様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　２ 年　 ６ 月　 ５ 日現在

機関番号：１３９０１
研究種目：挑戦的研究（萌芽）
研究期間：2018～2019
課題番号：１８Ｋ１９７８２

研究課題名（和文）ディープラーニングを利用した革新的自動チューニング基盤の創製

研究課題名（英文）Development of Innovative Auto-tuning Middleware With Deep Learning

研究代表者
片桐　孝洋（Katagiri, Takahiro）

名古屋大学・情報基盤センター・教授

研究者番号：４０３４５４３４

交付決定額（研究期間全体）：（直接経費）　　4,800,000円

研究成果の概要（和文）：GPUやメニーコアCPUに代表される約300スレッド実行が可能な先進計算機アーキテクチャがもたらすチューニング作業の困難性の爆発的増大から、ソフトウェア性能を人手を介さず最大限に引き出す仕組み（自動チューニング、AT）が求められている。一方、近年ディープラーニング（DL）の技術進展がはなはだしく、多くの分野へ適用がなされている。DLは本来AT方式を実現する手法の１つであるが、DLを用いたAT方式の開発は殆どなされていない。そこで本研究では、(1) 数値計算ライブラリの性能パラメタチューニング；(2) AT基盤インターフェース開発；(3) スーパーコンピュータへの適用；の研究を行った。

研究成果の学術的意義や社会的意義
（１）数値計算ライブラリにおいて収束性に影響し実行時間に大きな影響を及ぼす前処理選択がある。本研究では前処理選択へ活用できるDLを用いたAT方式を開発した。これにより、数値計算を低いコストで高性能実行できる環境に貢献し、ものつくり等の生産性の向上に資する。（２）提案するAT方式の実用化に向け実行時の性能の揺らぎに対しても追随できるように改良を行なったことで、より堅牢なATシステムの実現に資する。（３）GPUやメニーコア環境における数値計算コードの最適化行うことで、最新計算機環境における最適化と性能評価のためのコードやデータを集め、高性能数値計算プログラム開発のコスト削減に資する。

研究成果の概要（英文）：Dramatically increase of difficulties for tuning work caused by advanced computer architectures, such as GPUs and many-core CPUs with ability of approximately 300 threads execution, requires a method to obtain maximum performance of software without manual tuning works, which is known as auto-tuning (AT) technology. On the other hand, technology of Deep Learning (DL) is dramatically progressing in several fields. Although DL is one of AT methods, there is little study for AT method with DL. Hence in this research, we have studied as follows: (1) Tuning performance parameters for numerical libraries; (2) Developing an AT basic interface; (3) Adaptation of super-computers.

研究分野：高性能計算

キーワード：ディープラーニング　自動チューニング　前処理方式選択　疎行列反復解法　Xabclib　GpGPU　FIBER方式　実行時最適化

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

Graphics Processing Unit (GPU)や Intel Xeon Phi に代表される 1 CPU で約 300 スレッド実行が可能なメニーコア CPU などの先進計算機アーキテクチャがもたらすチューニング作業の困難性の爆発的増大から、ことにソフトウェア性能を、人手を介さず最大限に引き出す仕組みが求められている。とりわけ、数値計算ライブラリにおいては、個々のソフトウェアに特化した入力データに対しても高性能に動作することが強く要求される。

このような状況のもと、利便性の向上と多様な入力や計算機環境での高性能化のため、アプリケーションの性能パラメタを対象計算機のキャッシュサイズ、コア数、通信性能などの計算機アーキテクチャの特性に加えて、数値アルゴリズム選択に至る広範な要因を自動的にチューニングするソフトウェア自動チューニング技術（以降、「AT 技術」と記載する）が、国内外から注目を集めている。

一方、近年ディープラーニング（以降、DL）の技術進展がはなはだしく、多くの分野へ適用がなされている。機械学習は、本来 AT 方式を実現する手法の１つであるが、DL を用いた AT 方式の開発は殆どなされていない。また、数値計算ライブラリのパラメタチューニングへ DL を適用する方式自体の研究がされておらず、AT で有効となる DL の適用方式も明らかになっていない。

