

jh250015

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

高性能数値計算ソフトウェアを開発する際、性能チューニングには工数が必要となるだけでなく専門知識が必要とされ、開発コストの削減が望まれる。一方、ソフトウェア工学(SE)と HPC との学際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 SE 学者との学際研究を行う。工数削減に寄与する自動チューニング(AT)、GPU コード最適化、固有値・非線形ソルバ、計算化学、量子関連計算、に加えて本年度は新規に機械学習アプリとコード生成 AI も取り扱う。

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

東北大学サイバーサイエンスセンター

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

(A) ソフトウェア工学適用 (主: 片桐・森崎、担当: 星野・椋木・櫻村)

(B) AT 方式開発 (主: 片桐、担当: 星野・大島・椋木・河合・Marques・水木・百武・中谷・任・棗・秋葉・林・末永・樹神・磯部・阪口・内田・森田・三笠)

(C) アプリケーション展開 (主: 片桐・星野、担当: 望月・杉崎・椋木・加藤・中島・Hwang・伊藤・片桐 (里)・IHWANI)

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(SE)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。また、工数削減に寄与する自動チューニング(AT)を研究する。エクス時代のスーパーコンピュータで必須の AI と GPU 最適化に加え、応用として固有値・非線形ソルバ、計算化学と量子アルゴリズム、機械学習アプリケーションを取り扱う。数値解析を専門とする応用数理(計算科学)者との協業も行うことで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。

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学際性に富む。CS 分野においても、高性能計算(HPC)と SE の協業は、米国ではエネルギー省 IDEAS-ECP プロジェクトで前例があるが、我が国では皆無である。本年度提案は数値計算プログラムの大規模化・並列化に加え、データ駆動科学の中核をなす AI 処理が含まれており、複数拠点の計算機資源活用が必須となる。

■ソフトウェア工学の観点: 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、予測手法は明らかではない。本提案では、**数値計算ライブラリ LAPACK** を対象にし、**数値テストケース**において SE 専門家と課題解決を進めている。加えて、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 AT 研究は、**生産性の観点から SE 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、および IDEAS-ECP プロジェクトや標準の AT ツール等を活用する。

■**応用の観点**：将来のスパコン環境で想定される、**計算化学・量子アルゴリズム**への適用を想定し、**GPU コード最適化**と**量子関連計算**の高速化に寄与する性能パラメタチューニングを対象とする。また、**機械学習アプリケーション**を対象としてデータ駆動科学への対応を行う。一方、AT 適用分野の拡大に資する**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。

■**チューニングノウハウの共有**：本提案では、名大「不老」、東大「Miyabi-G」「Wisteria(Odessey)」を利用する。そのため、「富岳」の CPU である **ARM A64FX** と GPU の **V100**、**GH200** での AI・GPU 性能、数値計算の高性能実装技法、および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し、拠点や国家スパコンの利用技術の知見が共有化される。

■**国際連携**：Marques 博士(**IDEAS-ECP** 代表者)、王教授・黄教授(**固有値**・**非線形アルゴリズム**)は

専門家)は**応用数理分野**の著名な研究者である。Marques 博士は**名古屋大学招へい教授**の実績がある。王教授が兼務する台湾の NCTS と黄教授の国立中央大学とは**名古屋大学情報基盤センターとの共同研究契約を締結**済である。したがって、代表者の片桐とは強力な国際連携の実績を有す。

4. 前年度までに得られた研究成果の概要

(A)**ソフトウェア工学適用**：**SE 専門家**森崎・片桐・大学院生の樫村は、**STCollection** のテストシーケンスの効率化の研究を行った。その結果、最も利用される **BLAS** にバグが混入する事例で、特定のテストケースでエラー検出がされることを発見した。このことで、**最大で 5.27 倍のテスト時間の短縮**ができる可能性があることを明らかにした。

(B)**AT 方式開発**：河合開発の反復解法ソルバ **ICTCG** の性能パラメタの AT を可能とする AI モデルの構築に成功した。加えて、**XAI ツール SHAP** を利用し、数値計算アルゴリズムの観点で妥当性検証を世界で先駆けて成功した。また昨年参加の修士・博士学生と共に、サポートベクターマシン(SVM)を**疑似量子アニーリングに適用した**。その結果、**古典環境より高精度解が得られる可能性**を示し、AT 適用の見込みを付けた。加えて、量子アニーリング環境特有の AT 課題を議論した。

(C)**アプリケーション展開**：

■①望月開発 **ABIBIT-MP** の **GPU コード最適化**において星野と片桐は GPU 化の方式検討に参加し、**CPU に対して最大で 65.7 倍の速度向上**を得た。これらは、査読付論文誌で成果を出版した。

