深層学習の次の領域を切り開く NVIDIA DGX Station A100

東京大学 大学院情報理工学系研究科 創造情報学専攻 中山英樹研究室様

東京大学の中山英樹先生は、「画像処理」や「自然言語処理」の研究に取り組むとともに、その2つを組み合わせた領域やディープラーニングの新しい知的システムの研究にも挑んでいる。高性能GPU「NVIDIA A100」を搭載する「NVIDIA DGX Station A100」が、その研究をさらに加速させる。

1877年に創立された東京大学は、国内外の様々な分野で指導的役割を果たしうる「世界的視野をもった市民的エリート」の育成を自らの使命とする日本初の国立大学である。広範で深い教養とさらに豊かな人間性を培うことを目的にリベラル・アーツ教育を重視し、自ら考えて行動できる人材を育成している。

大学院情報理工学系研究科 創造情報学専攻の中山英樹先生は、画像や映像を中心に、マルチモーダルな技術を利用した認識・理解をAIが自動で行うための基礎技術を研究。ディープラーニング（深層学習）による画像認識のブレイクスルーは2012年頃とされており、トロント大学の畳み込みニューラルネットワークが圧倒的な性能で優勝した国際コンテスト「ImageNet Large Scale Visual Recognition Challenge（ILSVRC）2012」や、俗に「Googleの猫」や「キャットペーパー」と呼ばれるGoogleの発表はとても有名な出来事だが、中山氏はそれよりも前から画像認識の研究に取り組んできた。さらに、2012年以降はディープラーニングやそのほかの新しいデータ解析の研究にも精力的に取り組むなど、この分野では日本を代表する先駆者的な存在の1人である。

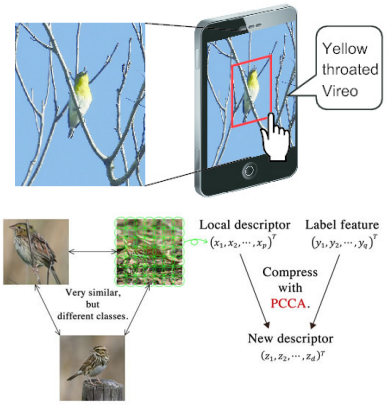
「2012年以前、画像認識は非常に難しい技術で、人間であれば子どもでも簡単にできるようなことが『コンピュータにはいつまで経ってもできない』という状況が何十年も続いてきました。しかし2012年、ディープラーニングによるブレイクスルーがその状況を大きく変えました。これは私にとっても非常に衝撃的な出来事でしたし、ディープラーニングに大きな興味と期待を感じた瞬間であったことは間違いありません」（中山氏）

現在、中山氏が取り組んでいる主な研究内容は「画像処理」と「自然言語処理」の2つ。画像処理は中山氏が長年研究してきた画像認識に通じる分野となるほか、自然言語処理は「自身の研究を“人工知能の研究”として捉えたとき、人間ならではの知能として『言語』も研究する必要があるのではないか」（中山氏）と考えて取り組むようになった。具体的な取り組みとして、画像処理では画像から物体やシーンを認識する「コンピュータビジョンとパターン認識」、自然言語処理では「機械翻訳」や「文章要約」などの典型的な研究が挙げられる。

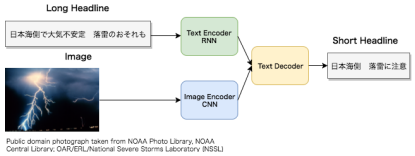
これらに加えて、中山氏が最近とくに興味を持って注力しているのが、画像処理と自然言語処理の2つを合わせて行うような領域の研究である。例えば、近年はAIを使ったチャットボットによって、AIと人間の簡易的な対話が実現されている。しかし、実際の人間同士の場合、お互いの会話の内容を言葉の字面のみで理解して対話しているのではなく、言葉以外の様々な情報（周りの状況や相手の表情など）も含め、総合的に判断して対話をしている。実際、人間同士の対話であれば「それを取って」と言われただけでも、視覚情報や状況などから“それ”が何であるのかを理解して行動に移すことが可能だ。これと同じようなことを、中山氏は「コンピュータでも可能にしたい」と考えている。

「この実現のためには、マルチモーダルに技術を組み合わせていくことが必要不可欠です。また、同様の取り組みとしては、機械翻訳で画像認識を活用する研究も進めています。例えばWebなどのニュース記事を翻訳する際、テキストだけなく写真や図解も認識し、その情報も活用することが可能になれば、翻訳の精度をより高められるわけです」（中山氏）

東京大学 大学院情報理工学系研究科 創造情報学専攻 中山英樹氏



数百種類の動植物の種類を識別する詳細画像識別のイメージ



Webニュース記事のマルチモーダル要約システムのイメージ

そのほか、ディープラーニングのブレイクスルーからもうすぐ10年が経ち、「いろいろなことがわかってきた一方で、状況としてはだいぶ落ち着いてきた」と感じる中山氏。特定の限られた範囲の中ではなく、状況が目まぐるしく変化するオープンな環境でも利用でき

るようなディープラーニングの新しい仕組みの研究にも着手している。具体的には、「少ないデータでどうやって学習させるか」や「過去の経験をどうやって新しい物事の学習に活用するか」といったことに加え、「人間がどうやって教えていくべきか」という点にも考えを巡らせて研究し、「世界で本当に役立つ新しい知的システム」（中山氏）の構築を目指している。