そこで本研究では、以下の３つを目的とし、革新的 AT 基盤の創製を目指す。

(1) 数値計算ライブラリの性能パラメタを DL を用いて AT する方式（以降 ATDL）開発

(2) ATDL が容易に利用できる AT 基盤インターフェースの仕様策定と参照実装

(3) DL 学習をスーパーコンピュータ（スパコン）を用いて行うツール開発

２．研究の目的

研究代表者らは、疎行列反復解法のための数値計算ライブラリにおいて簡便な利用ができ高性能を実現する実行時 AT を行う数値計算ライブラリ Xabclib を開発した。疎行列反復解法に限定した AT 機能の API 集 OpenATLib を開発した。疎行列反復解法に限定した AT 機能とは、たとえば、科学技術計算での基本演算の１つである疎行列-ベクトル積（SpMV）の実装および疎行列データ形式選択、解法数理に基づく性能パラメタ（リスタート周期など）の調整、である。

Xabclib は OpenATLib を用いて構築されている。OpenATLib を利用することで、ユーザ固有のプログラムに対して AT を適用したい処理に AT 機能を容易に付加できる。

本研究は、次世代計算機アーキテクチャにおいて AT を適用した数値計算処理、特に、疎行列解法ライブラリの適用時に AT の有効性を著しく増大させることを目的している。

疎行列反復解法ライブラリの全性能パラメタ（例えば、リスタート周期）に対し有効な方式は提案されていないばかりか、有効となる方法でさえ明らかになっていない。しかし我々の先行研究では、疎行列形状を画像化し、TensorFlow を用いて DL 学習させることで前処理方式選択に限定した AT 方式を提案し、最適な前処理方式を推定できる。

本研究では、この提案方式を発展させ、疎行列反復解法ライブラリの多くの性能パラメタに適用できる世界初の ATDL 方式の創製を目指すことを目的とする。

３．研究の方法

本研究では、以下の３グループに分け、協調的に進めることで目的を達成する。

(1)**AT 基盤設計グループ**:DL を用いた AT 機能を実現するに当たり、旧 OpenATLib に拡張を行い、新 OpenATLib を設計する。連携者は、Xabclib の設計・実装を行っているため、Xabclib の AT 機能拡張のための知識提供を行う。

(2)**AT モデル・実装方式**：d-spline による性能モデルの研究を活用し、ATDL における AT 時間の削減に資する方式を研究する。GPU およびメニーコア型 CPU 向きのコード最適化を研究し、ATDL 方式開発に寄与する。

(3)**ディープラーニング適用評価グループ**：本グループは、ATDL の方式評価を行う。TensorFlow などの DL ツールに、疎行列反復解法ライブラリ Xabclib から得られるデータを学習し最適解法を予測することで、性能評価を行う。

４．研究成果

（1）**DL を用いた自動チューニング方式開発の成果**：数値計算ライブラリには多数の性能パラメタがあるが、収束性に影響し実行時間に大きな影響を及ぼすものに前処理選択がある。

そこで本研究では、この前処理選択の AT 方式へ活用できる DL を用いた AT 方式を開発し予備

評価を行った。具体的には、以下である。

①疎行列計算のライブラリには実装選択に関わる多くのチューニングパラメータが存在し，パラメタ選択によって性能は大きく異なる。しかし、最適実装の選択は実行を伴う経験に基づいて行われ、時間的コストが高くなる。我々の先行研究では、疎行列構造を 3 次元カラー画像に変換し、ディープラーニングによる学習で前処理予測をする AT 方式を提案した。本研究では、新たな行列構造の画像変換手法を提案し、この AT 方式の改良を行った。

先行研究では、はじめにあらゆる前処理実装を実行して対象の疎行列の最適実装を調べて教師データを作成する。次に、行列構造をカラー画像に変換して学習の入力データを作る。これらのデータを用いて学習を行い、前処理選択用のニューラルネットワークを構築する。提案方式は、行列の特徴画像への変換アルゴリズムの改良である。

先行研究の画像変換方法では、青は行列サイズ、緑は疎度、赤は要素の大きさに注目して変換が行われている。その中で、青と赤は最大値と最小値によって正規化されている。提案手法では、極端な値をもつときに色の大きさが十分に反映されず、画像要素について学習率が低下する。そこで、画像要素の中央値を用いて正規化する手法を提案し、学習についてよりロバストになるように改良を行った。

