

スーパーコンピュータ「富岳」における HPC/AIへの取り組みと将来への展望

忽滑谷 淳史 赤尾 和俊 高橋 仁 福本 尚人 川上 健太郎
黒田 明義 南 一生 佐藤 賢斗 松岡 聡

あらまし

日本の科学技術政策として掲げられているSociety 5.0を実現するためには、AI技術の発展が必要不可欠である。そして、この発展とともに、AIの学習に必要な計算資源は増加し続けている。富士通は、スーパーコンピュータ「京」やスーパーコンピュータ「富岳」（以下、「富岳」）などの開発を通して、豊富な計算資源を持つHPC（High Performance Computing）システムを提供してきた。そして現在、HPCシステムの豊富な計算資源をAIの学習に活用するために、理化学研究所と共同で、「富岳」上にAIインフラを整備する取り組みを行っている。

本稿では、理化学研究所と共同で進めている、「富岳」におけるAI関連のソフトウェアの動作検証と性能評価の状況、およびHPCとAIに向けた今後の取り組みについて述べる。

1. まえがき

2016年、日本の科学技術政策としてSociety 5.0が提唱された[1]。Society 5.0で実現する社会では、AIによってビッグデータを解析することで、産業の高付加価値化を目指している。

AIを社会の中で活用し、Society 5.0を世界に先駆けて実現するためには、AI技術の発展が欠かせない。そして、この発展とともに、AIの学習・推論に必要な計算資源（演算機、メモリーなど）は増える傾向である。このような背景の下、計算資源が豊富なHPC（High Performance Computing）によるAIの活用が進められている。

本稿では、HPCとAIの関係、スーパーコンピュータ「富岳」（以下、「富岳」）で行っているAIに関する取り組みについて説明する。そして最後に、本取り組みの今後の展望について述べる。

2. HPCとAI

本章では、HPCにおけるAIの活用の観点に基づいて、HPCとAIの関係について述べる。

2.1 AI基盤としての「富岳」

Society 5.0を実現するために重要なAI技術の発展には、次の3点による貢献が大きいとされている[2]。

- ・ AIの学習に必要となる大量のデータ
- ・ 高性能なAI基盤
- ・ 高度なAI人材

「富岳」は、大量のデータを高速に処理する高性能なAI基盤を提供することで、Society 5.0の中心的なインフラとなることを目標としている。また、高性能なAI基盤を提供することで、世界から高度なAI人材を集め、AI技術の発展につながることも期待している。

2.2 HPCにおけるAIの活用

本節では、深層学習の高速化やHPC向けアプリケーションへの適用例を示し、HPCにおけるAIの活用について述べる。

(1) 深層学習の高速化

AIが近年急速に発展した主な理由として、深層学習技術の向上がある。深層学習は、従来の機械学習と比較して、より多くの特徴量を学習するため、学習に要する計算量が多い。また、学習に使用するデータも多いため、データを高速に処理する仕組みが必要である。

一方で近年、計算資源が豊富になり、現実的な時間で学習が行えるようになった。これによって、深層学習の活用が広がっている。スーパーコンピュータ「京」（以下、「京」）や「富岳」のCPUは、高い演算性能と並列性能、広いメモリーバンド幅を持つため、深層学習への活用が可能と考える。今後、豊富な計算資源を持つHPCシステムへの深層学習を用いたアプリケーションの適用と更なる高速化が期待されている。

(2) HPC向けアプリケーションへの適用

HPC向けのアプリケーションにAI技術を適用し、Society 5.0を実現する取り組みも始まっている。例えば、深層学習によって衛星画像の高解像度化や雲除去を行い、地表面の状況をより詳細に把握することで、大規模自然災害への対応に役立つと期待されている[3]。また、HPCの高度化のために、深層学習を用いたシステムソフトウェアのチューニングにも適用されている[4]。一方、深層学習自体の精度を上げる手法である、データ拡張の高速化へも活用が期待されている[5]。

HPC向けのアプリケーションの一環としてAIに関する演算を行うという点でも、HPCにおけるAI基盤の整備が重要になる。

3. 「富岳」における深層学習の取り組み

本章では、DL4Fugakuプロジェクトをはじめとした「富岳」における深層学習の取り組みについて述べる。

3.1 DL4Fugakuプロジェクトの立ち上げ

「富岳」のAI基盤としての提供に向けて、AIソフトウェアを高速に動作させることを目的として、DL4Fugakuプロジェクトが立ち上げられた[6]。このプロジェクトは、AIソフトウェアをチューニング

や動作検証によって整備することを目標としており、理化学研究所と富士通が覚書を結び、共同で研究開発を行っている [7]。また、効率的に分散学習を可能とするフレームワーク、アプリケーションの研究開発も進められている。

