

平成 30 年 5 月 31 日現在

機関番号：12601

研究種目：基盤研究(B)（一般）

研究期間：2015～2017

課題番号：15H02708

研究課題名（和文）複合的・階層的な自動チューニングを実現する数理基盤手法の研究とライブラリの開発

研究課題名（英文）Research on Mathematical Methods and Development of Libraries for Combined and Hierarchical Autotuning

研究代表者

須田 礼仁（SUDA, Reiji）

東京大学・大学院情報理工学系研究科・教授

研究者番号：40251392

交付決定額（研究期間全体）：（直接経費） 14,000,000円

研究成果の概要（和文）：自動チューニングは、ソフトウェアが内包するパラメタを自ら調整し、多様な条件下で良好な性能を達成することを目指す技術である。従来、複数のパラメタの調整が必要な場合、網羅試行か経験的枝刈りが広く用いられてきたが、本研究では、ベイズ統計に基づき、現実的に有効かつ漸近的に最適解を導く数理的手法を目指した。

従来研究の調査により、線形モデルと相関モデルが使われており、両者は同時に使うこともできることを示した。そのようなモデルを記述から、事前情報と実測結果から性能モデルを構築するプログラムを自動生成する仕組みを構築した。また、自動チューニング数理ライブラリの多様な計算に適用し、その有効性を確認した。

研究成果の概要（英文）：Autotuning is technology that aims to attain good performance under various conditions, by letting software controls its own parameter. In case multiple parameter exist, most of previous research chose either exhaustive search or heuristic pruning. In this research, we aim mathematically founded method using Bayesian statistics, which gives practically good and asymptotically optimal solutions.

In survey of previous works, we found that linear models and correlation models have such properties, and they can be combined. From description of such models, we create a software that generates a code that constructs performance model from a priori information and observations. Also we apply autotuning mathematical libraries to various computations, and confirms their effectiveness.

研究分野：高性能並列数値計算

キーワード：自動チューニング 複合的自動チューニング ベイズ統計 実験計画 線形モデル 相関モデル モデル
 ルフィッティング コード生成

1. 研究開始当初の背景

自動チューニングは、ソフトウェアが内包するパラメタを自ら調整し、多様な条件下で良好な性能を達成することを目指す技術である。パラメタには、変数のような簡単なものから、関数や手続きの選択、データ構造やアルゴリズムの選択など、性能に影響を及ぼすあらゆるものが含まれる。

従来の自動チューニング研究では、複数のパラメタを同時にチューニングする必要がある場合、(1) パラメタの組合せを網羅的に試す、(2) 経験的に試す組合せを絞り込む、のいずれかが用いられてきた。(1) ではパラメタ数が増加すると非現実的になり、(2) では最適な性能を与えるパラメタの組合せを見逃す可能性がある。

また、自動チューニング機能を有するライブラリプログラムなどを、自動チューニング機能を有するアプリケーションソフトウェアなどが利用すると、階層的な自動チューニングが発生する。このような場合に適切な自動チューニングの手法については従来研究には見るべきものがなかった。

2. 研究の目的

本研究では、このような自動チューニングを複合的・階層的自動チューニングと呼ぶ。そして、複合的・階層的な自動チューニングにおいて、効率的かつ確実に、良好なパラメタの調整を達成する手法、具体的には、数理モデル化手法および最適化手法を開発することを目的とした。

対象とするアルゴリズムとして、今後さらに重要性を増してゆく、メニーコアや大規模並列数値計算を主なターゲットとした。そして、提案手法を「自動チューニング数理ライブラリ」として実装・公開し、広く高性能計算の発展に寄与することを目指した。

3. 研究の方法

本研究では、3つのサブ課題を設定して上記の研究に取り組むこととした。

(1) 複合的・階層的な自動チューニングに対する確かな数理的基盤を与える数理的手法の開発。

(2) 上記サブ課題で開発した手法をライブラリとして実装。

(3) 上記サブ課題で開発したライブラリを高性能並列計算に適用。

4. 研究成果

(1) 本研究では、まず、複合的・階層的な自動チューニングに相当するような事例を調査し、それらを可能な限り一般化して考えた時に、どのような状況がありうるのかを分析した。その結果、当初の予想以上に複雑な状況がありうるが見いだされた。上述のように「ライブラリ」とそれを利用する「アプ

