若手研究者から「生物物理の未来に寄せて」

越境するのか？ されるのか？ 4次元ポケットを介した創造

根本 航¹, 内古閑伸之²

¹産業技術総合研究所生命情報工学研究センター

²東京工業大学大学院情報理工学研究科

1. 散在する多種多様な大規模データの統合化

昨今の実験技術の急速な発展により、核酸やタンパク質の配列・立体構造・相互作用・代謝パスウェイなど多種多様な生物学関連データがデータベースとして公開されている。この状況はデータを選択的に収集、解析し、仮説を構築することで研究を行っているわれわれには非常に望ましい。しかし大規模データが世界中に散在するため、それらの把握・取得・処理をどうすべきかという問題にも直面している。

実はこの問題は大規模データを取り扱うすべての分野において共通に認識されている。そこで、散在する大規模データを取り扱う科学研究の新たな方法論として、Data-intensive computing が提唱されている¹⁾。この方法論により大規模データから特徴や相互関係などが抽出され、新たな知識が創造されると期待されている。そのためにはまず、散在するデータ同士を関連づけ組み合わせる（統合する）必要がある。多種多様な大規模データの統合には Resource Description Framework (RDF) を中心としたセマンティック・ウェブの利用が有効とされている。そのためのオントロジー整備や記述様式の標準化を経て、さまざまな分野で統合作業が進められている²⁾。結果として、近い将来われわれの前にはあらゆる分野のデータが直接的／間接的に関連づけられた“統合情報空間”が出現するであろう。そこにはあらゆる分野のデータが、データの意味を表すデータと共に格納されており、SPARQLなどの言語を使用すれば誰もが自由に、さまざまな組み合わせで取得することができる。

2. 情報統合化以降

このように誰もが望むデータを無料で容易に取得できる近未来においては、2つの大変化が予想される。1つ目は分野間の垣根が限りなく低くなることである。たとえば地球環境に関連する分野では、これまで水質研究者は水質を、地質研究者は地質を研究して

いた。しかしデータ統合とその応用が先行的になされた結果、地球環境に関連するさまざまな分野由来のデータを横断的に取り扱うシミュレーションモデルが作成された³⁾。将来、ここに環境メタゲノムや疫学のデータが統合されると、疫学データの解析に際して気候変動データも考慮できることになる。このように考えると、50年前生物学へ越境した物理学者が創設した生物物理学が、統合情報空間を介して他分野から越境される可能性も十分に考えられる。たとえば、地球環境分野の研究者が統合情報空間を介し生物物理学的研究を行うことも可能になる。このように統合情報空間は分野間の垣根を低くし、各分野における問いを多様化させる。2つ目は、データ解析技術の発展に伴い、データの意味やデータ同士の関係をコンピューターが理解できるようになり、データ取得プロセスが自動化されることである。データ統合に利用されるセマンティック・ウェブでは、データにタグを付加することでデータの意味が明示され、コンピューターによる自動データ収集が可能になると考えられている。よって、パターン化された問いについては指標とするタグの組み合わせを記憶しておくことで、統合情報空間からのデータ取得が自動化されるだろう。

これらに加えて現在の最先端研究では、仮説構築から実験系の構築と実施、結果の解釈とそのフィードバックによる仮説再構築の自動化も行われている⁴⁾。2つの大変化とこのような流れを考え合わせると、近未来における研究の姿が見えてくる。パターン化された問いについては、ユーザが統合情報空間に何を知りたいかを伝えさえすれば、幅広い分野からさまざまなデータが自動で取捨選択される。そして取得されたデータに基づいて仮説が自動構築されるだろう。これはあたかも、使い手が望むだけでその目的に適した道具を提供する、ドラえもん の 4次元ポケットを手に入れたようなものではないだろうか。『頭でほしいと

We will Control the Integrated Information Space!