■②黄開発の **inexact Newton** 法におけるアルゴリズムの改良を行った。

■**国際連携**：Marques は固有値計算ライブラリ **LAPACK** の開発者であり **STCollection** の開発者でもある。王は固有値アルゴリズム、黄は非線

形ソルバ開発において成果を創出している。そのため、国際連携の準備は万全である。

以降、主成果について詳細を述べる。

(A) [ソフトウェア工学適用](#) ■ **LAPACK テストシーケンスの効率化**：片桐、森崎、檜森は、LAPACK で用いる BLAS に数値バグが混入する場合において、バグ混入度から特定のテストケースのみでバグが発症する事例を発見した。この事例から、テスト順序を入れ替えることで、**テスト時間の短縮が可能**である。この発見は **SE 分野**では自明ではないと考えられており、学術的に興味深い。

(B) [AT 方式開発](#)：■ **Optuna 適用**：片桐は昨年度協力者学生とともに、AI 業界で標準的に使われている **Optuna** を利用し、最小頂点被覆問題に対する量子アニーリングで必須の**パラメタチューニングを自動化する AT 方式を開発**した。

(C) [アプリケーション適用](#)：■ **量子関連技術と AT**：AT による工数削減観点から、現在ほとんど行われていない HPC による量子関連計算を取り扱った。**疑似量子アニーリング(QIA)**を対象に、フィックスターズ社の Web サービス Amplify を利用した。昨年度は **SVM の QIA 適用**を実現した。さらに NTT が開発した**コヒーレントイジングマシン(CIM)**に適用拡大した。Cyber-CIM (CIM-CACm) の性能パラメタ 5 種類 (<https://github.com/NTTRI-PHI-Algorithms/CACm>) のチューニングを行い、AT 適用の予備評価を行った。**約 3 万点の網羅的なパラメタサーチ**を行い、停留形状の底を抜けるところに最適解の集合がある特徴を、世界で初めて明らかにした。

5. 今年度の研究成果の詳細

(A) [ソフトウェア工学適用](#)

■①ソフトウェア(**LAPACK**)が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 Marques 博士が開発した固有値ソルバテストルーチン **STCollection** の**ソフトウェアテスト**を採用した。また、LAPACK の固有値問題ルーチン

に対して、実装から考えられるバグケースを 3 つ考案し、それに基づくバグを混入したデータを取得した。このバグデータによる固有値/固有分を入力として、複数の統計量データ計算の上で、機械学習に入力させる教師データの入力とし、バグのある/なしを出力とする機械学習のモデル生成手法を提案した。この手法の評価から、**バグを予想する AI モデルの作成に初めて成功した**。

次に、この AI モデルを用いて**任意のテストシーケンスを最適化する AT 手法**を提案した。本年度は、学習に使っていない未知のテストシーケンスによる評価を行った。並び替えられた各テストシーケンスについて、性能指標である APFD (Average Percentage of Faults Detected)を尺度に採用した。本テストは、評価対象である 3 つのルーチン(dgemm, dlaed3, dlaed4)からそれぞれ一つずつをランダムに選択した。その結果を表 1 に示す。表 1 より、全ルーチンにおいて提案手法による並び替え後の APFD 値はランダム実行(0.9751)を大きく上回っている。したがって、バグを含むテストケースを効果的に前方に集約できていることが確認できる。分類精度の低かった dlaed4 においても、他のルーチンと同様の APFD 値を出すことができ、実用的な順序付けが可能であることを示した。

表 1： 機械学習モデルによる未知のテストシーケンスに対する並び替えの性能評価

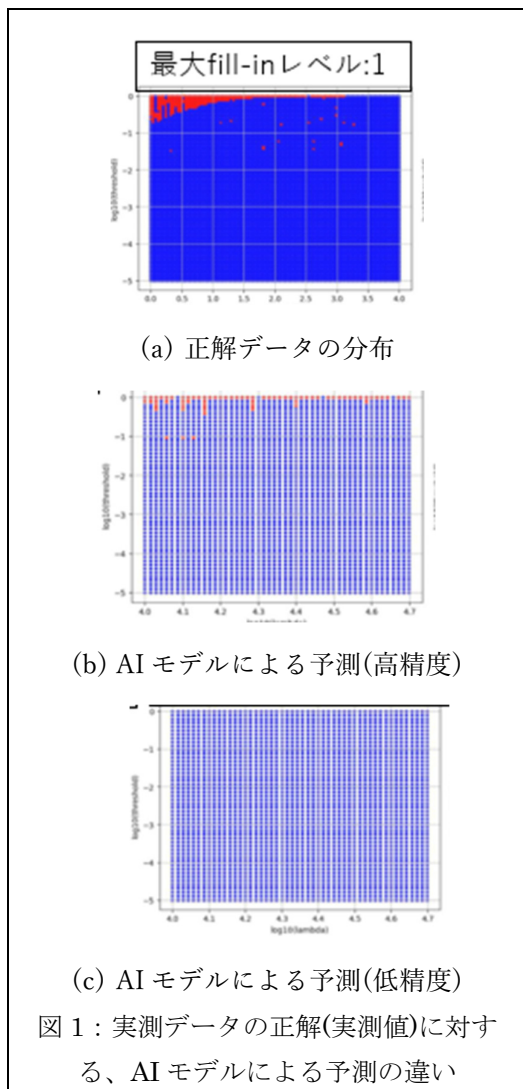
	ランダム	dgemm	dlaed3	dlaed4
APFD	0.9751	0.9988	0.9988	0.9988

