

TSUBAME 共同利用 令和 2 年度 学術利用 成果報告書

利用課題名 二次制約無し二値最適化の古典計算機による高速解法

英文: High-performance solver for quadratic unconstrained binary optimization problems with classical computers

利用課題責任者

中野 浩嗣

所属

広島大学大学院先進理工系科学研究科

URL: <https://cs.iss-j.org/>

邦文抄録(300 字程度)

あらゆる組合せ最適化問題は二次制約無し二値最適化問題(QUBO)に帰着することができる。そこで近年、FPGA、GPU、専用プロセッサ等を用いて QUBO を解く研究が盛んに行われている。アダプティブバルクサーチ(ABS)は複数 GPU を用いて並列に QUBO の解を探索する新手法である。本プロジェクトではたくさんの GPU を搭載したスーパーコンピュータに向けて、ABS のスケーラブルな実装を開発した。TSUBAME3.0 における実験の結果、GPU の数を増やすごとに QUBO のより最適解に近い解を探索することができ、探索速度が線形に向上することを示した。

英文抄録(100 words 程度)

A wide range of combinatorial optimization problems can be reduced to the quadratic unconstrained binary optimization (QUBO) problem. In recent years, researchers have proposed to solve QUBO on FPGAs, GPUs, and special-purpose processors. The adaptive bulk search (ABS) is a framework for solving QUBO in parallel on multiple GPUs. In this study, we propose a scalable implementation of the ABS. Our results on TSUBAME3.0 suggest that the solution quality improves as the number of GPU increases, and show linear improvement of the search rate.

Keywords: parallel computing, GPGPU, quantum annealing, Ising model, combinatorial optimization

背景と目的

二次制約無し二値最適化 (Quadratic Unconstrained Binary Optimization, QUBO) 問題のソルバは、あらゆる NP 困難の組合せ最適化問題を解くための汎用ソルバとして注目されている。特に近年研究が進んでいる量子アニーリングは QUBO 問題を量子現象によって解くことができるとされている。

しかし、量子アニーリングではハードウェア上の制約によって、扱える QUBO 問題に制限がある。QUBO 問題は二値変数(ビット)と、ビット間の相互作用によって定式化されるが、このビット数と相互作用の数に限りがある。最新の実用量子アニーリングマシン D-Wave Advantage では、ビット数が約 5000、相互作用はビットあたり 15 程度しか扱うことができない。また、必ずしも最適解が求められるわけではない。

このような量子アニーリングの限界から、従来の古

典計算機を活用して QUBO 問題を解く研究が行われている。FPGA、専用プロセッサ、GPU などを使って、量子現象をシミュレートする研究や、あるいは量子現象をまったく用いずに並列計算によって QUBO 問題の近似解を高速に求める研究などがある。中でも我々の開発したアダプティブ・バルク・サーチでは、CPU で遺伝的アルゴリズムを動作させながら、複数の GPU で並列に局所探索を行うことによって高速に最適解に近い解を探索することができる。

本プロジェクトでは、TSUBAME3.0 の豊富な GPU を活用することによって、QUBO 問題の解の探索速度を、GPU を増やすごとに向上させる手法を研究した。MPI と OpenMP を組み合わせたスケーラブルな実装方式を考案し、TSUBAME3.0 での実験で GPU の数に比例した性能向上ができることを実証した。

概要

量子コンピュータの一方式である量子アニーリングは、あらゆる組合せ最適化問題をイジングモデルまたは二次無制約二値最適化問題(QUBO)として定式化することで高速に近似解を求めることができる。しかし現在の量子アニーリングが対応する問題のサイズは限られており、また必ずしも最適解が得られるとは限らない。そこで本プロジェクトでは、多数のGPUを搭載したスーパーコンピュータを活用することで高速な解の探索を行う手法を開発した。

結果および考察

TSUBAME3.0を最大64ノード(=256GPU)使用して提案するQUBOソルバの解精度と探索速度、特にそれらのスケーラビリティを評価した。

評価のためのベンチマークとして、QUBO問題の入力をランダムに生成したランダム問題に加えて、一般的なNP困難の組合せ最適化問題として最大カット問題、巡回セールスマン問題、最大独立頂点集合問題をQUBO問題に変換したものを利用した。