Wataru NEMOTO¹ and Nobuyuki UCHIKOGA²

¹Advanced Industrial Science and Technology, Computational Biology Research Center

²Tokyo Institute of Technology, Graduate School of Information Science and Engineering

※図1は、電子ジャーナル (<http://www.jstage.jst.go.jp/browse/biophys-char/ja/>) ではカラー版を掲載しています。

思っただけで、体の中のコンピューターが作動して、いろんな部品を組み立てて道具が出てくるんだ。その間たったの3秒!』というドラえもん自身による4次元ポケットについての説明⁵⁾は、統合情報空間を想起させる。このような近未来の環境はとても快適そうではある(図1)。

3. 4次元ポケットの使い手として

問いの形式が定まり、それに答えるために必要なデータも既知の場合、データの取得はわれわれよりもコンピューターが得意とするだろう。予め設定された仮説群から1つを選択することも、コンピューター向きである。しかし未知の形式の問いを發し、それに答えるためのデータを取捨選択し、どのような切り口でデータを観察するかを決断し、新たな仮説を構築するという一連の創造的な作業は当面、人に委ねられるだろう。多種多様な情報が統合されてもそこに意志や価値判断は存在しないからである。4次元ポケット(統合情報空間)から取り出される道具(取捨選択されたデータ)は、われわれが巧く使いこなせば望ましい効果(よい仮説)を得ることができる反面、利用法が不適切な場合は大きな失敗を引き起こす。もちろん、4次元ポケット自体がユーザの望みを察してくれることはない。価値の低い問いに対してはそれ相応の答えが返ってくる。データの取捨選択を誤れば、問いに対して不適切なデータが取得される。したがって、われわれは4次元ポケットの賢明な使い手にならなければならない。価値ある問いを發し、適切なデータに到達

し、それらを組み合わせる必要がある。逆にパターン化された問いしか發することのできない者、解析技術偏重でデータの意味を軽んじる者、データ選択が得意な者はコンピューターに取って代わられる。また、統合情報空間を使いこなせない分野は一方的な越境を許すことになる。

価値ある問いを發し続けるためには、異分野に勇猛果敢に踏み込み、自らの分野に活かすための技術を習得する必要がある。革新的な技術を有する賢い使い手にとって統合情報空間は楽しみの尽きない愉悅の場所である。この点において、生物物理学者は数歩先んじていると思う。他の学会に属する研究者を見ると、データの生産者と解析者は必ずしも一致しない。これはデータ生産者の解析に関するリテラシーに起因する。次世代シーケンサーの普及に伴う解析技術者への急激な需要の増加は、このことを如実に表している。したがって、そのような学会から他分野への越境は起こりにくいかもしれない。一方、生物物理学会には、バックグラウンドが数学・物理学で、データ解析技術を身につけた研究者が比較的多いようだ。かつて物理学を武器に生物学に越境を果たしたように、生物物理学者は今またデータ解析技術を武器に統合情報空間を介して他分野へ越境し、価値ある問いを創造すると期待する。

謝 辞

本稿執筆にあたり、ライフサイエンス統合データベースセンターの岡本忍氏に貴重な情報をいただいた。

文 献

- 1) Tohy, Hey *et al.* (eds.) (2009) The Fourth Paradigm: Data-Intensive Scientific Discovery.
- 2) http://biosciencedbc.jp/gadget/rdprog_over/yonezawa.pdf
- 3) DIAS: <http://www.editoria.u-tokyo.ac.jp/dias/>
- 4) King, R. D. *et al.* (2009) *Science* **324**, 85-89.
- 5) 続ドラえもん全百科 (1979) 藤子不二雄監修, p. 19, 小学館, 東京.

図1 情報統合前後における大変化。統合前:さまざまな障壁により生物物理学者が取得できない情報が存在する。統合後:統合情報空間からの取捨選択を容易に行うことができる。

根本 航

根本 航(ねもと わたる)

産業技術総合研究所生命情報工学研究センター(CBRC)産総研特別研究員

2005年9月奈良先端科学技術大学院大学情報科学研究科博士後期課程修了(博士(理学))。日本学術振興会特別研究員(PD)を経て07年4月より現職。研究内容:立体構造情報生物学

連絡先:〒135-0064 東京都江東区青海 2-4-7

E-mail: w.nemoto@aist.go.jp

内古閑伸之

内古閑伸之(うちこが のぶゆき)

東京工業大学大学院情報理工学研究科計算工学専攻産学官連携研究員

2001年東京大学大学院総合文化研究科広域科学専攻修了(博士(学術))。日本学術振興会特別研究員、産総研CBRC共同研究員を経て現職。

研究内容:タンパク質複合体構造形成の生命情報学研究
連絡先:〒152-8552 東京都目黒区大岡山 2-12-1 W8-76

E-mail: uchikoga@bi.cs.titech.ac.jp