

jh220044

## 機械学習ソフトウェアへのソフトウェア自動チューニング技術の適用 田中輝雄（工学院大学）

### 概要

ソフトウェア自動チューニング(AT)におけるユーザプログラムの性能を決定する性能パラメタの最適組合せ探索に関する研究である。本研究では、機械学習ソフトウェアのハイパーパラメタの最適化をATで行う。題材として、(a)ロボット制御における人移動予測プログラム、(b)自然な画像拡大を実現する超解像プログラムを用いる。機械学習では膨大な学習時間を要する。これに対し、スーパーコンピュータの多数GPUを有効に用いた多重実行を制御するATツールを開発した。名古屋大学の不老TypeIIを用いて実行し、題材(a)に対しては、学習済みの類似データを用いることで、更なるATの実行時間の短縮および学習結果の精度の向上を実現した。題材(b)については、2段階学習方式と並列処理により、実用範囲のATとなる100-200倍の高速化を実現した。また、従来専門家が時間をかけて行った場合と同レベルの学習結果を得ることができた。

### 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  
名古屋大学 情報基盤センター

(2) 課題分野  
大規模計算科学課題分野

(3) 共同研究分野  
超大規模数値計算系応用分野

(4) 参加研究者の役割分担  
田中 輝雄（工学院大学・情報学部）  
・取り纏め、自動チューニング手法  
大島 聡史（名古屋大学・情報基盤センタ）  
・GPU, メモリア向け, アルゴリズム  
藤井 昭宏（工学院大学・情報学部）  
・並列処理アルゴリズム  
加藤 由花（東京女子大学・現代教養学部）  
・機械学習プログラム・データ提供  
藤家空太郎（工学院大学・大学院情報学専攻）  
・自動チューニングソフトウェア開発, 評価  
ヨウ ケン（工学院大学・大学院情報学専攻）  
・自動チューニングソフトウェア開発, 評価

### 2. 研究の目的と意義

我々は、多様な計算機環境での高性能化を実現する技術として、対象とするプログラム性能を決定する複数の性能パラメタを、キャッシュサイズやデータ通信性能等の計算機の特性に、自動的に最適化するソフトウェア自動チューニング(以下、ATと記載)の研究を進め、その手法を提案、実証している[a]。

現在、機械学習プログラムをターゲットに研究を進めており、その機械学習プログラムのハイパーパラメタに対してATを実施し、予測モデルの精度向上および実行時間の高速化を図る。この機械学習は1回の学習は30分から数時間かかる。ハイパーパラメタの組合せパターンは数万通りとなり、我々の開発したAT機構でもチューニングに数日かかる。これに対して、並列化によるハイパーパラメタの組合せ探索時間の削減を目指す。

対象とするアプリケーションは、東京女子大の加藤グループによる「人を回避しながら動くロボットの制御」に向けた人移動予測AI[b](以降、歩行者経路予測アプリ)である。対象とするシステムはPhysical worldとCyber

World(クラウド)から構成される。Physical world では、ロボットが予測モデルを使い周囲の人移動を予測し動作する。Physical world で得たセンシング情報を Cyber World に送り、予測モデルを機械学習により随時更新する。

もうひとつの応用先として、豊田工業大学の浮田グループによる機械学習を用いた超解像プログラムに適用する[c]。このプログラムでは、画像の画質を PI (Perceptual Index) と RMSE (Root Mean Square Error) で評価する。PI は人が見て綺麗かどうかを表す指標である。単に PI だけでは、元の画像の情報がじゅうぶんに取り込めないで、すべての画像のピクセル情報の平均二乗誤差 (MSE) の平方根である RMSE を用いて評価範囲を設定し、その範囲内での PI の最適化を図る。対象とする機械学習プログラムは 1 回の学習に 9 時間ほど要し、学習プログラムを何度も繰り返し実行する AT はそのままでの実現は不可能であり、この対策も必要となる。

本研究は、スーパーコンピュータを効果的に利用することにより、応用研究の専門家が時間をかけて行ってきた(学習)モデルの最適化を自動化し、応用研究者がモデル自体の研究開発に注力できるようにすることに意義がある。

[a]T. Tanaka, R. Otsuka, A. Fujii, T. Katagiri, T. Imamura, Implementation of d-Spline-based Incremental Performance Parameter Estimation Method with ppOpen-AT, Scientific Programming, Vol. 22, pp. 299-307, 2014.

