

## ディープラーニングを用いた タンパク質の構造予測や医用画像分類において Habana® Gaudi® AI プロセッサーが 学習速度の向上に貢献

国立研究開発法人 理化学研究所において、医療と創薬のAIをテーマに研究を進めている種石慶氏。ディープラーニングを活用した学習に、AIアクセラレーターとして、Habana® Gaudi® AI プロセッサーを採用しています。そこで、ヘルスケアとライフサイエンスの分野におけるHabana® Gaudi® AI プロセッサーの活用についてお話を伺いました。



国立研究開発法人 理化学研究所  
量子工学研究センター  
データサイエンティスト  
種石 慶氏

### Transformerの普及により、より高い計算能力が必要に

種石氏は現在、ディープラーニングを活用した研究として、胸部X線写真などの医用画像による疾患の解析や、分子標的に作用する低分子化合物の仮想スクリーニング、タンパク質の構造や機能変化の解明などに取り組んでいます。これらのAIを用いた研究には、大規模な計算基盤が必要不可欠で、AIアクセラレーターの役割が重要視されています。

「2017年に自然言語処理分野でTransformerが提案され、アテンション機構がディープラーニングの要素技術となってから、自然言語処理以外の分野でもこれまでの畳み込みや再帰型のニューラルネットワークで記述されていたモデルが書き換えられました。その結果、ディープラーニングに求められる計算の枠組みが大幅に拡大し、より高い計算能力が求められるようになりました」(種石氏)

理化学研究所には、共用の大規模計算基盤として「HOKUSAI SailingShip」(HSS)等がありますが、HSSは主にデータサイエンスでの活用を想定して最適化されているため、AIアクセラレーターは搭載されておらず、AIに特有の計算には計算科学研究センターの「富岳」や、革新知能統合研究センター(AIP)のAI研究用計算機システム「RAIDEN」などが利用され、加えて大手4社のクラウド利用が支援されることで、オンプレミスとクラウドを合わせて理化学研究所全体でヘテロジニアスな計算環境を構築しています。

### 高いコスト・パフォーマンスとモデルの移植の容易さを評価

クラウドサービスで提供されるコンピューティング・リソースの中で、種石氏が着目したのがHabana® Gaudi® AI プロセッサーを搭載した「Amazon EC2 DL1 インスタンス」です。Amazon EC2 DL1 インスタンスは、32GBの高帯域幅メモリーを備えた8個のHabana® Gaudi® AI プロセッサー、96 vCPUを提供する第2世代インテル® Xeon® スケーラブル・プロセッサー、768GBのメモリー、帯域400Gbpsのネットワーク、4TBのローカルストレージなどを主な仕様としています。Habana® Gaudi® AI プロセッサーには、8個のTensor Processing Cores (TPC)、32GBの広帯域メモリー、100Gbit Ethernet × 10などが統合されています。

「Habana® Gaudi® AI プロセッサーは、テンソル演算と行列積演算の両方にハードウェアで対応しており、ディープラーニングの計算グラフを最適化コンパイルすることで、これらのハードウェアの性能を十分に引き出している点に魅力を感じています。Transformerが標準的に使われるようになった2019年にリリースされたこともあり、Transformerの効率的な学習がサポートされている点も評価しています。さらに同世代のAIアクセラレーターの中でも後発ということで、コスト・パフォーマンスも高く、V100 GPUと同等のコストで2倍以上の学習スループットが期待できます」(種石氏)

TensorFlowやPyTorchのモデルをHabana® Gaudi® AI プロセッサーで利用するためには、スクリプトの書き換えが必要ですが、Habana Labからは開発環境として「Habana® SynapseAI® SDK」が提供され、基本的な文法の互換性がうたわれています。そのため、同じスクリプトで計算資源に応じてAIアクセラレーターを切り替えて使用でき、移行コストも安価となります。

「TensorFlowやPyTorchに備わっているAIアクセラレーターを使用するための構文がそのまま利用できるため、Habana® SynapseAI® SDKがサポートしているモデルやその派生モデルであれば移植は容易です。AIアクセラレーターの制御のために、数行程度のコードの追加が必要となりますが、データ、モデルの定義や学習や推論のプロセスで互換性は問題ありませんでした。研究者にとっては、論文などで公開されたモデルをいち早く使ってみてみたいといった時に、最もコスト・パフォーマンスの高い計算機を使えるということが最大のメリットです」(種石氏)

## タンパク質の二次構造予測で22%、 医用画像の疾患分類で18%高速化

種石氏が、Habana® Gaudi® AI プロセッサを利用して実施したディープラーニングの学習として、言語モデルのBERT-Largeを使ったタンパク質の二次構造予測と、胸部X画像に対するCheXNetを使った疾患分類の2つにおいて検証を実施しました。

