



## 医療研究の進化をディープラーニングで支援する 日立のスーパーテクニカルサーバ「SR24000/DL1」

高性能な並列演算機能を持つGPU<sup>\*1</sup>をディープラーニング(深層学習)に応用して、大規模シミュレーションや画像認識、医療研究などに役立てようという動きが活発化しています。国立研究開発法人 国立成育医療研究センターは、日立のスーパーテクニカルサーバ「SR24000シリーズ」のディープラーニング専用モデル「SR24000/DL1」を導入。超高速なビッグデータ解析環境を、医療研究に幅広く適用する取り組みを開始しました。 <sup>\*1</sup> Graphics Processing Unit

### 全ゲノム解析への対応を図るため 「SR24000/DL1」を導入

国立成育医療研究センターは、胎児期から新生児、小児、思春期を経て成人に至るまでに生じる疾患への医療と研究を推進する機関です。病院と研究所が一体となり、難病に悩む患者さんや家族に、安全性と有効性が十分に検証された高度先駆的医療の提供を行うため、基礎・臨床研究にも力を入れています。

研究所では2018年、最新GPUを搭載した日立のスーパーテクニカルサーバ「SR24000/DL1」を導入しました。研究所の中でも、遺伝子発現制御の分子メカニズムを研究し、成育疾患の原因解明や原因遺伝子の同定、新たな研究システムの開発などを進めているのがシステム発生・再生医学研究部です。導入の背景を、システム発生・再生医学研究部組織工学研究室長の岡村 浩司氏は「遺伝子の塩基配列を高速に読み出せる次世代シーケンサーの登場以来、医療・生物学の研究者にとって、日々膨大に出力される配列情報のビッグデータを、高速なコンピュータシステムによって解析する業務が不可欠となっています。本研

究所は、原因不明で治療の進まない症状を持つ小児患者さんに対し、ゲノム解析で原因や診断の手がかりを探す全国規模の研究プロジェクト「IRUD」<sup>\*2</sup>の拠点でもあるため、2013年に導入した日立の『HA8000サーバクラスシステム』を用いて、DNAの変容を見つけるためのエクソーム解析を行ってきました。ただしエクソーム解析が全ゲノムの1%程度だけを解析する手法であるのに対し、近年は次世代シーケンサーおよび計算機の性能向上によりほぼすべてをカバーする全ゲノム解析が研究の主流となりつつあります。そのデータを最大限に活用し、かつ研究スピードを高めるために、SR24000/DL1を導入しました」と語ります。

<sup>\*2</sup> Initiative on Rare and Undiagnosed Diseases:  
「希少・未診断疾患イニシアチブ」

### 1か月前後かかっていた解析時間が 1日半程度に短縮

SR24000/DL1は、脳の仕組みを模したニューラルネットワーク(NN)を多層に重ねたディープニューラルネットワーク(DNN)をサポートするシステムです。深層学習の学習時間を短縮し、医療研究だけでなく、流体計算や気象予測などの

大規模シミュレーション、大量データを用いた科学技術計算においても高い性能を発揮します。

すでに、システム発生・再生医学研究部ではSR24000/DL1を活用した研究成果が次々と生まれています。その一例が、研究者であり医師でもある瓜生 英尚氏による研究課題「機械学習法を用いたクロマチン高次構造の予測モデル構築」です。

「真核細胞内に存在するDNAとタンパク質の複合体であるクロマチンは、ループ構造をとることで遺伝子発現の制御や、スプライシング<sup>\*3</sup>に影響を及ぼしていると報告されています。またループ構造の破綻により、胎児が育つ段階で正常な発育が妨げられたり、ガンなどの原因になり得たりすることもわかっています。しかし、どのような条件で破綻するかの予測は、既存のシステムでは解析するのが困難でした。そこで今回われわれは、関連する因子すべてをSR24000/DL1のDNNに入れ、ループ構造の有無を予測するモデル構築に取り組みました。その結果、DNA配列、エピゲノム情報<sup>\*4</sup>を学習データとして各要素を組み合わせることで、高精度の予測モデルが構築できることを導き出