[b]R. Akabane and Y. Kato, Pedestrian Trajectory Prediction Using Pre-trained Machine Learning Model for Human-Following Mobile Robot, IEEE International Conference on Big Data Workshop (IoTDA 2020), pp. 3453-3458, 2020.

[c]M. Haris, G. Shakhnarovich, N. Ukita, Deep Back-Project Networks for Single Image Super-Resolution, in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 43, no. 12, pp. 4323-4337, 2021.

### 3. 当拠点の公募型研究として実施した意義

本研究課題で対象とする機械学習は膨大な学習時間を要する。AT では、この計算量に対してハイパーパラメタを変更しながら、応用プログラムを何度も実行する必要がある。また、機械学習においては、その計算の特性から GPU が有効である。そのために、多数の GPU を有するスーパーコンピュータである名古屋大学情報基盤センターの不老 Type II (以下、不老 Type II) を用いる。

本研究では、この不老 Type II の持つ大規模 GPU 並列環境を有効に利用する手段を提案し、実証する。

さらに、本研究では応用研究の専門家（ここでは、ロボット工学者あるいは高解像画像処理技術者）と高性能計算学者との学際研究として推進し、HPC 及びスーパーコンピュータの利用技術を実証することで、広く開発技術を提供、普及させていく。

### 4. 前年度までに得られた研究成果の概要

ロボット制御における人移動の予測モデルを、機械学習を用いてより精度の高い予測モデルにセンシングデータを追加し順次修正する。このときの機械学習のハイパーパラメタの最適化を AT で行う。機械学習では膨大な学習時間を要し、さらに大量のデータの利用にランダム性があり、結果が一意に定まらない。

以上のような課題に対し、

(1) 不老 Type II の持つ多数の GPU を有効に用いた多重実行手段を構築した。並列化により、AT のハイパーパラメタ探索時間を逐次実行に比べて 13-23 倍高速化した。

(2) 機械学習で生じる揺らぎに対応するため、得られた上位複数の学習結果を再度実行し、学習結果の安定性を確保した。その結果、評価指標である人移動の最終位置の誤差に対し、従来の値から 29-38% 改善した学習モデルを構築できた。

## 5. 今年度の研究成果の詳細

### (1) 概要

今年度は機械学習における以下の2つの適用先について評価を行った。

- (a) ロボット制御における「人を回避しながら動くロボット制御」に対する歩行者移動予測 AI プログラム、
- (b) 見た目が自然な画像拡大を実現する超解像プログラム。

機械学習では膨大な学習時間を要する。これに対し、スーパーコンの多数 GPU を有効に用いた多重実行を制御する AT ツールを開発した。

名古屋大学 不老 TypeII を用いて実測を行い、題材(a)に対しては、学習済みの類似データを用いることで、2021 年度に対してさらなる AT の実行時間の短縮および結果の精度の向上を実現した。題材(b)については、実用範囲での AT を実現するために、2段階学習方式と並列処理により 100-200 倍の高速化を実現し、さらに、従来専門家が時間をかけて行う場合と同レベルの結果を得ることができた。

### (2) 題材(a) : 歩行者移動予測 AI プログラムにおける学習済みデータの利用

歩行者移動予測 AI プログラムでは、学習データとして、いろいろな環境データを用いる。その際、類似の環境データ（つまり、類似の歩行者移動環境）では、近い学習モデルが得られると推察される。したがって、類似のモデルで学習した結果を用いて評価を行うことにより、新しい環境データに対する学習においても学習時間の短縮、あるいは、高い精度の学習などが期待できると考える。そこで、実験を行い、学習時のハイパーパラメータを初期値の設定を①従来の学習データで得られた結果と、②類似の環境データで学習した結果を比較した。具体的には、ハイパーパラメータの初期値として、①「従来」は経験的に良いとされている値、②「類似」は環境データ (Stanford Drone Dataset/hyang-4) で学習したハイパーパラメータを用いた。

表 1 初期ハイパーパラメータを変えたときの実測結果

ハイパーパラメータの初期値	①従来	②類似
FDE (m)	0.39	0.38
AT 実行時間 (h)	89	51
実測回数 (回)	236	183

- ・ここで、AT の対象とした環境データは (Stanford Drone Dataset/hyang-5)
- ・AT 実行前の FDE 値は 0.6[m] (FDE: Final Displacement Error 評価尺度であり、実際の歩行者到着地点と予測器の予測地点の誤差を表す)