研究のための計算資源はあればあるほど嬉しい

ディープラーニングにまつわる基礎的な研究から将来を見据えた一歩先を行く研究まで、様々な取り組みにチャレンジする中山氏にとって、GPUを搭載する計算資源は「必要不可欠」と言っても過言ではない。さらに言えば「いくらあっても足りることはなく、あればあるほど嬉しい」とも感じており、その思いは中山氏の研究室の学生も同様。仮の話とはいえ、現状の倍の設備があったとしても「学生には『足りない』と言われるかもしれませんね」と中山氏は思わず苦笑する。それだけに、高性能な最新モデルが登場したのであれば「できるだけ早く導入して活用したい」とつねづね考えてきた。

このような背景から2021年3月、中山氏は自身の研究室に最新のAmpereアーキテクチャを採用したGPU「NVIDIA A100」を搭載するワークステーション「NVIDIA DGX Station A100」を導入した。ラックマウント型のサーバータイプであれば、より高性能なモデルがあることは中山氏も承知しているが、「身近な環境に置いて利用する」という点を重視し、スタンドアローンタイプの「NVIDIA DGX Station A100」を選択。サーバータイプと違って「特別な電源や空調といったインフラ設備を用意する必要がない」ほか、「導入や管理がしやすい」「コストパフォーマンスに優れる」などを導入のポイントに挙げた。

性能も使い勝手も高評価導入した意義は非常に大きい

今回導入したシステムでは、GPUメモリが40GBの「NVIDIA A100」を4基搭載。GPUメモリの合計が160GBとなり、システムとしてのGPU性能は非常に優秀かつ魅力的だ。もちろん、これには中山氏も大満足だが、それ以外にも「各パーツの性能バランス」に着目。例えば、データの前処理などはCPUやストレージの性能に大きく依存するため、GPU以外のパーツの性能が不十分だと「GPUが持つ本来の性能を活かしきれない」というケースが起こり得る。その点、「NVIDIA A100」はCPUやストレージにも高性能なパーツを採用しており、スムーズな処理が可能だ。

「全体のパッケージとして、とても上手く設計されたシステムだと感心しました。『さすがはNVIDIA』といったところでしょうか。研究をスピーディに進められるとあって、学生にも好評です。また、計算や処理に必要なソフトウェアが充実している点もうれしいところ。仮想環境を構築するためのツールであるDockerなども、最新版をパッケージでそのまますぐに使える環境が整っているというのは、とても助かります」（中山氏）

さらに、スタンドアローンタイプならではの使い勝手として、「高い自由度で使える利便性の高さ」も見逃せないポイントの1つに挙げる。例えば、中山氏の研究室では産業技術総合研究所の人工知能処理向け計算インフラストラクチャ「AI橋渡しクラウド（AI Bridging Cloud Infrastructure；ABCI）」やそのほかのGPUクラウドサービスを利用することは可能だが、学生も含めて必ずしも積極的に利用しているわけではない。なぜなら、ジョブ管理システムの影響で利用制約がかかるほか、通信環境によっては思ったほどのスピードが出ないケースもあるからだ。

しかし、研究室専用の「NVIDIA DGX Station A100」であれば、クラウド型と違っていつでも自由に利用できるし、ストレージも含めて安定したスピードが出てくれる。そういった意味でも「導入した意義は非常に大きい」と中山氏は語る。

中山英樹研究室に導入された「NVIDIA DGX Station A100」は、GPUメモリが40GBの「NVIDIA A100」を4基搭載する

理論やソフトだけでは片手落ちハードについて考える必要もあり

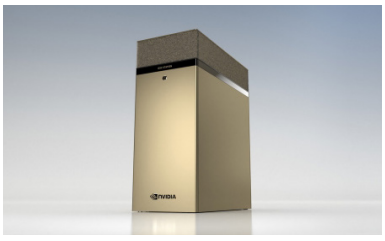
新たな計算資源を手に入れ、これまで以上に研究を加速していく中山氏。最終的な目標は「人間における知能のメカニズムを解明すること」にある。そのための取り組みとして現在は「人工知能の実現」を目指し、その理論やソフトウェアの研究を進めているが、その一方で「ハードウェアの重要性」についても言及する。

そもそも、ディープラーニングによるブレイクスルーは「ハードウェアの進歩」に引っ張られる形で起きた。そういった背景を踏まえれば、理論やソフトウェアとともに「ハードウェアについてもしっかり考えていかなければ片手落ちとなる」と中山氏は指摘。計算資源を含む最新テクノロジーを上手く使いこなすことで「これまでにない新しい知能のメカニズムに迫っていきたい」と考える中山氏は、NVIDIAにさらなる高性能なGPUや使い勝手の良いワークステーションなどの登場を期待した。

東京大学 中山英樹研究室の使用モデル

NVIDIA DGX Station A100

メモリバンド幅の広いHBM2E（High Bandwidth Memory 2E）のGPUメモリ40GBを採用したAmpere世代GPU「NVIDIA A100」を4基搭載するワークステーション。各GPU間は、200GB/Secの第三世代NVLinkで接続されている。デスクサイドに設置できるサイズ感で、100V駆動に対応する。



株式会社ジーデップ・アドバンス
〒105-0014　東京都港区芝1-12-7　芝一丁目ビル4階
TEL：03-6803-0620　URL：http://www.gdep.co.jp/