

AIを支えるコンピューティングプラットフォーム

石坂 一久 荒木 拓也 井上 浩明

要旨

AIではデータ量の大量化やアルゴリズムの複雑化により、膨大な計算性能が必要となっており、特定の用途に特化したハードウェアアクセラレータの活用が重要となっています。多岐にわたるAIの個々の開発者がこれらのアクセラレータを使いこなすのは困難で、フレームワークと呼ばれるアクセラレータをサポートしたソフトウェアも重要となります。本稿では、NECが開発しているアクセラレータをサポートしたことを特徴とする統計的機械学習向けのフレームワークFrovedis、統計的機械学習と深層学習の両者をサポートしたアクセラレータVector Engineを紹介します。



AI/プラットフォーム/フレームワーク/アクセラレータ/Frovedis/Vector Engine/SX-Aurora TSUBASA

1. はじめに

AIでは、統計的機械学習、深層学習、強化学習、組み合わせ最適化、画像解析などさまざまなアルゴリズムが利用されます。また、AIは大量のデータからの学習に膨大な計算時間がかかり、アルゴリズムの複雑化やデータの大量化により必要な計算量は年々増加しています¹⁾。一方、近年では約18カ月で性能が2倍になるというMooreの法則の鈍化などにより、コンピューティング分野では特化型アーキテクチャを使いこなすことが重要になってきています。今までも処理に特化した特化型アーキテクチャが幾度となく登場してきましたが、Mooreの法則によって汎用プロセッサであるCPUに駆逐されてきた歴史があります。しかし、2005年ごろからチップの消費電力制約により動作周波数が限界を迎え、CPUの単一スレッドの処理性能が頭打ちとなりました。これによりマルチコア化が進展するものの、プログラムのなかで逐次的に実行しなければならない部分によって高速化が制限され、これ以上コア数が増加しても性能向上が図れない状況になりつつあります(アムダールの法則)。そのため近年では、AI処理の特徴を踏まえ、GPUやベクトルプロセッサ、深層学習チップ、FPGAなど、さまざまなアクセラレータ(ハード

ウェア)を適材適所に使い分けることがトレンドとなっています。

複雑化する社会課題に対応するため将来の社会ソリューション実現におけるAIプラットフォームの課題は、設計工数が膨大になることです。その理由は前述した通り、特化型アーキテクチャの進展です。特化型アーキテクチャで性能を向上させるためには、ソリューション実現に必要なAI処理を完全に理解したうえで、更に処理に適したハードウェアを選択・組み合わせる必要があります。アクセラレータの種類が増加すればするほど設計工数が増加します。

このような膨大な設計工数に対応するために重要となるのが、AIのためのソフトウェアフレームワーク(以下、フレームワーク)です。フレームワークは、前述したようなさまざまなAIのアルゴリズムを実装するために利用される、ライブラリーと呼ばれる部品を提供するソフトウェアです。多くの場合、アルゴリズムはこれらのライブラリーを組み合わせることで実現されます。フレームワークが、あらかじめアクセラレータに対応したライブラリーを提供することで、AIの開発者は容易にアクセラレータに最適化されたAIを開発することが可能となります。

以降では、アクセラレータとフレームワークについて説明します。

2. アクセラレータ

前述したように、AIでは、汎用性が高くいろいろな種類のソフトウェアに対応しているCPUに加えて、汎用性よりも性能を重視したプロセッサを搭載したアクセラレータが利用されます。アクセラレータについて紹介する前に、まずプロセッサの性能向上の手法について説明します。近年のプロセッサは、ベクトル処理を活用して性能を向上させています。ベクトル処理とは、複数の異なるデータに対して、同時並列に演算を行う処理です。プロセッサ内に複数の演算器を搭載し、これらの演算器を同時に動かすことでベクトル処理を行います。ベクトル処理は、演算器の数を増やし、その利用率を高めることで性能を向上させることができます。AIのためのライブラリーにはベクトル処理が可能なものが多くあります。

