

jh240002

ソフトウェア工学による自動チューニング技術の新展開

片桐 孝洋（名古屋大学）

概要

高性能な数値計算ソフトウェアを開発する際、性能チューニングは大きな開発コスト（工数）が必要となるだけでなく専門知識が必要とされ、誰もが行えるわけではない。一方、高性能数値計算ソフトウェアの開発工数を削減する目的のソフトウェア工学の研究は、我が国ではほとんど行われていない。そこで本提案では、ソフトウェア工学を専門とするコンピュータサイエンス（CS）学者との連携を目的とする。工数削減に寄与する自動チューニング（AT）、および、科学技術計算で必須の GPU 最適化、固有値・非線形ソルバ、および、最先端の計算科学事例として計算化学と量子アルゴリズムを取り扱う。さらに、数値解析を専門とする応用数理（計算科学）者との協業も行い強力に学際領域研究を推進し、新しい AT 研究の境地を開拓する。

1. 共同研究に関する情報

(1) 共同利用・共同研究を実施している拠点名

東京大学 情報基盤センター

名古屋大学 情報基盤センター

九州大学 情報基盤研究開発センター

(2) 課題分野

大規模計算科学課題分野

(3) 参加研究者一覧と役割分担

- 代表者: 片桐孝洋（研究統括、AT 開発）
- 副代表者: 星野哲也（GPU 最適化、AT 開発）
- 共同研究者: 森崎 修司（ソフトウェア工学）、大島聡史（GPU 最適化、AT 開発）、中島研吾（連立一次方程式ソルバー）、Osni Marques（AT 開発、固有値問題）、Weichung Wang（AT 開発、固有値問題）、Feng-Nang Hwang（非線形ソルバ）、望月祐志（化学計算、量子計算）、杉崎研司（化学計算、量子計算）、河合直聡（連立一次方程式ソルバー、量子計算）
- 共同研究者（大学院生）: 森下誠（AT 開発、量

子計算）、羽生達郎（AT 開発、AI 適用）、湯淺義尚（AT 開発、AI 適用）、IVAN LUTHFI IHWANI（非線形ソルバ）、任軒正博（AT 開発、AI 適用）、植野貴大（AT 開発、AI 適用）、水島慎吾（AT 開発、AI 適用）、百武尚輝（AT 開発、AI 適用）、樫村寛大（AT 開発、ソフトウェア工学）、水木直也（AT 開発、量子計算）、中谷崇真（AT 開発、AI 適用）、棗恒輝（AT 開発、AI 適用）、秋葉康佑（AT 開発、AI 適用）、林俊一郎（AT 開発、AI 適用）

2. 研究の目的と意義

高性能な数値計算ソフトウェアを開発する際、性能チューニングは大きな開発コスト（工数）が必要となるだけでなく専門知識が必要とされ、誰もが行えるわけではない。一方、高性能数値計算ソフトウェアの開発工数を削減する目的でのソフトウェア工学の研究は、我が国ではほとんど行われていない。そこで本提案では、自動チューニング（AT）、GPU 最適化、ソフトウェア工学を専門とするコンピュータサイエンス（CS）学者と、計

算化学と量子アルゴリズム、固有値・非線形ソルバを専門とする応用数理（計算科学）者との協業による学際領域研究を推進し、新しい AT 研究の境地を開拓する。

目的Ⅰ) ソフトウェア工学：数値計算ソフトウェアにおけるソースコードの更新、コンパイル、静的解析、ビルド、自動テストの実行といった一連の更新手順を自動化し、効率化する**継続的インテグレーション**において、**自動テストの実行順序を工夫**することにより、デバッグを含む開発効率を高める手法が提案されている。これまで、**Fault-prone module prediction** といったバグを含みやすいソフトウェアモジュールを予測する研究はあったが、予測の対象はソフトウェアモジュールであり**テストケース**ではない。また、実行する前にどのようなテストケースによって問題が見つかりやすいかを予測することは対象ソフトウェア、開発プロセス、開発者に依存する可能性が高く、その予測手法は明らかではない。

本課題では、**数値計算ライブラリ LAPACK** をまずは対象にして、**数値テストケース**において本課題の解決を狙う。加えて、AT 技術の研究は、高性能化のための**パラメタチューニング**や**コード自動生成**で工数削減に寄与するのが自明であり、**生産性の観点からソフトウェア工学上の意義**をなす。

本提案では、代表者の片桐が開発した AT 言語の **ppOpen-AT フレームワーク**、および **IDEAS-ECP プロジェクト開発の AT ツール**を活用する。