プロジェクトの取り組みによって整備されたソフトウェアは、OSS (Open-Source Software) として順次公開される [8]。

3.2 深層学習フレームワークの整備

本節では、DL4Fugakuプロジェクトの成果である、深層学習フレームワークの概要および性能評価について述べる。

(1) 深層学習フレームワークの動向、構造

PyTorchをはじめとして、TensorFlow、Chainer、MXNetなどの深層学習フレームワークは、OSSとして公開されている。PyTorchとTensorFlowは、Pythonなどの各種言語向けに提供されるAPI (Application Programming Interface) 部と、C++で実装された演算部とを分けた設計を採用している (図-1)。演算部がC++で実装されている理由は、高速な動作が求められる深層学習の学習や推論処理の大半を演算部で行っているためである。

深層学習における演算部の主要な処理内容は、

行列演算である。このため、Intel CPU向けに最適化されたoneAPI Deep Neural Network Library (oneDNN) [9] や線形演算ライブラリー (BLAS) などの演算ライブラリーを演算部から呼び出すことで、学習や推論を高速化している。

(2) Chainerによる「富岳」の深層学習性能の実証

「富岳」によるAIソフトウェアの性能向上の実現可能性を確かめるために、日本国内で広く普及しているChainerを用いて、性能チューニングおよび大規模並列実行時の性能評価を行った。

深層学習で頻繁に行われる畳み込み演算には複数のアルゴリズムが存在するが、ChainerではGEMM (行列行列積) 演算を用いた実装が多い。このため、深層学習ライブラリーdnnl_aarch64や「京」のシステムソフトTCS (Technical Computing Suite) の数学ライブラリーと結合することを視野に入れつつ、「京」や「富岳」のCPUでも使うことができるようにGEMM演算による汎用的な高速カーネルのプロトタイプを作成し、「京」上でチューニングを実施した。実施したチューニングは、TCSの数学ライブラリーによる畳み込み演算のスレッド並列化、Optimizerの1つであるAdam演算の最適化である。これによって、「京」上で学習の演算を36.4倍高速化した。

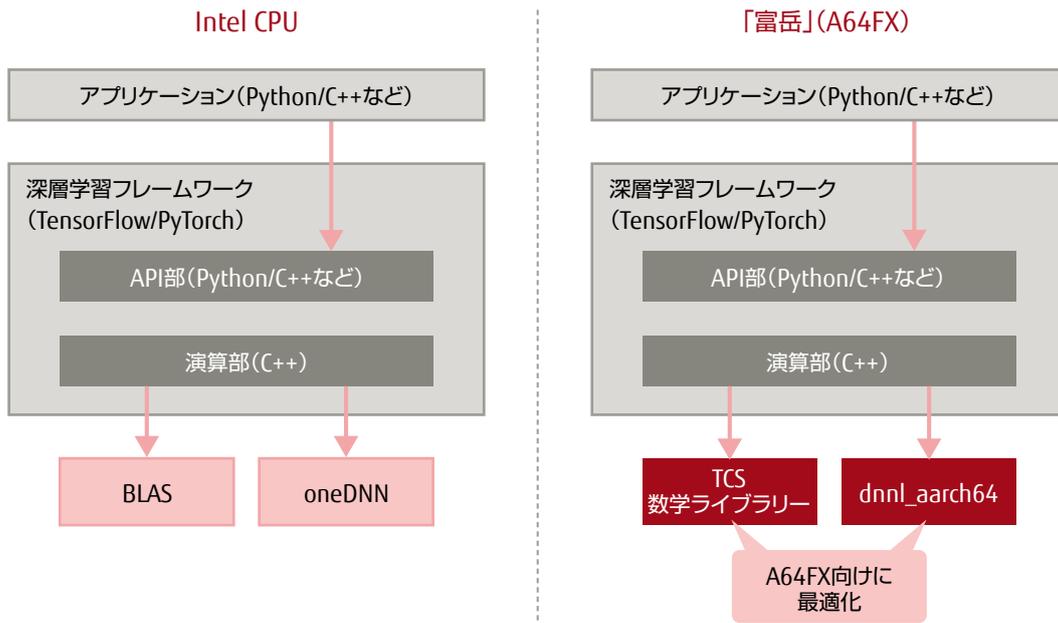


図-1 深層学習フレームワークの内部構造

「富岳」上では、新規に設計したファイルシステムLLIO (Lightweight Layered IO Accelerator) を活用した疑似ステージングによるI/O処理の高速化、Pythonパッケージのimport処理高速化など、更なる並列チューニングを実施した。これらのチューニングによって、72ラック (27,648ノード) 上で画像認識ニューラルネットワークResNet-50 v1.5を用いて学習を行った結果、410,726 ips (1秒あたり410,726枚) の学習処理性能を達成した。図-2に示すように、ラック数に比例して性能が向上しており、より大規模な構成での性能向上が期待できる。