リケーション」という階層があったとしても、キャッシュなど内部状態への影響のため、各階層が他方の階層に与える影響は、ソフトウェアの階層よりも複雑なものになりうる。また、これまではチューニングの対象となるソフトウェアの実行は単純に開始して終了し、かつ、終了から次の開始までの間にパラメタを変更できるものと仮定していたが、複合的なソフトウェアではより複雑な事態が考えられ、パラメタの変更が制限される可能性もあることが明らかとなった。これらの考察から、一般性のある階層的な自動チューニングの精密な数理モデル化はかなり複雑なものとなり、本研究課題が当初想定していた研究範囲を超える難しい課題であることが明らかとなった。そこで、本研究では、複数のパラメタがチューニングするソフトウェアの部分かどのような関係であるかは無視して、ソフトウェア全体を1まとまりとして扱い、複数のパラメタをどのようにチューニングするかという問題に集中することとした。これにより、一般性は失われないが、ソフトウェアの構造を自動チューニングに活用することができないため、効率の面で制限がかかることになる。

(2) 次に、複数のパラメタを持つ複合的自動チューニングにおける様々な従来研究における性能モデル化について調査を行った。特に我々はベイズ的性能モデルを用いることで、開発者が持っている知識を事前情報として組み込み、効率的な最適化を可能にするとともに、測定された性能をベイズ的に分析することで、事前情報の精度の良しあしに関わらず、最適なパラメタを選択できる手法を開発してきた。このような観点に沿うような従来の自動チューニングの性能モデルを分析した結果、「線形モデル」と「相関モデル」の2種類に大きく分けられることがわかった。

線形モデルは、性能の平均値をパラメタの関数としてモデル化し、モデルの中に含まれる未定係数を測定データから推定する手法である。パラメタと性能の関連性がある程度わかっている場合には、それに応じて「計画行列」を適切に設定することで効率的にモデル化できるし、そのような事前知識がない場合も、パラメタの主効果をとそれ以外に分割するANOVA型のモデルが役に立っていることがわかった。

相関モデルは、パラメタが近いと性能が近いという仮説に基づき、近いパラメタでの性能から未知のパラメタでの性能を推定する方法である。この種の汎用的なモデルであるガウス過程モデルや、Krigingモデルが自動チューニングに使われたことがある。

従来研究の分析からこれら2種類のモデルの有用性が確認されたが、さらに検討した結果、これらの2種類の性能モデルは相反するものではなく、組み合わせることでできることも明らかとなった。また、これらの

性能モデルを設定したときに、測定された性能からパラメタをフィッティングする手法についても明らかにした。

(3) 次に、これらの知見を基に自動チューニングのための数理基盤ソフトウェアの構築を行った。上記の線形モデルと相関モデルの組合せを考えると、極めて多様なモデルが構築できる。この多様性を維持したままで汎用性のあるライブラリにすると、汎用性のために性能を犠牲にすることになる危険性が大きいことが危惧された。自動チューニングライブラリの性能が低いためにチューニングされるソフトウェアの性能が劣化しては意味がない。そこで、モデルを記述したプログラムから、そのモデルに適した自動チューニング用のプログラムを自動生成するという方針を取ることにした。我々は滝沢らが開発している Xevolver というコード変換システムを利用して、モデルの記述から自動チューニングのための性能推定を行う数理ソフトウェアを自動生成することを行った。これはかなり複雑なコード変換を必要とし、我々が分析した線形モデルと相関モデルのすべてをカバーできるだけのものには到達できなかったが、一定の問題について適切に性能モデル化ができることを実証した。しかし完成度は十分ではなく、公開には至らなかった。