画像変換方法について、既存手法と提案手法の精度を比較した。対象となる疎行列は、業界標準の疎行列データを用い、最適実装のラベリングは、名古屋大学情報基盤センター設置のスーパーコンピュータ Fujitsu PRIMEHPC FX100 を使用した。

数値実験では、疎行列ライブラリ Xabclib を利用して、連立一次方程式を GMRES 法で解く Xabclib_GMRES を使った。解ベクトルの要素が全て 1 となるように右辺ベクトルを設定し、解が収束するまでの時間を測定した。Xabclib_GMRES には前処理実装が 6 つ用意されており、もっとも速く解を得る実装を最適な前処理実装とした。

ニューラルネットワークと最適実装の教師データは共通データとする。既存手法と提案手法の画像変換方法で生成した画像をそれぞれ入力として学習して性能評価を行った。

評価結果を表 1 に示す.

表 1　最適実装予測精度の比較

既存手法(%)	提案手法(%)
予測精度：75.4	予測精度：79.8

表 1 より、既存手法と比較して 4.4 ポイント最適実装の予測精度が向上することが明らかとなった。そのため、より効率的な DL を用いた AT 方式の開発に成功した。

②本科研の基礎技術となる AT について、実用化に向けて評価対象プログラムの実行時の性能の揺らぎに対しても追随できるように改良を行なった。

AT の適用先として、機械学習の分野における学習モデル CNN の層のノード数などのハイパーパラメタのチューニングに適用し、有意な構成を導き出せることを示した。さらに、この機械学習の考え方を AT に用いて、疎行列ソルバの最適前処理予測を複数の計算機環境で行い、計算機環境間での実測データの再利用性について評価を行なった。

(2) **先進計算機環境への適用の成果**：GPU やメニーコア環境における数値計算コードの最適化に関する研究を進め、ATDL 方式の開発に必要な最新の GPU やメニーコア環境における最適化と、性能評価のためのコードやデータを集めた。また、具体的な ATDL 方式開発のための検討を行った。具体的には、以下のとおりである。

ディープラーニングを含む機械学習・AI 処理の多くにおいては、従来の科学技術計算（高性能並列数値計算）にて性能向上が求められてきた単一の大規模な数値計算処理ではなく、多数の小規模な数値計算を高速に行う要求が大きいことが知られている。そこで本研究では、多数の小規模な数値計算を高速に行う方法を検討し、主に GPU 環境に向けた実装と性能評価を行った。

多数の小規模な数値計算を高速に行いたい需要は機械学習分野以外にも幾つかの数値計算問題・科学技術計算にて需要がある。それらの需要がある研究者・研究グループと連携して研究を進めた。

例えばこの計算方法は近年注目されている階層型行列計算法においても要求が大きいことから、階層型行列計算法を扱う研究グループとも連携して研究を実施し、その成果を国際会議などで発表した。また精度保証計算においても一部の計算を高速化できる可能性を明らかにし、計算手法を適用して成果を研究会等で発表した。

(3) **通信方式の最適化の成果**：AT 方式の評価対象のコードとして、「京」コンピュータ／Fujitsu PRIMEHPC FX100 に搭載されている機能である RDMA を用いノード間のデータ通信の高速化を実現する方式を提案した。