タンパク質の二次構造予測は立体構造予測の初期段階で、タンパク質を形成する20種類のアミノ酸配列を入力として、構造の決定されたタンパク質のデータベースでディープラーニングのモデルを学習して、タンパク質の二次構造配列を推論するものです。検証の結果、一回の繰り返し当たりの学習に要した時間は、V100 GPUで5.9秒だったものが、Habana® Gaudi® AI プロセッサでは4.6秒と22%の高速化が確認できました(図1)。

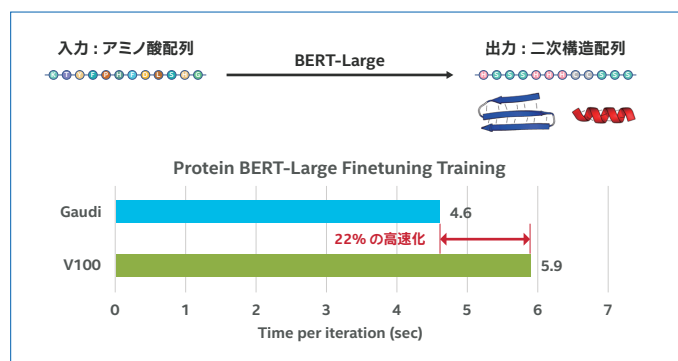


図1: タンパク質の二次構造予測

CheXNetを使った胸部X画像の疾患分類は、患者の胸部X線画像を入力として、30,805人の患者から得た述べ112,120枚の胸部X線画像をデータセットとしたデータベースから、肺炎などの疾患と部位を推定するものです。その結果、CheXNetの一回の繰り返し当たりの学習に要した時間は、V100 GPUで1047.7秒だったものが、Habana Gaudi AI プロ

セッサでは859.1秒と18%の高速化が確認できました(図2)。複数のHabana® Gaudi® AI プロセッサを分散データ並列で実行した際のオーバーヘッドも少なく、単一ノード内での高いスケーラビリティを確保しています。

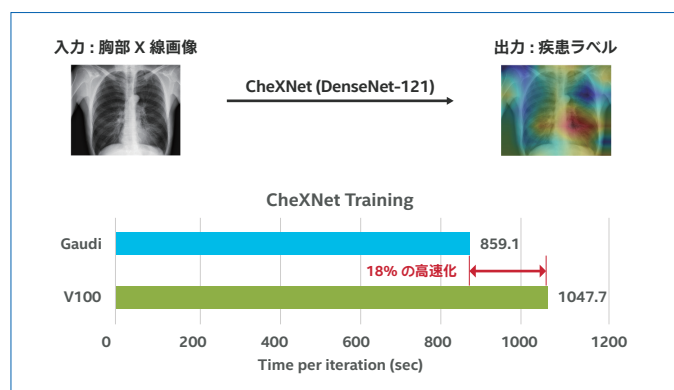


図2: 胸部X線画像の疾患分類

## 前世代より約3倍性能がアップした Habana® Gaudi®2 AI プロセッサに期待

今後については、創薬AIの分野ではタンパク質中のリガンド結合の有無によって生じるタンパク質の立体構造と機能の変化の解明を進めるため、現在はMDトラジェクトリからアミノ酸残基間の距離を学習し、リガンド結合で生じる変化を検出することに取り組んでいます。医療AIの分野については、数万人規模の電子カルテ、医用画像、ゲノム、生活習慣など、大規模かつ複合的なデータが利用可能になってきていることから、医用画像に加えてマルチモーダルなデータを用いたモデルの研究に取り組んでいくといえます。

Habana® Gaudi® AI プロセッサは、2022年5月に第2世代の「Habana® Gaudi®2 AI プロセッサ」が発表されました。Habana® Gaudi®2 AI プロセッサは、ノードが従来の16nmから7nmへと微細化され、TPCの数も8から24へと3倍となり、演算効率を改善しています。メモリーは32GBから96GB、ネットワークも10×100Gbitから24×100Gbitと大幅に強化。これらの改善により、ResNet-50の学習スループットはA100 GPUの約2倍、前世代のGaudiの約3倍と大幅に性能がアップしています。

「Habana® Gaudi® AI プロセッサの採用実績が増えて対応したモデルやスクリプトが普及すれば、Habana® Gaudi®2 AI プロセッサの導入のハードルも下がると思いますので、今後クラウド環境で利用しやすいサービスが出てきたらぜひ試してみたいと思います」(種石氏)



インテルのテクノロジーを使用するには、対応したハードウェア、ソフトウェア、またはサービスの有効化が必要となる場合があります。

絶対的なセキュリティを提供できる製品またはコンポーネントはありません。

実際のコストや結果は異なる場合があります。

Intel、インテル、Intel ロゴ、その他のインテルの名称やロゴは、Intel Corporation またはその子会社の商標です。

その他の社名、製品名などは、一般に各社の表示、商標または登録商標です。

インテル株式会社

〒100-0005 東京都千代田区丸の内 3-1-1  
http://www.intel.co.jp/

©2022 Intel Corporation. 無断での引用、転載を禁じます。  
2022年10月

353411-001JA  
JPN/2210/PDF/SE/BCG/YS