Chainerは、2019年にメンテナンスモードへ移行し、PyTorchへの移行を促している[10]。そこで、今後の大規模並列実行の性能評価はPyTorchを用いて実施する。

(3) 「富岳」における高速化の取り組み

富士通は、ARMアーキテクチャを採用する「富岳」のCPU (A64FX) 向けに最適化したdnnl_aarch64とTCSの数学ライブラリーを開発している。Intel CPU向けに最適化されたoneDNNとBLASをそれぞれdnnl_aarch64とTCSの数学ライブラリーに置き換えることで、「富岳」上における

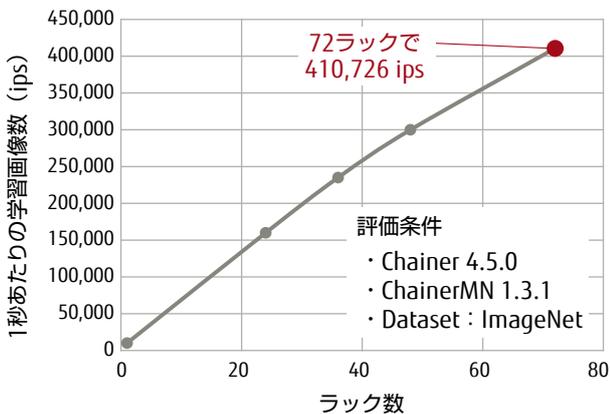


図-2 Chainerを用いた分散学習の性能 (ResNet-50)

深層学習フレームワークの高速化が期待できる。

(4) PyTorch/TensorFlowでのResNet-50評価

dnnl_aarch64とTCSの数学ライブラリーをPyTorchおよびTensorFlowに組み込み、A64FXを用いてResNet-50 v1.5の性能評価を行った。

PyTorchでは、1ノードにおける学習で98.8 ips、推論で383.7 ipsを達成した (表-1 赤字部分)。また、深層学習フレームワーク向けの分散学習ライブラリーHorovodを利用し、分散学習時の性能評価を行った結果、8ノードを利用した分散学習では、781.5 ipsを達成している (図-3)。

また、PyTorchとともに広く普及しているTensorFlowについても動作を確認し、1ノードにおいて学習で86.9 ips、推論で295.6 ipsを達成している (表-1 赤字部分)。更に、Horovodを利用した分散学習では、651.5 ipsを達成している (図-3)。

今後は、Chainerにおける大規模並列実行の実績を踏まえ、8ノードを超えるより大規模な環境において性能評価を実施する。

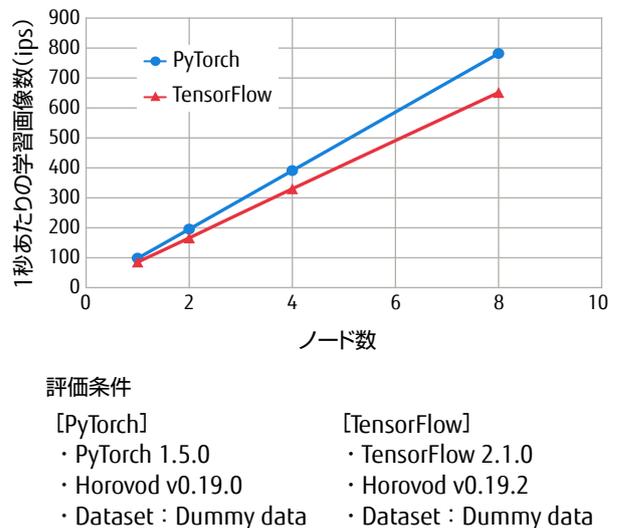


図-3 分散学習の性能

表-1 PyTorch/TensorFlowの単ノード性能

フレームワーク	データセット	プロセス数	学習性能 (ips)	推論性能 (ips)
PyTorch 1.5.0	Dummy	1	82.3	337.6
	Dummy	4	98.8	383.7
TensorFlow 2.1.0	Dummy	1	60.5	176.8
	Dummy	4	86.9	295.6

4. 今後の取り組み

(1) ARMコミュニティとの協調

A64FX向けに開発したdnnl_aarch64は、OSSとして公開している [11]。また、ARMアーキテクチャーを採用したCPUでPyTorchやTensorFlowを利用する方法は、Arm HPC Users Groupから情報発信されている [12]。今後は、「富岳」における性能評価や動作検証で得られた知見をフィードバックし、ARM CPUにおけるAI基盤の整備に貢献していく。

(2) 新技術の適用

富士通は、実行時に演算の内容に応じて演算精度を自動的に調整するCAC [13] など、AIの処理を高速化する技術を開発中である。開発した技術をAIソフトウェアに適用し、より高性能なAI基盤を提供したい。また、A64FXが得意とするアプリケーションの調査についても、引き続き行う。