目的Ⅱ) 実用アプリケーション評価：本課題では、数値計算法、および、量子計算への AT 適用を、実用アプリケーションの観点から評価する。特に、**計算化学における量子アルゴリズム**を想定した量子回路シミュレーション（特に GPU で稼働する **NVIDIA cuQuantum**）の高速化に寄与する性能パラメタチューニングを検討する。

一方、AT の適用アルゴリズムを拡大するため、**非線形ソルバのアルゴリズム上に現れる性能パラメタチューニング**も取り扱う。

目的Ⅲ) チューニングノウハウの共有：本提案では、名大「不老」と東大「Wisteria/BDEC-01」

の CPU と GPU を利用する。そのため、「富岳」の CPU である **ARM A64FX** と GPU の **V100** と **A100** の高性能実装技法やパラメタチューニングのノウハウが集約される。そのノウハウを論文出版等で公開することで、拠点や国家スパコンの利用技術の知見が共有化される。

目的Ⅳ) 国際連携：参画する **Marqus 博士**、**王教授**と**黄教授**は応用数理分野の著名な研究者であるため、本課題の成果が国際的にも周知されることで国際連携を強力に進めることができる。

3. 当拠点の公募型共同研究として実施した意義
本提案は CS／計算科学／応用数理の研究者との協業であり学際性に富む。また CS 分野においても、高性能計算 (HPC) とソフトウェア工学の協業は、米国ではエネルギー省 **IDEAS-ECP プロジェクト**など大規模ファンドの支援があるが、我が国では皆無である。本課題は、実用ソフトウェアの大規模化／並列化した数値計算アルゴリズム開発に加えて AI 処理が含まれ、複数拠点の計算機資源活用が必須であるため、拠点の公募型研究として実施した意義を有す。

4. 前年度までに得られた研究成果の概要

■2023 年度成果のまとめ

昨年度、代表者の**片桐**、副代表の**星野**、協力者の**河合**と博士学生の**任**は、NICAM ベンチマークの物理計算層の長大ループ中の計算処理に**混合精度計算を入れ込む最適化の AT**の新方式を開発した。

博士学生の**森下**と共に**疑似量子アニーリングと量子回路シミュレータのハイパパラメタ最適化の AT 方式**の提案を行った。

ソフトウェア工学の専門家の**森崎**は**片桐**と学生**の樫村**とともに、LAPACK の固有値ソルバテストルーチン (STCollection) のテストシーケンスの効率化に関し、特定ケースでエラー検出ができる事例を発見し、**テストシーケンスの最適化に資する基礎成果**を得た。

星野は共同研究者の**望月**開発の ABIBIT-MP に関

するコード最適化と GPU 最適化の検討を行った。アプリケーション適用では、片桐と大島は望月と杉崎ら計算化学分野の研究者と、名古屋大学の「不老」Type II で cuQuantum を利用した量子アルゴリズムの研究成果を創出した

国際連携の観点では、米国 IDEAS-ECG プロジェクトの代表者の Marques 博士は固有値計算ライブラリ LAPACK の開発者で、固有値テストルーチンの STCollection の開発者でもある。また王教授は固有値アルゴリズム開発連携しており、かつ黄教授は非線形ソルバ開発において成果を創出した。

■ソフトウェア工学上の成果

(1) LAPACK テストシーケンスの効率化：片桐、森崎、榎森は、LAPACK で用いる BLAS に数値バグが混入する場合において、STCollection のテスト順序を入れ替えることで、テスト時間の短縮を大幅に改善可能である事例を発見した。

(2) AT ツール GPTune の LAPACK 適用：片桐、Marques、森下は、LAPACK のパラメタチューニングに米国 ECG 開発物の AT ツール GPTune を適用し性能チューニングする基礎成果を創出した。

(3) AT の AI 活用と説明可能 AI：AT 機構に最新 AI を適用することは喫緊の課題である。しかし AI モデルが妥当な選択をするかの解釈（説明可能 AI (XAI)）は依然大問題である。特に数値計算に対する XAI の研究は限定される。本課題に対して昨年度、河合の開発した閾値付不完全コレスキー前処理付き CG 法ライブラリ中の性能チューニングの AT に深層学習を適用した事例の XAI を研究した。その結果、XAI ツール SHAP を用いて、当該事例でも妥当な AT ができることを世界で初めて示した。