## 国立研究開発法人 国立成育医療研究センター

所在地 東京都世田谷区大蔵2-10-1  
 設立 2002年3月  
 職員数 1,594名(非常勤含む/2018年1月1日現在)  
 病床数 490床(2017年)



しました。今回のような予測モデルを活用すれば、将来的には各発達段階や分化段階でのクロマチン構造の異常による疾患の診断につながる可能性があると思います」と瓜生氏は語ります。

瓜生氏によれば、既存システムでは1回の解析結果が出るまでに1か月前後はかかっていたとのこと。それが新システムでは「早ければ1日半、データ量が多くても3日前後で結果が出ます。この速さには本当に助けられています」と評価します。

※3 遺伝情報の中から不要な部分を取り除く分子的な編集作業  
 ※4 DNAの塩基配列を変えずに遺伝子の働きを決める情報の集まり

## 深層学習を活用し、研究者をサポートするシステムを構築

一方、岡村氏と学生の片桐 沙紀氏がSR24000/DL1を使って行った研究が「ゲノム塩基配列における特定塩基長の頻度情報を利用した深層学習による種分類」です。

「現在、国内外でさまざまな生命科学系のデータベースが存在します。そのデータを研究に活用する際は、データの付帯情報であるメタデータを頼りに検索しますが、データをアップロードする側の研究者にとっては、このメタデータ登録が非常に煩雑な作業となるため、データ共有がなかなか進まない要因のひとつとなっています。そこでわれわれは、メタデータの登録作業を深層学習で効率よく自動で行う分類システムをつくるため、SR24000/DL1を使った検証実験を行いました。生物種としてヒトとマウスなど事前に学習した12種の認識が可能か、また全ゲノム、エクソーム、トランスクリプトーム、ヒストン修飾、それぞれの解析手法を



正しく判別できるかを検証したところ、適切な学習により、ミスを起こしうる手作業よりも精度の高い分類と問題検出が自動で行えるシステムを構築することに成功しました。この技術の応用で将来的には、メタデータを入力しなくても自動で次世代シーケンサーからデータベース登録できる仕組みができるようになるはずですよ」と岡村氏は語ります。

片桐氏も、「研究者の先生方を見ると、限られた時間の中でメタデータ登録のような研究以外の作業に多くの時間と手間を取られているのが、とてももったいないなと日頃から感じていました。今回つくった自動登録の仕組みが、人がより新しいことを考える時間を生み出すことにつながれば本当にうれしいですね」と笑顔で語ります。

## 深層学習は医療研究に有益なツールだと実感

SR24000/DL1を活用した深層学習の効果について瓜生氏は「研究プログラムのプロトタイプは当初うまく動かないの

が当たり前です。しかし丸1日動かして失敗に気づくと、1分で結果が出て何度も修正を繰り返せるのとでは仕事の進め方がまったく違います。また、DNNはこれまでの統計モデルに比べて精度と柔軟性がともに高く、既存の研究やデータベースを活用して新しい知見が得られることも期待できます」と語ります。

岡村氏も「深層学習は非常に有益なツールだと実感しました。医療分野においてはDNNに必要なサンプル数の確保が困難であることが多く、またGPUを活用するソフトウェアが少ないことが悩みですが、他の研究者や医療機関とも協力しながら、医療革新につながる取り組みを進めていきたいと思います。そのためにも日立さんには継続的なサポートをお願いしたいですね」と熱く語ります。

これからも日立は、多数の研究機関や企業の先進的な研究を支えてきた実績と、高度なシステム構築・運用ノウハウを活かし、「SR24000シリーズ」をはじめとする高性能・高信頼なスーパーコンピュータシステムの提案を行っていきます。

お問い合わせ先・情報提供サイト

(株)日立製作所IoT・クラウドサービス事業部  
[http://www.hitachi.co.jp/Prod/comp/hpc/SR\\_series/sr24000/](http://www.hitachi.co.jp/Prod/comp/hpc/SR_series/sr24000/)