さらにアプリ―ションレベル（共役勾配法）での性能評価を行い、有効性を評価した。

５．主な発表論文等

〔雑誌論文〕　計5件（うち査読付論文　2件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 大島聡史，鈴木惣一朗，坂下達哉，荻野正雄，片桐孝洋，安藤嘉倫	４．巻 2018-HPC-166
２．論文標題 512bit SIMD環境における分子動力学アプリケーションMODYLASの性能評価	５．発行年 2018年
３．雑誌名 研究報告ハイパフォーマンスコンピューティング（HPC）	６．最初と最後の頁 1-9
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Satoshi Ohshima, Soichiro Suzuki, Tatsuya Sakashita, Masao Ogino, Takahiro Katagiri, Yoshimichi Andoh	４．巻 -
２．論文標題 Performance evaluation of the MODYLAS application on modern multi-core and many-core environments	５．発行年 2019年
３．雑誌名 Proc. of IEEE IPDPSW2019	６．最初と最後の頁 787-796
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 范 谷瑛，関 直人，多部田 敏樹，藤井 昭宏，田中 輝雄	４．巻 2019-HPC-168
２．論文標題 ソフトウェア自動チューニングにおける反復2次元d-Spline探索法の提案と評価	５．発行年 2019年
３．雑誌名 研究報告ハイパフォーマンスコンピューティング（HPC）	６．最初と最後の頁 1-8
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 Satoshi Ohshima, Ichitaro Yamazaki, Akihiro Ida, Rio Yokota	４．巻 -
２．論文標題 Optimization of Numerous Small Dense-Matrix-Vector Multiplications in H-matrix Arithmetic on GPU	５．発行年 2019年
３．雑誌名 2019 IEEE 13th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC)	６．最初と最後の頁 9-16
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/MCSoC.2019.00009	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

１．著者名 関直人，范谷瑛，多部田敏樹，藤井昭宏， 田中輝雄	４．巻 2019-HPC-169
２．論文標題 性能パラメータ推定における評価対象プログラムの実行時間の揺らぎに対応した自動チューニング手法の提案	５．発行年 2019年
３．雑誌名 研究報告ハイパフォーマンスコンピューティング（HPC）	６．最初と最後の頁 1-8
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

〔学会発表〕　計22件（うち招待講演　3件／うち国際学会　11件）

１．発表者名 Satoshi Ohshima, Ichitaro Yamazaki, Akihiro Ida, Rio Yokota
２．発表標題 Optimization of hierarchical matrix computation on GPU --- accelerating many small matrix calculation
３．学会等名 Sapporo Summer HPC Seminar 2018（国際学会）
４．発表年 2018年

１．発表者名 大島聡史，藤井昭宏，田中輝雄，深谷猛，須田礼仁
２．発表標題 マルチコア・メニーコア計算機環境における Chebyshev基底通信削減CG法の性能評価
３．学会等名 日本応用数理学会2018年 年会
４．発表年 2018年

１．発表者名 Teruo Tanaka, Fan Guuing, Akihiro Fujii, Takahiro Katagiri
２．発表標題 Enhancement of Performance Parameter Search Method for Multiple Parameter Estimation
３．学会等名 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing (ATAT in HPSC 2019)（国際学会）
４．発表年 2019年

１．発表者名
多部田敏樹，田中輝雄，藤井昭宏，関　直人，范　谷瑛

２．発表標題
ソフトウェア自動チューニングにおける複数性能パラメータを同時推定する手法の効率化

３．学会等名
第81回情報処理学会全国大会

４．発表年
2019年

１．発表者名
出蔵英真，藤井昭宏，田中輝雄

２．発表標題
共役勾配法におけるダブルバッファリング利用したRDMA通信の性 能評価

３．学会等名
第81回情報処理学会全国大会

４．発表年
2019年

１．発表者名
Takahiro Katagiri

２．発表標題
Toward Auto-tuning of Preconditioners for Sparse Iterative Solvers by Deep Learning

３．学会等名
2019 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing (ATAT2019)（国際学会）

４．発表年
2019年

１．発表者名
Takahiro Katagiri, Kenya Yamada

２．発表標題
Auto-tuning of Preconditioners with Deep Learning

３．学会等名
SIAM Conference on Computational Science and Engineering (CSE19)（国際学会）

４．発表年
2019年

１．発表者名
Satoshi Ohshima

２．発表標題
Trying to accelerate many small BLAS calculations on GPU

３．学会等名
ATAT in HPSC (2019 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing)（招待講演）（国際学会）

４．発表年
2019年

１．発表者名
Fumiya Ishiguro, Takahiro Katagiri, Satoshi Ohshima, Toru Nagai

２．発表標題
Performance Evaluation of Accurate Matrix-matrix Multiplications on GPU Using Sparse Matrix Multiplications

３．学会等名
International Conference on High Performance Computing in Asia-Pacific Region (HPCAsia2020)（国際学会）