この成果は、情報処理学会で発表したほか、名古屋大学の修士論文となった。

(B) [AT 方式開発](#)

■①疎行列反復解法 **ICTCG** における混合精度計算の **AI モデル**生成と、AI モデルの妥当性検証の成果を推進した。具体的には、説明可能 AI (XAD)のテーマとして、まず ICTCG 法の性能パ

ラメタ、および、疎行列の画像情報を入力として、出力に ICTCG 法による実行時間を出力する AI モデルを作成した。その AI モデルを基に、倍精度、単精度、混合精度（倍精度+混合精度）の 3 種類の実行時間を予測し、最適な演算精度を予測する自動チューニング手法を提案した。AI 出力の検証の観点での性能評価の結果、学習データにない性能パラメタ入力時に予想できない原因を明らかにした(図 1)。

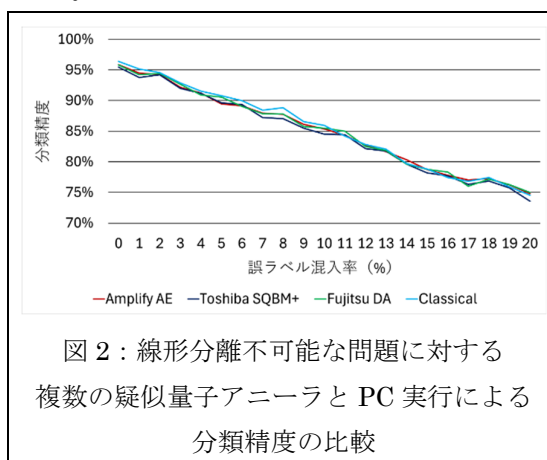


また、混合精度ソルバにおける演算精度の新たな選択肢を検討するため、最新の 4 倍および 6 倍精度演算アルゴリズムを、CG 法に適用し性能を評価した。

以上の成果は、情報処理学会で発表し、名古屋大学の修士論文となった。

■②量子アニーリングの性能パラメタチューニングに関して、昨年度は成果を複数の疑似量子アニーラに展開を行い、性能評価を行った(図 2)。本成果を、情報処理学会研究報告で発表したほか、名古屋大学の修士論文となった。

また、光量子アニーラのパラメタチューニングに関する新方式を提案した。この新方式は、単純な複数パラメタの同時ベイズ探索に対しての優位性を示した。本成果は、国際会議 Candar のワークショップにおいて査読付国際会議論文として発表した。



(C) アプリケーション展開

■①計算化学ソフトウェア ABINIT-MP(FMO 法)の GPU 化について、GPU 対応は MP2 部分を検討中である。また、RI 積分のルーチンについては、プロトタイピングを推進した。加えて、④(i)の LLM を利用したコード生成について、ABINIT-MP の関連の HF ミニアプリを提供した。

■②量子アニーリングによるタンパク質量み込みモデルの研究を推進した。

■③深層学習アプリケーションに関して、以下の進展があった。

(i)大規模言語モデル(LLM)：本年度、LLM によるコード生成 AI により HPC コードを自動生成する、マルチエージェント AI システムである VibeCodeHPC を開発した。

VibeCodeHPC は、現在、ClaudeCode で標準的な機能として提供されているが、その機能よりも半年前に実現した、世界初のマルチエージェント AI による HPC コード生成システムである。単なる実装ではなく、エージェントによるコンテキストメモリ制御による継続実効性を高めるなど、先進的な実装を行ったものである。また、HPC 分野初のマルチエージェントによる並列化 (OpenMP、MPI、OpenACC) のほか、単体最適化 (ブロック化、SIMD 化など) を完全自動で行う自動チューニングシステムである (図 3)。HPC 分野では初めて、マルチエージェント構成の有効性を示した。