これと並行して、従来から研究を進めている d-spline による性能モデル化を複数パラメタの場合に拡張する研究を行った。単純には多次元 d-spline を用いてモデル化することになる。2 パラメタの場合には効率的なアルゴリズムを構築できたが、3 パラメタ以上では計算量がかなり大きくなってしまふことが確認された。また、2 パラメタの場合について詳しく調査したところ、探索も大きな計算量がかかることが判明したので、1次元化した探索を反復することで2パラメタの探索を効率化する手法を提案、実装して良好な性能を得た。また、従来は d-spline モデルでは性能のばらつきを考慮していなかったが、測定した性能にばらつきがあっても安定して最適化できる手法を開発した。

また、チームメンバーが開発している自動チューニングフレームワークである ppOpen-AT において階層的な自動チューニングを組み込むことを行った。

(4) また、自動チューニングの数理手法および数理ライブラリを実アプリケーションに適用することを目指した。我々が開発している自動チューニング数理ライブラリ ATMathCoreLib を、マルチコア計算機を用いた特異値分解に適用し、ユーザが性能特性を把握している場合、把握していない場合、アルゴリズムと性能パラメタの同時選択を行う場合の3つの累計を考察して、適切な計画行列の選択を提案し、また実験により自動チューニングの効果を検証した。また一般固有

値問題アルゴリズムに意図的に実行時間の誤差を加え、40% ものばらつきを与えても ATMathCoreLib が実行時間最短の手法を選択することを確認した。また Xeon Phi 上での三角行列の積にも ATMathCoreLib を適用し、高い性能が達成できることを示した。

また、AMG (代数的マルチグリッド)法における階層的な性能パラメタと、それに伴う性能の挙動を分析し、階層的・複合的な自動チューニングの必要性が確認された。ただしこの事例では、あるパラメタの設定値によって、新たなパラメタが生じるなど、本研究で検討してきたよりもさらに複雑な状況であることもわかった。ひとまずモデルを無視して100万とおりのパラメタ値を我々が開発している自動チューニングライブラリ ATMathCoreLib で最適化させ、自動チューニングの効果が得られることを確認した。また、変動させるパラメタを制約することで、d-spline による自動チューニングを適用することを可能にし、良好な性能を達成した。

ただし、今回開発した新しい手法はまだ ATMathCoreLib に組み込まれておらず、これは今後の課題として残された。

(5) また、自動チューニングおよび高性能計算の新たな地平線を開拓する研究を行った。

GA を用いた連立一次方程式解法 GA-BiCG における自動チューニングを検討した。このような反復法においては、反復が終了しないという場合が発生する。これは従来のベイズ性能モデルでは適切に表現することができないものであった。そこで、新たに反復が収束する確率をモデルの中に組み込み、これを考慮した新しい性能のモデル化と、ベイズ統計を用いた自動チューニングの実装を行い、収束確率が高く、反復回数が少ないパラメタを効率的に見つけ出すことに成功した。

また、このほかに大規模並列高性能アルゴリズムの研究も行った。通信削減型の連立一次方程式の解法である CBCG 法において、疎行列・ベクトル積の通信削減手法を提案して有効性を実証した。またブロック赤黒順序付けにより並列化した MILU(0) 前処理において、著名なライブラリである MAGMA を越える性能を達成した。頻出アイテム集合を発見する LCM アルゴリズムの大規模並列化、その応用として多重検定アルゴリズム LAMP の並列化や遺伝子関連解析 GWAS の実装、制約充足問題の大規模並列化、RNA inverse folding や材料科学へのモンテカルロ木探索の応用など、数値計算以外の大規模計算の並列化に関して目覚ましい成果を挙げた。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文](計 4 件)

Yosuke Kumagai, Akihiro Fujii, Teruo Tanaka, Yusuke Hirota, Takeshi Fukaya, Toshiyuki Imamura, Reiji Suda, Performance Analysis of the Chebyshev Basis Conjugate Gradient Method on the K Computer, Lecture Notes in Computer Science, 査読有, 9573, 2016, 74-85. https://doi.org/10.1007/978-3-319-32149-3_8

工藤周平、山本有作、Xeon Phi における DSYRK の並列化手法と性能解析、情報処理学会論文誌コンピューティングシステム、査読有、9(4), 2016, 15-24.