5. むすび

本稿では、HPCとAIの関係、および「富岳」におけるAIの取り組みについて述べた。

「富岳」における深層学習フレームワークの取り組みの成果により、HPCシステムでのAI基盤の整備が進むことが期待される。今後は、より大規模な分散学習の性能評価、およびA64FXが得意とするアプリケーションの調査を行い、HPCにおけるAIの活用を推進していく。また、ARM社やその他のOSSコミュニティと協調することで、ARMにおけるAI基盤の整備を進め、Society 5.0の実現に貢献していきたい。

本稿に掲載されている会社名・製品名は、各社所有の商標もしくは登録商標を含みます。

参考文献・注記

- [1] 内閣府：Society 5.0.
https://www8.cao.go.jp/cstp/society5_0/
- [2] 松岡 聡：スーパーコンピュータ「富岳」. p. 17 (2020年5月).
https://www.r-ccs.riken.jp/wp-content/uploads/2020/05/20200515_matsuoka.pdf

- [3] B. Adriano et al. : Cross-domain-classification of tsunami damage via data simulation and residual-network-derived features from multi-source images. Proceedings of 2019 IEEE International Geoscience and Remote Sensing Symposium, p. 4947-4950.
- [4] T. Dey et al. : Optimizing Asynchronous Multi-Level Checkpoint/Restart Configurations with Machine Learning. 2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), New Orleans,USA, 2020.
- [5] R. Hataya et al. : “Faster AutoAugment: Learning Augmentation Strategies using Backpropagation.” arXiv:1911.06987, 2019.
- [6] S. Matsuoka : Fugaku as the Centerpiece of Society5.0 Revolution. p. 42 (Feb 2020).
https://www.r-ccs.riken.jp/R-CCS-Symposium/2020/shared/images/under/program/1-01_Matsuoka.pdf
- [7] 理化学研究所 計算科学研究センター：「富岳」上のAI（人工知能）フレームワーク構築に向けた覚書を富士通株式会社と締結しました。
<https://www.r-ccs.riken.jp/library/topics/191126.html>
- [8] GitHub : DL for Fugaku.
<https://github.com/dl4fugaku>
- [9] Intel Open Source Technology Center : oneAPI Deep Neural Network Library (oneDNN).
<https://01.org/onednn>
- [10] Preferred Networks : Preferred Networks、深層学習の研究開発基盤をPyTorchに移行。
<https://preferred.jp/ja/news/pr20191205/>
- [11] GitHub : Fujitsu.
<https://github.com/fujitsu>
- [12] GitLab : Arm HPC Users Group.
<https://gitlab.com/arm-hpc>
- [13] 富士通研究所：計算精度を自動調整する「Content-Aware Computing」により、AI処理を10倍高速化（技術者インタビュー）。
<https://www.fujitsu.com/jp/group/labs/about/resources/article/202002-cac.html>

著者紹介



忽滑谷 淳史 (ぬかりや あつし)
富士通株式会社
プラットフォームソフトウェア事業本部
HPC向けのAIソフトウェア研究開発に
従事。



南 一生 (みなみ かずお)
国立研究開発法人理化学研究所
計算科学研究センター 運用技術部門
「富岳」向けアプリケーション開発基盤
の整備に従事。



赤尾 和俊 (あかお かずとし)
富士通株式会社
プラットフォームソフトウェア事業本部
HPC向けのAIソフトウェア研究開発に
従事。



佐藤 賢斗 (さとう けんた)
国立研究開発法人理化学研究所
計算科学研究センター 高性能ビッグ
データ研究チーム
HPC向けAI/ビッグデータ処理のための
システムソフトウェア研究開発に従事。



高橋 仁 (たかはし じん)
富士通株式会社
プラットフォームソフトウェア事業本部
HPC向けのAIソフトウェア研究開発に
従事。



松岡 聡 (まつおか さとし)
国立研究開発法人理化学研究所
計算科学研究センター センター長/
高性能人工知能システム研究チーム
DL4Fugakuプロジェクトを統括。



福本 尚人 (ふくもと なおと)
株式会社富士通研究所
ICTシステム研究所
HPC向けのAIソフトウェア研究開発に
従事。



川上 健太郎 (かわかみ けんたろう)
株式会社富士通研究所
プラットフォーム革新プロジェクト
HPC向けのAIソフトウェア研究開発に
従事。



黒田 明義 (くろだ あきよし)
国立研究開発法人理化学研究所
計算科学研究センター 運用技術部門
「富岳」向けアプリケーション開発基盤
の整備に従事。

この記事は、富士通の技術情報メディア「富士通
テクニカルレビュー」に掲載されたものです。
他の記事も是非ご覧ください。

富士通テクニカルレビュー

<https://www.fujitsu.com/jp/technicalreview/>