一方で、多数の演算器を同時に動かすことができない処理をスカラ処理と呼びます。スカラ処理の性能を上げるためには、演算器の数ではなく、演算器の動作速度を向上させることや、高機能な演算器を用いることが重要となるため、ベクトル性能を向上させるよりもコストがかかります。AIのフレームワークでは、ライブラリーの実行管理などライブラリー以外でのスカラ処理が多くみられます。

スカラ処理の性能を重視したプロセッサがCPUであり、ベクトル処理の性能を重視したものがアクセラレータと言えます。CPUでもベクトル処理を導入していますが、アクセラレータは、シンプルな演算器をより多数搭載することやメモリ帯域幅を増やすことで、ベクトル処理に必要な演算性能やメモリ性能を上げており、CPUに比べて数倍のベクトル性能を実現しています。AIでは、ライブラリーの処理をアクセラレータで実行し、それ以外の処理をCPUで実行するという形態がとられます。なお、メモリ性能とは演算器にデータを供給する能力であり、これが不十分だと演算器がアイドル状態になりますので、プロセッサの重要な性能指標です。

アクセラレータは、汎用型と深層学習特化型に分けることができます。汎用型は、CPUほどの汎用性はありませんが、プログラマブルなプロセッサコアを持ち、プログラムを書き換えることでさまざまな処理を行わせることができます。このようなアクセラレータとして、演算器の数を重視したGPU、メモリ性能を重視したVector Engineが挙げられます。消費電力やチップ面積の制約から、演

算性能とメモリ性能の両者を最高レベルにすることは難しく、そのバランス配分にアクセラレータの性格が表れます。

一方、画像認識などに使われる深層学習では、Convolutionと呼ばれる処理がボトルネックになります。このConvolutionに特化した演算器を搭載することで、搭載可能な演算器の数を増やし、性能を向上させるアクセラレータが登場してきています。このようなアクセラレータを深層学習特化型と呼ぶこととします。例としてはTPUが挙げられます。GPUでも、Convolutionに特化した演算器を搭載するモデルが登場しており、深層学習特化型の性質も持つようになっています。

3. フレームワーク

フレームワークは、AI向けのライブラリーを提供することに加え、ライブラリーの実行管理、メモリ管理、アクセラレータ管理などを行います。また、複数のライブラリーを結合するなどの最適化、サーバ間の通信の最適化といった処理速度を速めるための最適化も行います。更に、学習プロセスの可視化や性能のプロファイリング機能も提供します。このようにフレームワークはAIを開発するうえで非常に多くの機能を提供し、AI開発には欠かせないものとなっており、活発に技術開発が行われています。

フレームワークの特徴は、

- ・ サポートするライブラリーの種類
- ・ プログラミングインタフェースのデザイン
- ・ 対応するプログラミング言語
- ・ サポートするアクセラレータの種類
- ・ 分散処理への対応
- ・ ベンダーによるサポートの有無

などのさまざまな観点をもち、開発するAIアルゴリズムの性質や開発チームのスキル・経験などによって選択されます。また、フレームワークが必要なライブラリーを提供していない場合は、自分で実装することになりますので、そのような場合の拡張性も重要となるでしょう。

これらの特徴のなかでも、特に重要なものは「サポートするライブラリーの種類」と「サポートするアクセラレータ」と言えます。フレームワークが、開発したいアルゴリズムに十分なライブラリーを備えているかは、開発効率に大きな影響を与えます。ライブラリーの種類の観点では、フレームワークは統計的機械学習重視型と深層学習重視

型に分類できます。深層学習は画像分類を中心に注目されているため、多くの深層学習重視型のフレームワークが登場しています。そのようなものとしてTensorFlow、Keras、PyTorch、Caffe、Theano、MXNet、RAPID 機械学習などがあります。一方で、統計的機械学習を重視したものとしては、Frovedis、Spark（機械学習lib）、Scikit-learnなどがあります。

処理性能の観点では、フレームワークがサポートするアクセラレータが重要となります。第2章で述べたように、AI向けには複数のアクセラレータがあります。それぞれ異なった性能特性やプログラミング特性を持っているため、すべてのアクセラレータに対応するのは難しいです。複数のアクセラレータに対して統一的な枠組みで実行管理やプロファイリングを行うこと、アクセラレータを効率的に使うための最適化などに関する技術開発が行われています。