熊谷洋祐、藤井昭宏、田中輝雄、深谷猛、須田礼仁、共役勾配法への種々の通信削減手法の適用と評価、情報処理学会論文誌コンピューティングシステム、査読有、9(3), 2017, 1-13.

Reiji Suda, Cong Li, Daichi Watanabe, Yosuke Kumagai, Akihiro Fujii, Teruo Tanaka, Communication-Avoiding CG Method: New Direction of Krylov Subspace Methods towards Exa-scale Computing, 京都大学数理解析研究所講義録、査読無、1995, 2016, 102-111.

〔学会発表〕(計 9 件)

村田陸、入江純、藤井昭宏、田中輝雄、片桐孝洋、二次元パラメタ空間への標本点逐次追加型性能パラメタ推定法の拡張、日本応用数学会 2015 年度年会、2015 年 9 月 11 日、金沢大学(石川県金沢市)

Seiji Nagashima, Takeshi Fukaya, Yusaku Yamamoto, On constructing cost models for online automatic tuning using ATMathCoreLib, IEEE MCSoc-16, 2016 年 9 月 21 日、リヨン(フランス)

田中輝雄、望月大義、村田陸、藤井昭宏、複数性能パラメタ空間における実行時 AT 機構、2016 年ハイパフォーマンスコンピューティングと計算科学シンポジウム、2016 年 6 月 6 日、東北大学(宮城県仙台市)

須田礼仁、複合的自動チューニングのための数理とソフトウェア、第 21 回計算工学講演会、2016 年 5 月 31 日、朱鷺メッセ(新潟県新潟市)

Reiji SUDA, Embedded-DSL-like code generation and optimization of Bayesian estimation routines with user-defined source-to-source code

transformation framework Xevolver, International Workshop on Legacy HPC Application Migration, 2017 年 11 月 20 日、ASPAM(青森県青森市)

須田礼仁、複合的自動チューニングのための数理ライブラリの構築、第 22 回計算工学講演会、2017 年 5 月 31 日、ソニックシティ(埼玉県大宮市)

山本堅太郎、松本正晴、須田礼仁、遺伝的アルゴリズムに基づく疎行列解法のパラメタに関するオンライン自動チューニング、情報処理学会第 159 回 HPC 研究会、2017 年 4 月 17 日、東京大学(千葉県柏市)

Guqing Fan, Masayoshi Mochizuki, Akihiro Fujii, Teruo Tanaka, D-Spline Performance Tuning Method Flexibly Responsive to Execution Time Perturbation, PPAM2017, 2017 年 9 月 10 日、ルブリン(ポーランド)

Masayoshi Mochizuki, Akihiro Fujii, Teruo Tanaka, Takahiro Katagiri, Fast Multidimensional Performance Parameter Estimation with Multiple One-Dimensional d-spline Parameter Search, International Workshop on Automatic Performance Tuning (iWAPT2017), 2017 年 6 月 2 日、オーランド(米国)

〔図書〕(計 0 件)
なし

〔産業財産権〕
なし

〔その他〕
国際研究集会(共催)

- 2016 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing (ATAT2016), 2016/2/19-20, 国立台湾大学(台湾)
- 2018 Conference on Advanced Topics and Auto Tuning in High-Performance Scientific Computing (ATAT2018), 2018/3/26-27, 国立成功大学(台湾)

6. 研究組織

(1) 研究代表者

須田 礼仁(SUDA, Reiji)

東京大学・大学院情報理工学系研究科・教授

研究者番号: 40251392

(2) 研究分担者

藤井 昭宏(FUJII, Akihiro)

工学院大学・情報工学部・准教授
研究者番号：10383986

美添 一樹 (YOSHIZOE, Kazuki)
国立研究開発法人理化学研究所・革新知能
統合研究センター・ユニットリーダー
研究者番号：80449115

(3)連携研究者

田中 輝雄 (TANAKA, Teruo)
工学院大学・情報工学部・教授
研究者番号：90622837

山本 有作 (YAMAMOTO, Yusaku)
電気通信大学・情報理工学研究科・教授
研究者番号：20362288

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