題など多くの組合せ最適化問題は、QUBO問題として記述することができ、幅広い応用があります。そのため多くの大学や企業が専用集積回路や量子効果に基づいた量子アニーリングマシンなどを用いてQUBO問題を解くシステムや手法の開発を競っています。組合せ最適化問題を解く従来手法のシミュレーテッド・アニーリングなどでは、同じ解を何度も探索するなど無駄が多く、良い解が得られない場合があります。また、手法によってうまく解ける問題と解けない問題が存在します。そこで本研究では、重複がないように大量の解を並列に探索する新しいフレームワークを開発しました。さらに、探索アルゴリズムを問題に応じて適応的に変化させることが可能で、幅広い問題を解くことができます。

本研究成果の詳細は、広島大学大学院先進理工系科学研究科の安戸僚汰特任助教が8月17日から開催される国際会議 International Conference on Parallel Processing (ICPP)にて発表します。

【背景】

社会における問題の多くは、制約の下で多くの選択肢の中から最も良い選択の組み合わせを求める組合せ最適化問題として定式化できます。たとえば、訪問したい都市とその移動コストが与えられたときに「どの順序で回るのが一番低コストで済むか」という巡回セールスマン問題が典型的な例です。都市数が増えるにつれて順序の総数は飛躍的に増加し、5都市であれば120通りですが、20都市で200京通りに達するため解くことが難しい問題です。近年、膨大なデータの処理やAI・機械学習においても組合せ最適化問題を解く場面が増え、ますます重要な課題となっています。

近年注目されているのがイジングモデルです。量子アニーリングマシンを用いるとイジングモデルで記述された組合せ最適化問題を高速に解くことが理論上可能なた

め、そのようなマシンの開発・実用化が競われています（図1）。このイジングモデルは、従来のコンピュータで扱いやすい 0/1 の二値変数を用いて、二次無制約二値最適化（Quadratic Unconstrained Binary Optimization、QUBO）に等価な変換が可能です。