4. NECでの取り組み

NECでは、今後深層学習だけでなく統計的機械学習などのさまざまなAIの高速化が重要になると考えています。本章では、NECが開発しているアクセラレータVector Engine、統計的機械学習用のフレームワークFrovedisについて説明します。

4.1 アクセラレータVector Engine

Vector Engineは、ベクトル型スーパーコンピュータSXシリーズの技術を継承したアクセラレータです。Vector Engineのハードウェアとしての特徴は、現時点で世界最高のメモリ帯域（1.2TB/s）を誇ることです。ソフトウェアとしての特徴は、C/C++といった一般的な言語でプログラム可能であることと、強力な自動ベクトル化機能を持つコンパイラを備えていることが挙げられます。これらはライブラリーの最適化に重要であり、フレームワーク開発者や最先端のアルゴリズムを開発する技術者にとって、最適なアクセラレータと言えます。

なお、Vector Engineを搭載したサーバとしてSX-Aurora TSUBASAシリーズがあります。タワー型、ラック型、大型機と多様なラインアップを揃えており、さまざまなユースケースで利用可能となっています。写真にVector Engineの外観を示します。



写真 Vector Engineの外観

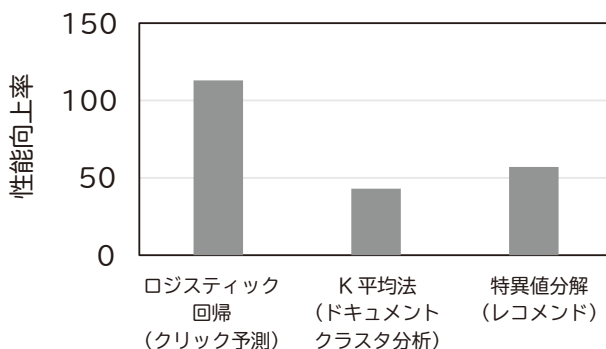


図 Vector Engine上のFrovedisの性能 (CPU上のSpark比)

4.2 統計的機械学習用フレームワークFrovedis

Frovedisは、データ分析用の高性能なフレームワークです。CPUとVector Engineをサポートしており、分散処理にも対応しているのが特徴です。このため、複数のサーバを必要とするような大規模な処理も高速に実行できます。また、統計的機械学習向けのライブラリーが充実しています。プログラミングインタフェースとしては、C/C++だけでなく、プログラミング言語のPython、Scalaに対応しており、統計的機械学習向けライブラリーであるPythonのScikit-learnやSparkのMLlib互換のAPIを提供しています。したがって、これらを利用しているソフトウェアであれば、少しのプログラム修正で利用することができます。

統計的機械学習のライブラリーにはメモリ帯域依存の処理が多く含まれますが、Frovedisはメモリ帯域の高い

表 フレームワークの対応アルゴリズムとプロセッサ

アルゴリズム	代表的な フレームワーク	対応する アクセラレータ		
		汎用型		DL特化型
		GPU	Vector Engine	TPU
深層学習	TensorFlow	○	○	○
	PyTorch	○	×	×
統計的機械学習	Frovedis	×	○	×
	Scikit-learn	×	×	×

ポイント 1：統計的機械学習では Frovedis のみアクセラレータをサポートしている
 ポイント 2：Vector Engine のみが深層学習と統計的機械学習向けのフレームワークでサポートされている

Vector Engine をサポートしており、これらのライブラリーを非常に高速に実行できます。図に Frovedis の性能を示します。この図は、ロジスティック回帰、K平均法、特異値分解と呼ばれる機械学習向けのライブラリーの性能を、Vector Engine 上での Frovedis と CPU (Intel Xeon) 上での Spark で比較しています。この結果は、Vector Engine 上の Frovedis は数十倍から百倍程度高速で、従来数十時間かかっていた処理を、数分で終わらせるポテンシャルを持っていることを示しています。

