

ビッグデータは 自然科学のあり方をも変える

中西崇文（国際大学 GLOCOM 主任研究員/准教授）

ビッグデータという言葉が使われ始めてから、すでに5年以上が経つ。ビッグデータという言葉が、誰がいつ使い始めたのかについては諸説あるが、2011年頃から急激に広がったことは確かである。そして、今、ビッグデータがビジネス界において、非常に大きなインパクトを与えていることに疑いの余地はない。さらにそれだけではなく、ビッグデータの流れは、自然科学のあり方をも変えようとしている。本稿では、ビッグデータが、実際、どのように自然科学のあり方を変えつつあり、それによってどのような進展が期待されるかを述べることにする。

◆ データが現実を表す解像度を手に入れた

現在、Internet of Things (IoT) により、ビッグデータは、より増大している。このような環境によって、何がかわるのだろうか。

端的に言えば、「データが現実空間を表すだけの十分な解像度を手に入れた」ということができる。大量のデータによって、現実の事細かな様子まで表すことが可能となった。多様なデータによって、現実を様々な角度から多角的に表すことが可能となった。高頻度に更新されるデータによって、現実で起こる様々な変化を随時くみ取り、表すことが可能となった。現実空間で起こった事実は、サイバー空間において、データにより、詳細に、そして連続的に観測できるようになったのである。データにより、逐次、現実社会を蓄積することが可能となれば、現実空間においては誰も気づかなかった事象をも、サイバー空間に蓄積されているデータによって、仮説や推定なしに捉えられる可能性がある。

これにより、データ分析手法も変わる事となる。従来のデータ分析は、仮説に基づいてサンプリングしたデータを対象とする。つまり、母集団に含まれる数と、サンプリングされたデータ分析対象の数に

は大きな差があるため、分析をした後、その結果が母集団でも成立する有意なものであるかを検証する必要があった。それに対して、ビッグデータの場合は、現実空間を表すだけの十分な解像度を持っているため、データ分析対象を母集団と見なすことができるようになった。つまり、分析結果をそのまま現実の結果として見るができるということである。

◆ 新たなパラダイム、データ中心科学

データ分析手法が変わることで、新たな科学が生まれつつある。データ中心科学¹である。データ中心科学とは、最初からデータを分析することで問題を解決するアプローチを指す。最初からデータを分析するとは、モデルを作ることなく、データスフィア（データ全体）が現実を表現する全てと捉え、なぜ、そういう事象になったのかを検証していく科学である。

データ中心科学に移行することにより、もっともらしい仮説を立てることや、その仮説を証明するためのサンプルを集めるような必要がなくなる。その代わりに、現実をもれなく表現する膨大で様々なデータを随時取得し、蓄積、分析する必要がある。これは、母集団と取得したデータがイコールにならない限り成立しない。また、それらの膨大なデータを直感的に把握するための可視化手法も非常に重要な要素となってくる。データ分析と新たな知見の発見が同時に行われていくのである。

◆ 人工知能技術はモデルを超えていく

さらに、人工知能技術は、データ中心科学をドライブするツールのひとつになり得る。特に、「教師あり学習」と呼ばれる手法に属する、過去のデータから現状を判断したり予測したりするものが注目され

ている。

これまでのシミュレーションを中心とした手法では、人間がモデリングを行う。ここでいうモデリングとは、ある着目している現象を定式化することと考えてよい。モデリングされたものにデータを当てはめることにより、現状で起こっていることの把握や予測を行っていた。

これに対し、人工知能技術は、上記のようなモデリングを必要としない。人工知能がデータを与えるだけで、現状を示したり、予測を示したりしてくれる。人間は、着目した現象について、どのようになっていくか、ある意味、仮説的な定式化を行わなくてよくなる。モデリングなしで新たな知見の発見があり得る。つまり、特に自然科学系の研究者が、これまで行ってきたことを必要としなくなるということだ。もちろん、人工知能技術自体がブラックボックスになってしまうという欠点があるが、あきらかに、人間の従来型の研究者を必要としないソリューションが生まれつつある。人工知能技術は、これまでの研究のあり方を変えるツールとなり得る。

これからの自然科学研究者は、ある現象を定式化、モデリングするだけではなく、過去のデータを基に、人工知能技術というツールを用いて、現在・未来について表現する力が必要となるだろう。

◆ シミュレーションも PDCA サイクルで進化

センサーが現実的に設置可能であり、データによって観測可能な部分においては、データ中心科学の波がこれからも広がっていくと考えられる。

もちろん、実際には、センサーが十分に設置できない、もしくは現状のセンサーの性能では十分に観測できない、ということも多い。例えば、気象の場合、台風を観測できるのは、おもに地上、もしくは衛星からである。もし、台風のときの雲の断面を実際に観測しようとする、現状では不可能、もしくは多大なコストがかかってしまう。

そのような場合は、シミュレーションを用いて、その値を推定することが必要になるだろう。ただし、シミュレーションは、従来の仮説や作り上げたモデルの妥当性を検証するためのものではなく、新たなデータを生成し、より現実空間の事象を解像度高く表すためのものに、その目的が変容しつつある。また、シミュレーションによって生成されたデータは、新たに取得した別のデータによって、妥当な値であ

るかを検証することができる。妥当な値であれば、シミュレーションから出されたデータもビッグデータの一部として用いることができる。また、この妥当性の検証によって、差異が生じた場合、その差異を補間するように、モデルを書き直すことも考えられる。データがリアルタイムに入手でき、検証が随時行われ、さらにその検証に応じて自動的にモデルが書き直される。ここに、データ分析の PDCA サイクルが生まれる。このようなデータ分析の PDCA サイクルにより、より詳細な現実を描くモデルが随時生成でき、それによって、これまで見えてこなかった詳細な知見を得ることも可能となるだろう。

◆ まとめ

データはまだまだ語る。ビッグデータの収集・分析を行うことにより、現実空間をサイバー空間で表現する際の解像度はさらに上がっていくだろう。これまでは、ぼやけて見えていなかった事象が、くっきりと見えるようになる。

これまでの科学のほとんどは、マクロで観測し得る事象を表すモデルと、ミクロで観測し得る事象を表すモデルとが、独立して構築されており、これらを同時に扱うことが難しかった。データが十分に取得可能となり、解像度が十分になれば、まるで解像度の高い写真を綺麗にズームイン・ズームアウトするかのよう、マクロで観測し得る事象とミクロで観測し得る事象を同時に扱うことが可能になるかもしれない。実践例としては、Multi-Scale Simulator for the Geoenvironment (MSSG)ⁱが挙げられる。MSSGは、地球シミュレータ上で実現される大気と海洋を結合した予測シミュレータであり、全球スケールから日本周辺、建物を解像した都市・街区スケールまでの予測計算を階層的に精度よく実現するものである。

ビッグデータは単なるバズワードではなく、自然科学を変えうるパラダイムシフトである。

ⁱ 中西崇文 (2015) 『スマートデータイノベーション』、翔泳社

ⁱⁱ Keiko Takahashi, Akira Azami, Yuki Tochihara, Yoshiyuki Kubo, Ken'ichi Itakura, Koji Goto, Kenryo Kataumi, Hiroshi Takahara, Yoko Isobe, Satoru Okura, Hiromitsu Fuchigami, Jun-ichi Yamamoto, Toshifumi Takei, Yoshinori Tsuda, and Kunihiko Watanabe(2011). "World-highest resolution global atmospheric model and its performance on the Earth Simulator. In State of the Practice Reports," (SC '11). Association for Computing Machinery, New York, NY, USA, Article 21, 12 pages.