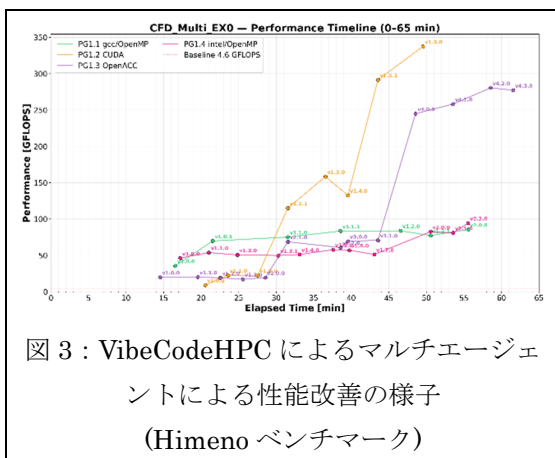


図 3 : VibeCodeHPC によるマルチエージェントによる性能改善の様子 (Himeno ベンチマーク)

これらの成果を arXiv で論文発表するとともに、GitHub でソースコードを開発した。また、国内会議 xSIG2025 でポスター展示し、**優秀学生ポスター賞を受賞**した。加えて、多くの国内、国際会議での発表を行った。

加えて、またローカル LLM を活用した軽量なコード最適化システムの試作も行った。

一方、その他の LLM によるコード生成 AI エージェントの研究も推進した。まず、ローカル LLM を想定した、強化学習によるファインチューニング手法を提案し、その有効性を評価した。次に、AI エージェントによる研究の全自動化として、アイデア策定、プログラミング、データ取得と図表の作成、論文生成までを行う AI Scientist の研究がある。ただし既存の AI Scientist は AI 研究の研究ジョブフローに特化していた。そこで、HPC 研究のジョブフローに特

化した AI Scientist for HPC の機能開発と有効性評価を行った。

(ii) **ロボットサービス** : マルチモーダルデータを用いた感情推定モデルを人とロボットのインタラクションに活用する手法を開発し、VR 環境を用いた評価システムに組み込んだ[3]。また、自律移動ロボット制御のための環境情報を統合した歩行者経路予測モデルの構築、および深層強化学習に基づく経路探索手法を開発し、スパコンを用いたモデル学習および評価実験を行った (成果については国内学会でポスター発表を行った)。歩行者経路予測モデルに関する成果は、現在、査読付き国際会議に投稿中である。

(iii) **教育ビッグデータ解析** : 高等教育機関における学生の退学リスクを早期に検出し、学修の支援を行うことを目的に、線形判別分析による次元削減処理と、二値分類確率に基づく補助特徴量の生成を組み合わせた独自のパイプラインを構築することで、特定の教育ドメインやデータ構造に依存しない柔軟かつ頑健な予測モデルを開発した。スパコンを用いたモデル学習および評価実験を行い、成果を国内学会で発表した (**学生奨励賞を受賞**)。

■④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性能評価と改良を進めており、性能改良を行った。

■⑤国際連携:2025年12月5日に名古屋大学で、**JHPCN Field Workshop · State-of-the-Art in Code Generative AI for High-Performance Computing** を開催した。米国の HPC 分野の研究者を招聘し、本課題のメンバーと AI の HPC 適用に関する講演を行い、情報交換を行った。

6. 進捗状況の自己評価と今後の展望

本提案は多岐のテーマにわたるが、全分野において複数の論文成果が出ている。このことから、

自己達成率は 100%と判断する。

(A) ソフトウェア工学適用

本年度、未知のテストデータに対する最適化が可能な AI モデルの作成に関する新規手法の提案を行った。加えて、未知のテストケースでの有効性を評価した。この観点から、当初の研究目的を達成したものと自己評価する。

(B) AI 方式開発

本年度は、世界初の LLM によるコード生成 AI を用いたマルチエージェント AI システム VibeCodeHPC の開発に成功した。それだけではなく、行列積や流体などの典型的 HPC ベンチマーク適用を行い、マルチ AI エージェントシステムの有効性を示した。加えて、多数の学術発表を行い、受賞もされた。以上から、LLM によるコード生成 AI による AI 方式においては、初期の目的を完全に達成したと判断する。

量子アニーラにおいては、(i) 複数のアニーラ評価、(ii) アニーラのパラメタ向けに特化したパラメタ調整手法の高度化 (AI 手法の提案) を行った。これらは、学会発表に加えて、査読付論文の発表を行った。そのため、当初の目的を完全に達成したと判断する。

(C) アプリケーション展開

計算化学ソフトウェア ABINIT-MP の GPU 化の高度化を実現し、学術論文発表を行った。今後、コード生成 AI への展開など、発展が期待できる。

深層学習のアプリケーション (ロボットサービス、教育ビッグデータ解析) については、スパコン上でデータ検証を行った。これらは、スパコンによる同分野の研究推進であると評価する。その結果に対し、複数の学術論文発表を行い、複数の受賞を得た。

以上から本研究について、当初目的を実現したと自己評価する。