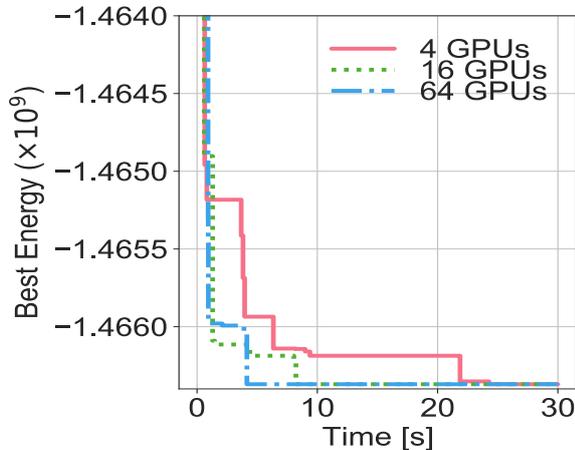


図 1. ランダム問題の探索結果

図 1 に 4096 ビットのランダム問題の探索結果を示す。時間経過とともにエネルギーが下がっていく様子が見られる。最も GPU 数が多い 64GPU の場合でより早く低いエネルギーに到達していくことがわかる。最大カット問題、巡回セールスマン問題、最大独立頂点集合問題においても同様の結果となり、GPU の数を増やすことで解精度が上がり、より最適解に近い解を出す確率が高いといえる。特に図 2 に示す最大独立頂点集合問題では GPU の数による解精度の差が顕著に現れている。

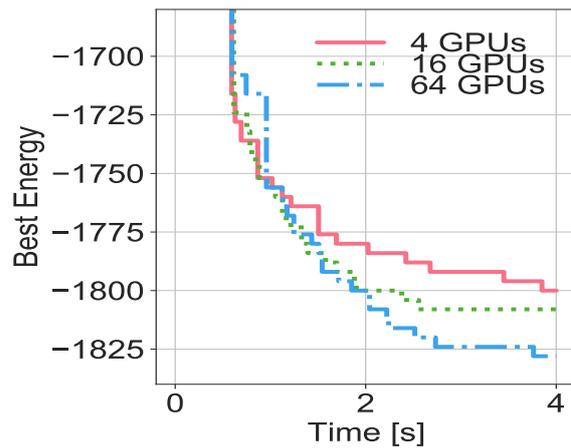


図 2. 最大独立頂点集合の探索結果

次いで探索速度の評価を行った。1秒間に目的関数を評価できる解の数を表すサーチレートを評価指標とした。

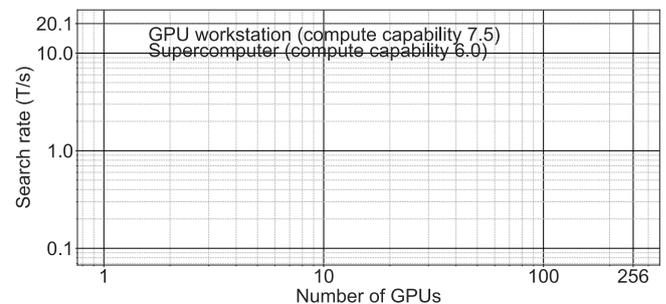


図 3. サーチレートの線形的向上

図 3 は GPU 数を 256 まで増やしたときのサーチレートを示している。NVIDIA RTX 2080 Ti を 4 基搭載した GPU ワークステーションでは 1.2 T/s のサーチレートであるのに比べ、TSUBAME3.0 では最大 20.1 T/s のサーチレートまで線形的に向上することがわかる。なお、GPU 1 台あたりのサーチレートが異なるのは GPU が異なるためである。

まとめ、今後の課題

MPI と OpenMP を用いた効率的な並列計算手法によって、QUBO の探索速度を GPU 数に比例して線形的に向上させられることを明らかにした。TSUBAME3.0 において動作させた結果、256GPU の場合で 1 秒間に 20 兆を超える解のエネルギーを計算することができた。探索速度の向上によって、解の質も向上する。

今後は探索速度の向上だけでなく、より大規模な QUBO 問題を効率よく解ける手法を研究していく予定である。