5. 今年度の研究成果の詳細

本提案は 3 年計画の 2 年目である。ソフトウェア工学、AT、アルゴリズム研究の各側面から研究を行なった。

① ソフトウェア工学

目的 I) ソフトウェア提供する機能や能力を期

待通り満たすかどうかを確かめるためにソフトウェアテストを実施する。ソフトウェアに与えられる入力値を幅広くカバーし、様々な入力に対して期待通りの出力が得られるかを確かめるために様々なテストケースを実行する。バグ等の原因で期待通りの結果とならない場合、その部分を修正して再びテストケースを実行し、すべてのテストケースにおいて問題がないことを確かめる。このとき、バグが検出されやすいテストケースを選ぶことができれば、テストケースの実行時間を短縮することが期待される。

そこで前年度は、LAPACK におけるテストプログラム STCollection のテストケースを優先順位付けし、問題を検出しやすいテストケースが実際にあることを発見した。本年度は、この前提を活用し、固有値分布を入力としたディープラーニングにより生成したモデルと、固有ベクトルを入力とした CNN モデルを作成した。どちらのモデルもバグ発生率を出力する。

この学習のデータセット、および、検証データには、STCollection で用意されているテスト問題を活用した。テスト問題の、固有値分布、三重対角行列を、学習データとして使用する。生成した行列のサイズは N=60 である。合計 360 個のデータを用意し、その 8 割を訓練データとした。結果を図 1、図 2 に示す。

図 1 と図 2 から、訓練の回数増加に伴い損失は減った。また精度は常に 80% 以上と高い値を示しており、一定の精度でバグ発生率を予想出来ている。

ここでの実験は、直接的に未知のテストデータに対応するものではない。しかし、少なくとも STCollection におけるテスト問題においては、バグが生じやすい固有値分布と固有ベクトルを予想できるという結果を得ている。すなわち、未知のテスト問題でも、この学習セットによく似た固有値分布や固有ベクトルであれば、バグが生じやすい問題を予想できる可能性を示唆している。

以上のように本年度は、未知のデータセット対応に向けた深層学習適用の予備評価に成功した。

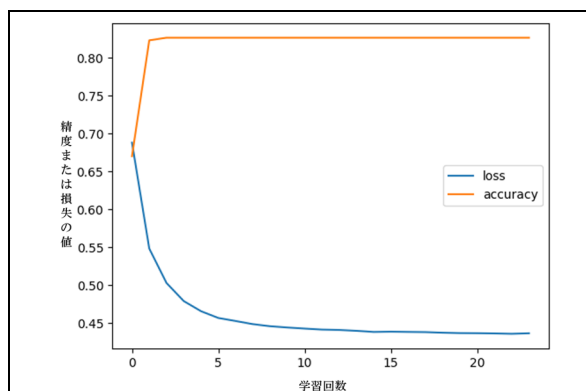


図 1 入力固有値分布の場合の学習推移

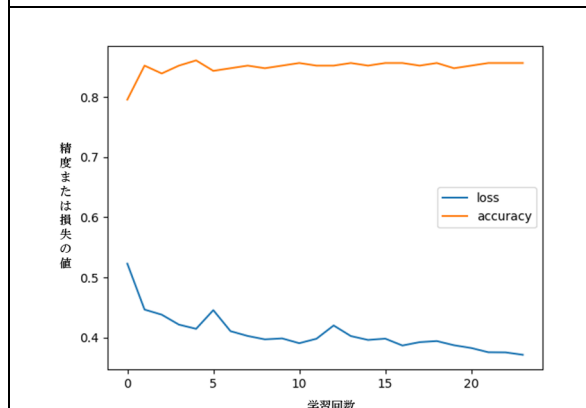


図 2 入力固有ベクトルの場合の学習推移

② AT 研究

目的Ⅱ) 実用アプリケーション評価では、量子関連研究（量子アニーラ）への AT 適用で大きな進展があった。量子アニーラでは QUBO と呼ばれる形式に帰着して最適化問題を解く。その際、解の精度に影響する性能パラメタがあり、チューニング工数が増える。また、解の性能に影響する。本年は量子アニーラへの AT 適用で、多方面にわたり大きな成果を上げた。

まず、**疑似量子アニーラの性能パラメタチューニング問題**に対し、**サポートベクターマシン (SVM)**における有効性検証を行い、査読付国際会議論文[3]を発表した。また SVM を、**我が国の主要な疑似量子アニーラ**全般における解の精度と実行時間の観点での網羅的評価[10]を行った。図 3 は、我が国の**主要疑似量子アニーラ 4 種**（日立 CMOS、NEC VA、東芝 SQBM、Fixstars Amplify AE）および、PC 実行（古典）における Time To Solution (TTS) の結果である[10]。図 3 から、TTS の観点で大きな違いがあ