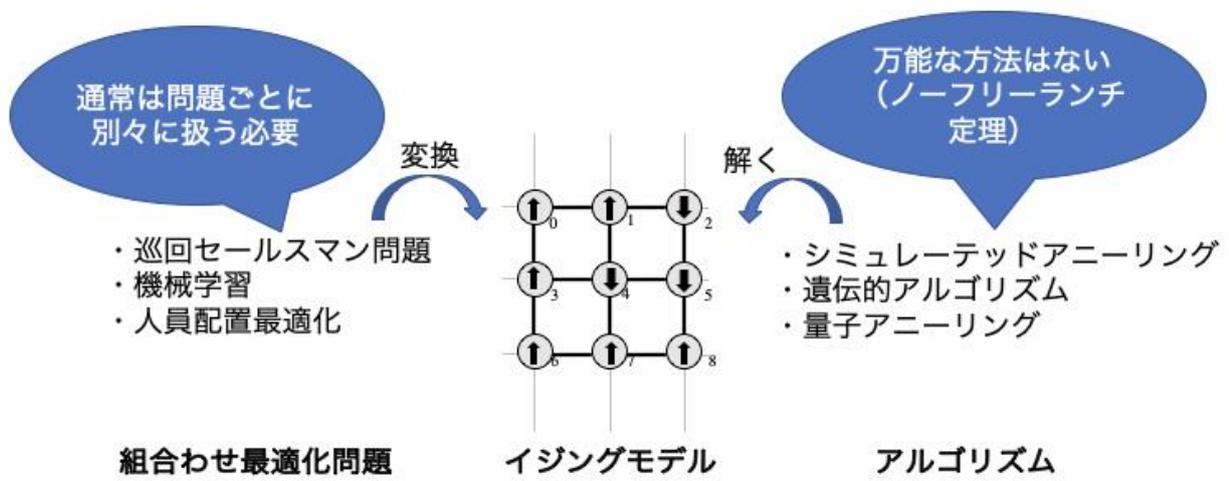


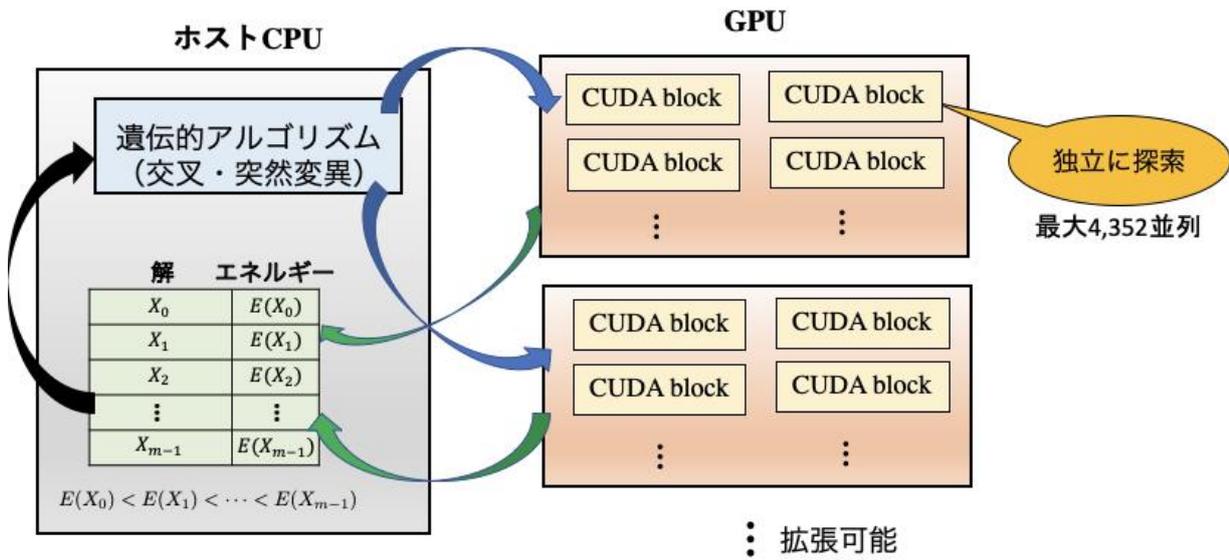
図 1: イジングモデルへは様々な最適化問題から変換することができる。しかし、その効率的な解き方は問題によって異なる。

量子効果に基づいたマシンではなく、専用ハードウェアを用いて QUBO 問題を解くシステムが一定の成功を収めており、量子効果をシミュレートするものや、シミュレーテッド・アニーリング、遺伝的アルゴリズムなどを用いた手法が提案されています。

それでは、どの手法が最も良いのでしょうか。ノーフリーランチ定理が示唆するように、万能な解法は存在せず、問題の種類によって適したアルゴリズムは異なります。つまり、イジングモデルや QUBO で統一的に表したとしても、元の問題の性質によって使うべきアルゴリズムが違うということを意味します。そこで本研究では、複数のアルゴリズムを用いて大量の解を GPU で並列に探索するという新しいフレームワークを提案します。

【研究成果の内容】

アダプティブ・バルク・サーチのコンセプトは、ホスト CPU による遺伝的アルゴリズムと GPU による近傍探索をシームレスにつなぎ、膨大な数のコアを持つ GPU による並列計算でなるべく重複がないように大量の解を探索するというものです（下図）。



QUBO 問題は、 n 個のビット変数の全組み合わせからコストが最小となるものを見つける問題です。QUBO 問題を解く場合、最適候補のコストを繰り返し計算する必要があります。通常このコストの計算量は、 n の二乗に比例して増加していくため時間がかかります。本研究では、最適候補のコストを求めるのに必要な計算量が一つあたり $O(1)$ の並列探索法を GPU に実装しました。

この並列探索法はシミュレーテッド・アニーリングをベースとしてなるべく重複する探索を行わないように、かつ GPU で探索性能を最大化するように設計されています。実験では、最大カット問題、巡回セールスマン問題、ランダム問題で 1 秒以内に最適解を得ることができ、1 秒間に 1 兆を超える解の探索を達成しました。現状では、32、768 変数の QUBO 問題まで扱うことができます。

【今後の展開】

今回は GPU を 4 基搭載した計算機システムに実装しましたが、より大規模なシステムや、将来のより高性能な GPU を使うことで性能をさらに上げることが可能です。また、32、768 変数の問題よりも大規模な問題を扱えるように拡張中です。

今後、本手法は流通、金融、製造分野の重要な問題や、AI・機械学習の高性能化に適用していきます。

【用語説明】

● GPU

Graphics Processing Unit の略で、グラフィック処理のための集積回路です。計算処理能力の高さから、グラフィック処理以外の様々な処理を高速化することができます。多くのスーパーコンピュータに搭載されています。

● QUBO 問題

n ビットの変数をもつ二次式の値が最小になるように各変数に 0 と 1 の割り当てを求める組合せ最適化問題。

● シミュレーテッド・アニーリング

金属の焼きなましにヒントを得た組合せ最適化問題の解法。最適解の探索範囲を温度に依存して決定することで最適解を見つける確率が高くなります。

● 遺伝的アルゴリズム

生物の進化にヒントを得た最適解探索アルゴリズム。解を遺伝子とみなして、交叉と突然変異を繰り返しながら適応度が高い解を残していくものです。

【発表情報】

- ・ 国際会議名：International Conference on Parallel Processing (ICPP) 2020
- ・ 論文タイトル：Adaptive Bulk Search: Solving Quadratic Unconstrained Binary Optimization Problems on Multiple GPUs
- ・ 著者：Ryota Yasudo、Koji Nakano、Yasuaki Ito、Masaru Tatekawa、Ryota Katsuki、Takashi Yazane、Yoko Inaba
- ・ DOI 番号：10.1145/3404397.3404423

【お問い合わせ先】

大学院先進理工系科学研究科 中野 浩嗣 Tel：082-424-5363 FAX：082-424-5363 E-mail：nakano@hiroshima-u.ac.jp

発信枚数：A4版 3枚（本票含む）