初期パラメータを変えることで我々の開発した AT ツールの推定結果がどう変化するかを表 1 に示す。ここでのチューニング対象は、初期パラメータ推定時に使用したデータセットと同じマップ構造だが、学習データとして別の歩行者の移動軌跡が記録されているシーン “hyang 5” を用いる。

実測の結果、FDE 値は、類似の推定した初期パラメータを用いた自動チューニングでは、従来の経験的な初期パラメータを用いた場合と比較して、実行時間が 57%、実測回数が 78% に減少した。FDE 値はチューニング前に比べていずれも約 0.2m 短縮された。この結果から、同一シーンの異なるシーケンスを自動チューニングする場合、推定したハイパーパラメータを初期パラメータとして用いた自動チューニングは、予想通り実行時間と実測回数を削減することができた。

### (3) 超解像への適用

今回、新たに、超解像の機械学習プログラムを対象とした。超解像は、低解像度画像を高解像度画像に変換する技術である。本研究で用いた機械学習を用いた超解像度プログラムは、

豊田工業大学の浮田グループが開発した Dense Deep Back-Projection Networks (D-DBPN) である。ここでは、2 つの性能値 Perceptual Index (PI) と Root Mean Square Error (RMSE) を使用して、結果の画像を評価する。学習モデルの評価のために、異なる RMSE の範囲を設定し、範囲内の最小の PI 値をそれぞれ調べる。

この学習プログラムは1回あたりの学習時間が長く、2000 エポックで9時間を要する。自動チューニングでは学習を繰り返し行う必要がある。したがって、このまま自動チューニングを行うことは不可能であり、大幅な実行時間の削減が必要となる。

本研究の目的は、自動チューニングの時間を大幅に削減することで、実用的な時間内に超解像度の機械学習プログラムのハイパーパラメータを自動チューニングすることができるようにすることである。そのために、(1)学習を事前学習とファインチューニング2段階学習とする手法、(2)複数回の学習を効率的に行うための GPU クラスタ上で複数ジョブの多重実行制御手法を提案し、実装する。

### (3-1) 2段階学習

1回の学習時間を大幅に削減するために、事前学習をした結果を用いることで毎回の学習時間を抑えることを試みる。そのために、エポック数 2000 回の途中の G\_epoch\_loss (毎回のエポックの Generator loss) の挙動を調査し、1000 回で安定していること、また、1000 回学習し、さらに 250 回の学習を行うことで、安定した画像が得られることを確認した。そこで、まず一度だけ1000回の学習を行い(事前学習)、それを AT では毎回 250 回の学習(ファインチューニング)の2段階で行うこととした。毎回 2000 回の学習を行う場合に比べて 1/8 の回数となる。

### (3-2) 並列実行

スーパーコンピュータの多数の GPU ノードを有効に活用するために、AT で必要となる複数の学習プログラムの実行を同時に行い、それらの

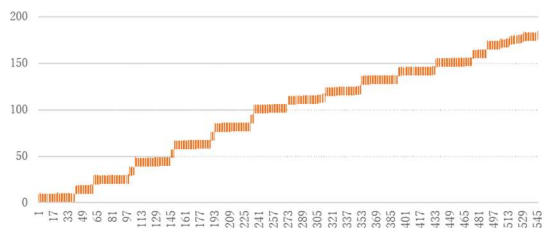


図1 並列実行状況

結果を分析し、さらに AT による最適パラメータの探索を進める並列実行環境を構築し、適用した。並列実行状況を図1に示す。

図1において、縦軸は実行時間、横軸は AT により実行された各ジョブである。縦方向で、ひとつひとつのジョブの開始時間と終了時間を示している。AT ツールで必要な数の学習プログラムが順次評価されながら実行される。その際、実行するスーパーコンピュータシステムの運用で使用可能な最大数のジョブを同時に実行する。この「使用可能な最大数」は、システムにより異なり、また、システム運用時の利用状況(混雑状況)で変化するが、我々の開発した AT ツールはそれに柔軟に対応することができる。

2段階学習と GPU クラスタ上で複数ジョブの多重実行を組み合わせることにより、今回対象とした超解像プログラムは 1/100~1/300 に短縮された。これにより、不可能と考えられていた AT を数日の規模で実現することができた。