る。この結果はは、アニーリングパラメタ設定方式の違いや、扱う問題の性質の違い、問題規模の違いで差が出ると予想されるが、PC 実行と比較した網羅的な評価として意義がある成果である。

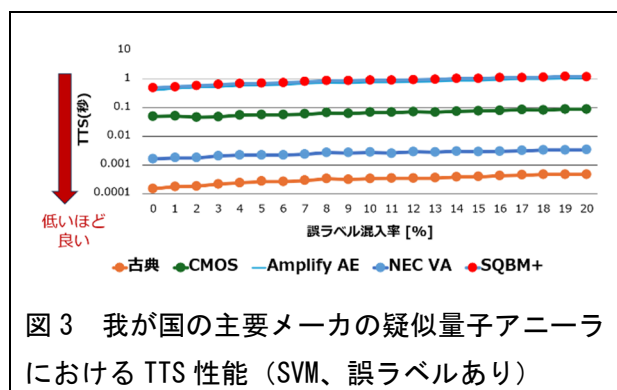


図 3 我が国の主要メーカの疑似量子アニーラにおける TTS 性能（SVM、誤ラベルあり）

以上に加えて、疑似量子アニーラのパラメタチューニングに **AT ツール Optuna** を適用し AT 効果を評価した[6][7]。加えて、同様の性能チューニングが光量子アニーラにも有効であることを示した[6][8][11]。また、スパコンと量子計算環境を融合する **HPC-Centric Quantum Computing** という新しいコンセプトを発表した[14][18]。

③ アルゴリズム・実装技法研究

目的Ⅱ) の対象アプリである **ABINIT-MP の GPU 化**[17]を行った。加えて、高性能化に関する成果[1][2][19]がある。このことで、**目的Ⅲ) チューニングノウハウの共有**がなされた。

目的Ⅱ) に含まれる、**計算化学と量子アルゴリズムの研究**を創出した。さらに目的Ⅱ) の非線形ソルバのアルゴリズムでは、黄開発の inexact Newton 法による非線形ソルバの並列アルゴリズムの改良の性能評価を実施している。

④ 国際連携

目的Ⅳの国際連携では、米国 ECP におけるソフトウェアエンジニアリングの専門家と連携し、**国際会議 IEEE eScience2024 のワークショ**

ップにおいて、ソフトウェア工学の観点での招待講演を行った。具体的には、持続的な HPC ソフトウェア構築に関する **AT 方式**[8] および **評価メトリクス**[9]に関する招待講演を行った。

⑤ 計算リソース活用

機械学習や性能向上のデータ取得のため、② AT 研究を中心に十分に活用された。

ムにおける量子回路シミュレーションの利活用時の GPU 活用による高速化事例も検証済みである。加えて、非線形最適化の国際共同研究成果も継続していることから、AT 適用によるさらなる研究進展が期待できる。

④ 計算リソース活用

成果創出が飛躍的に進んでいる②AT 研究を中心に十分に利用された。

6. 進捗状況の自己評価と今後の展望

① ソフトウェア工学

今回実施した STCollection のテストケース最適化事例、および未知のテストケースにおけるテスト優先度を評価する深層学習モデルの適用評価は、当該分野では研究事例が無く、高い学術的価値があると自己評価する。数値計算ソフトウェアのテスト手法のソフトウェア工学上の研究がほとんどないことを鑑みると、そのインパクトは少ない。本年度成果を拡張することで、汎用的なテスト手法の開発が可能となるため、予備評価としては十分であると自己評価する。

以上を総合して、本年度の最終評価時の達成度は 100%と自己評価する。

② AT 研究

AT 研究においては、多方面で成果の創出があった。本年で最も成果が出たのは、すでに説明したように、疑似量子アニーラの性能チューニングに関する成果である。この成果は、査読付論文、査読無論文および口頭発表であり、[3][6][7][10][11][12]と 7 件に及ぶ。今後も、飛躍的な成果創出が期待できる。

一方 AT 方式の観点では、大気海洋コードにおけるループ変換[5]、疎行列反復解法における説明可能 AI による AT 機能の検証[4]などの成果展開が行われ、さらなる発展が期待できる。

③ アルゴリズム・実装技法研究

ABINIT-MP の GPU 化による高速化は最大で 30～50 倍と極めて効果が大きい事例[16][17]ほか、高速化に関する査読付論文[1][2]もあり、成果の創出がなされた。また、計算化学と量子アルゴリズム