

特集3

クラウドコンピューティングは 研究のパラダイムシフトを 加速するか

遠い昔から、人類は好奇心に突き動かされ、自然界のしくみを明らかにする取り組みを続けてきた。様々な偶然と、ひらめきと、不断の努力によって、研究者たちは多くの自然現象を理学的に説明付け、工学的に応用することに成功した。さらに探求の矛先は人類社会や脳活動にも向けられ、あらゆる事物を解明せんとする科学の活動は広がり続けている。

今、その活動に転換期が訪れつつあるかもしれない。仮説を立て、検証し、結論付けるこれまでの研究手法は、前提としてその仮説やモデルを“ひらめく”ことが必須であった。だが、多量のセンサー群から得られるデータ、大量のゲノム情報、千数百億個あるニューロン接続、Web上で行われる無数の活動など、いわゆるビッグデータの取得が可能になった現在、我々の頭脳だけではそれら进行处理しきれなくなっているのだ。仮説駆動型の科学(hypothesis-driven science)からデータ駆動型の科学(data-driven science)への転換が、分野ごとに起き始めている。

本特集では、第30回リバネス研究費日本マイクロソフト賞で本賞および副賞の採択を受けた4人の若手研究者に取材を行った。それぞれに分野は異なるが、今後Microsoft Azureによるコンピュータリソースを活用し、大量のデータを処理して、それぞれに思い描く研究を進めようとしている。クラウドコンピューティングがデータ駆動型科学をどのように押し進めるのか、本特集を通じてその実際を感じていただきたい。

第30回リバネス研究費日本マイクロソフト賞とは

「データ分析を行う、あらゆる研究テーマ」に対して、総額約1億円分のクラウド利用権が提供された。必要に応じて数千台の仮想マシンを稼働させ、大量のデータを高速に分析することで研究をサポートするべく設置された研究助成。すでにマイクロソフト(本社)は「Azure for Research」というプログラムを世界中に展開しており、Microsoft Azureが研究の時間を短縮するだけでなく新しい発見をするためのコストを大きく低減することを確認している。

Microsoft Azureが、あなたの研究を加速します

Microsoft Azureは、分析、コンピューティング、データベース、モバイル、ネットワーク、ストレージ、Webなどを統合した、現在急成長を遂げているクラウドサービスです。

マイクロソフトの研究部門は、Artificial intelligenceをはじめとした様々な研究活動を行っています。Microsoft Azureはそういった研究活動の知見を生かし、開発された研究者のデータ収集・分析処理をサポートします。コンピュータサイエンスが専門ではない方でも使いやすいツールとなっており、Microsoft Azureを使ってこれまでできなかったデータ分析に挑戦できます。Microsoft Azure Webサイト <https://azure.microsoft.com/ja-jp/> クラウドを利用した研究事例や、研究者向けの情報 <http://aka.ms/rdcloud>

ユーザビリティの高いツール開発で、 分野を超えた研究者の協働を実現する

東京工業大学情報理工学院情報工学系

大上 雅史 助教

がんや自己免疫疾患の治療薬としてタンパク質間相互作用を阻害する薬剤が注目されているが、創薬ターゲットとなるタンパク質やタンパク質間相互作用の探索は容易ではない。例えば、遺伝子発現解析などで疾患に関連が深いタンパク質が500種示唆されたとき、相互作用の有無を網羅的に解析するには125,250ペアの検証が必要だ。東京工業大学の大上雅史助教は、10万種あるヒトタンパク質の立体構造データを効率的に利用して相互作用を予測する新しいクラウドシステムの開発に取り組んでいる。



タンパク質間相互作用予測ソフトウェア MEGADOCK

タンパク質間相互作用を予測するため、詳細な分子シミュレーションを行う方法や、既知のタンパク質複合体構造との類似性を利用する方法などが開発されてきた。これに対して大上氏が開発した“MEGADOCK”は、タンパク質をボクセルという立体ドット絵のようなラフな構造に変換することの特徴とする。この状態で形状や物理化学的性質を関数化し、相手となるタンパク質も同様にボクセルから関数に変換する。2つの関数どうしを平行移動して重ね合わせを行うことで、立体構造どうしがカチッとはまるかどうかを評価するのだが、このときに高速フーリエ変換を施すことで大幅に演算を削減できる。この方式により、東工大 Tsubame スパコンの計算ノードを420基利用した計算で、100万ペアの相互作用予測を約半日で完了することができた。

誰もが使えるツールへ昇華する

実際に MEGADOCK を利用して非小細胞肺癌に関係する未知のタンパク質間相互作用を探索した結果、約370万件の計算から新たな相互作用が複数予測され、うち6ペアは実験でも相互作用が認められるという成果に繋がった。しかし、大上氏は満足していない。現状の MEGADOCK はコマンドライン（CUI）上での操作が必

要であり、CUIに慣れていない生物学者にはハードルが高いのだ。これをパブリッククラウド上で直感的に動かせるフリーソフトウェアの形にして多くのユーザーに使ってもらえれば、大学のスパコンを使うよりも気軽に予測計算を進められるし、生物学者の専門的知見をもとにしたシステムへのフィードバックもかけられるはず。本当に役立つ技術とは、ユーザーが必要だと思ったとき即座に使えるものだというのが、同氏の考えだ。「誰でも、気軽に、ひらめいた瞬間にすぐに試してみることができる。そんな予測ツールを目指したい」。そう考え、今後 Microsoft Azure 上にシステムを構築する予定だ。

情報科学と生命科学の橋渡し役に

ユーザー視点に立ったツール開発の背景には、異分野の研究者が協働することの重要性を実感した経験がある。「例えば、予測計算でどんなに高い相互作用スコアをマークしても、細胞内局在が違って、実際にそのタンパク質どうしが出会わないのであれば相互作用が起きることはない。でも、そういったことは生物学の知識がないと気づけないんですね」。開発側とユーザー側両方の視点が混ざり合うことで、より深い考察が可能になる。生物学者と情報学者が専門用語や知識の垣根を超えて、真の協働を実現できるシステムを大上氏は新たに生み出そうとしている。（文・中嶋香織）