### (3-3) 画像の評価

PI のみを考慮した場合、良好な画像効果が得られるわけではない。より良い画像結果を探索するために、再構成精度 (Reconstruction accuracy) の指標が考慮される。PIRM2018 チャレンジ[d]では、異なる RMSE 範囲を設定し、それぞれの範囲内での最小の PI 値を探索する。それにしたがって、範囲を限定した、つまり、制約条件付きの AT をを実現する。PIRM チャレンジでは RMSE の 3 つの範囲 :  $RMSE < 11.5$ ,  $11.5 < RMSE < 12.5$ ,  $12.5 < RMSE < 16$  をそれぞれ region 1, region 2, region 3 に設定している。今回はその評価に合わせた。

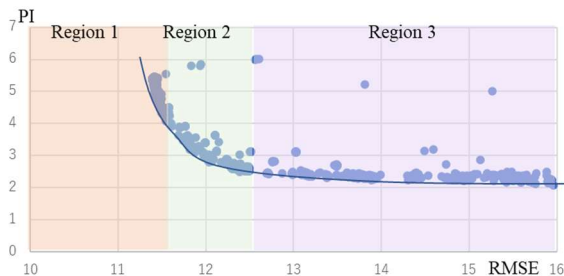


図2 AT 結果の PI と RMSE の関連

ATにおいて、対象とする機械学習プログラムを実行して得られた結果のPIとRMSEの組み合わせを図2に示す。この図でわかるように、ほとんどがRMSEの制限の中でPIが最小となるような組み合わせを選択していることがわかる。

[d] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, L. Zelnik-Manor, The 2018 PIRM challenge on perceptual image super-resolution, Proceedings of the European Conference on Computer Vision (ECCV) 2018.

## 6. 進捗状況の自己評価と今後の展望

特に実行時間を要する機械学習プログラムに適用するために、機械学習プログラムの多重実行制御を可能とするようにATツールを強化した。このATツールを2021年度に実施した歩行者移動予測AIプログラムへの適用に引き続き、2022年度は、さらに学習時間を要する超解像プログラムにも適用し、その効果を示すことができた。

2022年度は、名古屋大学のスーパーコンピュータ不老Type-IIを利用した。我々の開発したATツールは不老Type-IIに固有に開発したのではなく、他のスーパーコンピュータシステムでも汎用的に利用することが可能である。

2023年度は、東京大学および九州大学のスーパーコンピュータシステムを利用することができるため、これらのスーパーコンピュータでも実証実験を行い、我々の開発したATツールの汎用性を示したいと考えている。

## 7. 研究業績

### (1) 学術論文 (査読あり)

無し

### (2) 国際会議プロシーディングス (査読あり)

[1] Sorataro Fujika, Yuga Yajima, Teruo Tanaka, Akihiro Fuji, Yuka Kato, Satoshi Ohshima, T. Katagiri, Parallelization of Automatic Tuning for Hyperparameter Optimization of Pedestrian Route Prediction Applications using Machine Learning, HPC Asia 2023, 2023.3.

### (3) 国際会議発表 (査読なし)

[2] Xuan Yang, Sorataro Fujika, Yuga Yajima, Akihiro Fuji, Teruo Tanaka, Kazutoshi Akita, Norimichi Ukita, Satoshi Ohshima, Applying Automatic Tuning to Hyper-parameter Optimization of Machine Learning Programs for Super-Resolution, poster, HPC Asia 2023, 2023.3

(HPC Asia 2023 BEST STUDENT POSTER AWARD 2nd Placeを受賞)。

[3] Xuan Yang, Sorataro Fujika, Yuga Yajima, Akihiro Fuji, Teruo Tanaka, Kazutoshi Akita, Norimichi Ukita, Satoshi Ohshima, Reduction of Time Spent on Hyperparameters Estimation through Automatic Tuning Parallelized by Multiple Job Execution ITBL Booth Poster, Super Computing 22 (SC22), 2022.11.

### (4) 国内会議発表 (査読なし)

[4] 楊暄, 藤家空太郎, 矢島雄河, 秋田和俊, 藤井昭宏, 田中輝雄, 浮田宗伯, 大島聡史, 超解像のための機械学習プログラムのハイパーパラメタ最適化に対する自動チューニングの適用, 情報処理学会, 研究報告HPC-184-5, 第184回HPC研究会, 2022.5.

### (5) 公開したライブラリなど

無し

### (6) その他 (特許, プレスリリース, 著書等)

無し