４．発表年
2020年

１．発表者名
Toma Sakurai, Takahiro Katagiri, Satoshi Ohshima, Toru Nagai

２．発表標題
Autotuning by Changing Directives and Number of Threads in OpenMP using ppOpen-AT

３．学会等名
International Conference on High Performance Computing in Asia-Pacific Region (HPCAsia2020)（国際学会）（国際学会）

４．発表年
2020年

１．発表者名
Takahiro Katagiri

２．発表標題
Towards Auto-tuning Technology in Exascale Era

３．学会等名
CANDAR'19 (The Seventh International Symposium on Computing and Networking)（招待講演）（国際学会）

４．発表年
2020年

１．発表者名
森下誠，大島聡史，片桐孝洋，永井亨

２．発表標題
OpenACCを用いたGKVベンチマークの並列化

３．学会等名
情報処理学会第８２回全国大会

４．発表年
2020年

１．発表者名
山梨祥平，大島聡史，片桐孝洋，永井亨

２．発表標題
外乱のある環境での分散深層学習の性能評価

３．学会等名
情報処理学会第８２回全国大会

４．発表年
2020年

１．発表者名
杉浦拓未，大島聡史，中島大地，片桐孝洋，横田達也，本谷秀堅，永井亨

２．発表標題
医用画像処理におけるLDDMMのGPU高速化

３．学会等名
情報処理学会第８２回全国大会

４．発表年
2020年

１．発表者名
片桐孝洋、櫻井刀麻

２．発表標題
ポストムーア時代に向けた自動チューニングに向けて　～スレッド数の動的最適化

３．学会等名
第24回計算工学講演会

４．発表年
2019年

１．発表者名
片桐孝洋

２．発表標題
ポストムーア時代の計算機環境における数値計算カーネル実装の

３．学会等名
2019年並列／分散／協調処理に関する『北見』サマー・ワークショップ（SWoPP2019）

４．発表年
2019年

１．発表者名
片桐孝洋

２．発表標題
相対的高メモリバンド幅環境における密行列固有値ソルバの実装方式について

３．学会等名
日本応用数理学会2019年度年会

４．発表年
2019年

１．発表者名
Hayate Hasegawa, Masao Ogino, Takahiro Katagiri

２．発表標題
Initial particle distribution based on the centroidal Voronoi tessellation for two-dimensional particle method

３．学会等名
The 7th Asia-Pacific Congress on Computational MechanicsProgram (APCOM2019)（国際学会）

４．発表年
2019年

１．発表者名
北澤修太，沼波政倫，大谷寛明，片桐孝洋，大島聡史，永井亨

２．発表標題
Windows MR＋Unityの環境におけるプラズマ乱流シミュレーションの可視化

３．学会等名
先進的可視化環境を用いた可視化情報の研究会(VR2019)

４．発表年
2019年

１．発表者名
Naoto seki, Toshiki Tabeta, Akihiro Fujii, Teruo Tanaka

２．発表標題
Stable Automatic Tuning Method for Performance Fluctuation

３．学会等名
2020 SIAM Conference on Parallel Processing for Scientific Computing (PP20)（国際学会）

４．発表年
2020年

１．発表者名
Toshiki Tabeta, Naoto Seki, Akihiro Fujii, Teruo Tanaka, Hiroyuki Takizawa

２．発表標題
An Optimization technology of Software Auto-Tuning Applied to Machine Learning Software

３．学会等名
International Conference on High Performance Computing in Asia-Pacific Region (HPCAsia2020)（国際学会）

４．発表年
2020年

１．発表者名
関直人

２．発表標題
ppOpen-ATによる自動チューニングの実行

３．学会等名
第20回AT研究会オープンアカデミックセッション（ATOS20）（招待講演）

４．発表年
2020年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

―

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	大島　聡史 (Satoshi Ohshima) (40570081)	名古屋大学・情報基盤センター・准教授 (13901)	

６．研究組織（つづき）

	氏名 （研究者番号）	所属研究機関・部局・職 （機関番号）	備考
研究分担者	田中　輝雄 (Tanaka Teruo) (90622837)	工学院大学・情報学部（情報工学部）・教授 (32613)	