4.3 フレームワーク対応状況

表に代表的なフレームワークのサポートするアルゴリズムとアクセラレータを示します。TensorFlow は、深層学習向けですが世界で最も広く利用されているフレームワークであり、対応するアクセラレータが多いという特徴があります。統計的機械学習向けフレームワークでは、Frovedis のみがアクセラレータに対応しています。一方、アクセラレータを見ると、Vector Engine が統計的機械学習向けフレームワーク (Frovedis) と深層学習向けフレームワーク (TensorFlow) の両方でサポートされている唯一のアクセラレータです。これらのことからフレームワークでは Frovedis、アクセラレータでは Vector Engine が AI では重要な立場にあります。

5. むすび

本稿では、AIを支えるコンピューティングプラットフォームとして、特に重要なアクセラレータ Vector Engine とフレームワーク Frovedis について説明しました。NEC は、多様な AI 処理に対してさまざまなアクセラレータを適材適所で活用するためのフレームワークを整備していきます。

参考文献

- 1) AI and Compute, 2018.5
<https://blog.openai.com/ai-and-compute>

執筆者プロフィール

石坂 一久

データサイエンス研究所
主任研究員

荒木 拓也

データサイエンス研究所
主幹研究員

井上 浩明

データサイエンス研究所
研究部長

関連 URL

次世代イノベーションプラットフォーム SX-Aurora TSUBASA
<https://jpn.nec.com/hpc/sxauroratsubasa/index.html>

Frovedis

<https://github.com/frovedis/frovedis>

NEC 技報のご案内

NEC 技報の論文をご覧くださいありがとうございます。
ご興味がありましたら、関連する他の論文もご一読ください。

NEC技報WEBサイトはこちら

NEC技報 (日本語)

NEC Technical Journal (英語)

Vol.72 No.1 新たな社会価値を生み出すAI特集

新たな社会価値を生み出すAI特集によせて
AIとデータ活用によるデジタイゼーションの拡大

◇ 特集論文

AIの社会実装に向けた取り組み

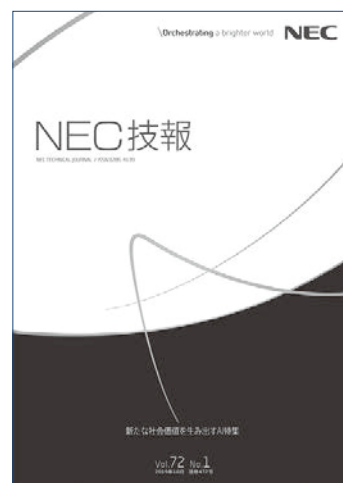
「NECグループ AIと人権に関するポリシー」とその実践に向けた取り組み
AI時代の人材育成

デジタルトランスフォーメーションを加速するAI活用サービス・ソリューション

「みんなで創るAI」を支えるNEC Advanced Analytics Platform (AAPF)
物体指紋認証技術による個人識別機能の活用
画像処理コントローラへのディープラーニング活用による外観検査ソリューション
通信予測制御技術を活用した車両の遠隔監視ソリューション
働き方改革や健康経営を支える「NEC感情分析ソリューション」
オフィスのセキュリティと利便性を向上する「顔認証ソリューション for オフィス」
業務自動化・省力化を実現する自動応答ソリューション (AIチャットボット) の概要
ビジネス創造へのワークシフトを加速するソリューション (AI for Work Shift Support) の概要と実証事例
AIを有効活用する「NEC Energy Resource Aggregation クラウドサービス」
容体変化予兆検知技術による早期退院支援の取り組み
予防・健康領域に対するデータ活用による効果的なアプローチ
AIを活用したインサイトマーケティング事業の共創
時代のムードを味わえる「あの頃はCHOCOLATE」の開発

人とともに未来を創る最新のAI技術

あらゆる小売商品を認識可能にする多種物体認識技術
ネットワークインフラを活用して実世界を見える化する光ファイバセンシング技術
熟練者の意思決定を模倣する意図学習技術
グラフベース関係性学習 (GraphAI)
時系列データ モデルフリー分析技術
社会インフラの最適運用を支援する論理思考AI
少量データ向け深層学習技術
AIを支えるコンピューティングプラットフォーム



Vol.72 No.1
(2019年10月)

特集TOP