

データサイエンスのための人材育成

東京大学数理・情報教育研究センター

明治大学先端数理科学インスティテュート

北川 源四郎

学術分野および社会において大量・大規模なデータが集積し、ビッグデータ時代が到来している。これに伴って、データから価値を引き出すための方法論としてのデータサイエンスが今後の社会発展の鍵として重要になっている。そのため、高度なデータサイエンスを推進できる専門人材育成とリテラシーの涵養が重要となっている。本稿では、データサイエンスのための人材育成の取り組みを報告する。

1. ビッグデータとデータサイエンス

情報通信技術や計測技術の飛躍的発展と普及によって社会や科学技術の多くの分野で大規模データが取得され、ビッグデータの時代が出現している。近年のデータ増加は、過去の物質世界の増加とはけた違いの早さで進行しており、1.5年で2倍、10年で100倍になるといわれるIT機器の集積度の増加（ムーアの法則）をも凌ぐ速度で蓄積している。その結果、社会では超スマート社会が実現し、Googleなどのデータ駆動型の産業の出現、すべての産業のICT化、エネルギーや交通システムなどの社会インフラのスマート化、医療・創薬・教育などにおける個別サービスなどが実現しようとしている。また、科学技術においては今後すべての研究はデータサイエンス化するとも言われている。

実際、社会においては既に通信、情報検索、情報発信、販売、流通、犯罪捜査などの分野で既に劇的な変化が起こっており、特に近年は深層学習や人工知能の応用により、コンピュータが人間の専門技能をも超える場面がしばしば生じている。このような急激な変化は、蒸気機関の出現によって惹き起こされた産業革命の再現ともいえ、しかもその変化は馬車がT型フォードに僅か10年程度で置き換わってしまったように、多くの人々が考えているよりも遙かに急速に進展する可能性が高い。

これに伴って過去数百年の歴史を持つ研究の方法も研究スタイルも急激に変化しようとしている。ビッグデータ時代にはそれにふさわしいデータ駆動型の研究方法論の確立が必要である。20世紀までの科学研究は実験科学と理論科学の二つの方法論に支えられてきたが、20世紀後半には計算科学が確立し、複雑な非線形システムの理解とシミュレーションにおいて大きな成果を挙げつつある。そして現在では、大規模データの出現によってデータサイエンス（データ科学）が第4の科学的方法論として重要になっている。実験科学と

理論科学がそれぞれ研究者の個人的能力に依拠した帰納的方法と演繹的方法なのに対して、計算科学とデータサイエンスは計算機の圧倒的な能力が拓いた新しい演繹的方法と帰納的方法と位置づけることができる。

この変化は「認識の科学」から「設計の科学」への転換を後押ししているものということができる。19世紀までの研究は主に物理現象を対象としたものであったが、進化論のインパクトもあって、20世紀に入ると生命や経済現象など進化し変化する対象も重要な対象となった。更に21世紀の現在、実世界とサイバー世界が一体となったCPS (Cyber-Physical System) が重要な研究対象となっている。このように研究対象の変化に伴って、研究目的も、従来の普遍の真理の探究から、予測、意思決定、価値創造などが重要になっている。更に、研究に用いられるモデルも第1原理から導かれた物理モデルから、直接の目的を達成するために対象の実態ではなく対象の特定の機能を目的に応じて表現するモデリングへと変わりつつある。

更に近年には、画像認識、音声認識、分類、自動翻訳、トラッキング、自動運転などで深層学習やAIが華々しい成果を挙げ、爆発的なブームとなっている。しかし、ブラックボックス型のこれらの方法が帰納的方法の限界をどこまで突破できるかは、数理的方法やデータサイエンスの今後の方法にも係る問題であり、注目していく必要がある。

2. データサイエンティスト育成のための施策

このような状況の中で、海外では今後のデータサイエンスを担うべき統計学やデータサイエンスの人材育成が急速に始まっている。これに対応し我が国では、日本学術会議では2014年に提言「ビッグデータ時代に対応する人材の育成」において、①データ中心科学を専門とする教育組織の設置、②基幹的研究組織内における恒久的なデータ解析部門の設置、③日本版インサイト・プログラムの早急な設置、④データサイエンティストの資格の認定を提案している。

さらにこれを受けて情報・システム研究機構に設置されたビッグデータの利活用に係る専門人材育成に向けた産学官懇談会ではその報告書において、データサイエンティストへの産学官からの要請を調査し、データサイエンス人材のレベルをデータリテラシー、見習い（基礎能力）レベル、独り立ちレベル、棟梁レベル、指導的データサイエンティストの5段階に設定し、それぞれ育成すべき人数を年間50万人、5万人、5000人、500人、50人としている。

また、この報告書では、それぞれのレベルの人材育成のための具体的施策を示すとともに、我が国においては棟梁レベルの人材が育っていないことが最も深刻な問題であることから、早急に棟梁レベルの人材育成に注力すべきことを提言している。また、棟梁レベル

の人材育成が実現すれば、彼らが人材育成を支援することによって、独り立ちレベル以下の人材育成が促進され、スケールアウトが進むというトリクルダウン効果についても指摘している。

これらの報告の下で、この懇談会の報告書「ビッグデータの利活用のための専門人材育成について」では、①年間 500 名の棟梁レベルの人材育成、②リテラシーレベルから独り立ちレベルの大学教育を加速させるために、全国 10 大学程度で人材育成をスタートし、全国への波及効果を狙う、③国家レベルのフラグシップ・プロジェクトを推進すること、の 3 項目の提言を行っている。

3. 数理・データサイエンス教育強化拠点コンソーシアム

このような動きを受けて、文部科学省高等教育局では「数理及びデータサイエンス教育の強化に関する懇談会」を設置して、数理・データサイエンス教育の強化方策を定め、2016 年度に数理・データサイエンスの教育強化のためのセンターを全国 6 大学に設置した。これらのセンターは、全学的・組織的に一般教育および専門基礎教育を実施して、数理的思考やデータ分析・データの活用能力を持ち、課題発見、問題解決、データからの価値創出ができる人材を育成するための教育体制構築を行うとともに、地域や分野の拠点として他大学への展開・波及を目指すことを目的としている。

さらに、データサイエンス教育の全国普及のために、これらの 6 大学は数理・データサイエンス教育強化拠点コンソーシアムを設立して

- ・標準カリキュラムの設定（スキルセットおよび参照基準の策定）
- ・共通教材の作成（教科書シリーズの編集や e-ラーニング教材、講義動画の作成・公開）
- ・教育用データベースの構築

に着手し各大学内における数理・データサイエンスの全学教育を効果的に推進するとともに、全大学への普及を目指している。

参考文献

- 日本学術会議 情報学委員会 E-サイエンス・データ中心科学分科会、「ビッグデータ時代に対応する人材の育成」、2014 年 9 月。
- 情報・システム研究機構 ビッグデータの利活用に係る専門人材育成に向けた産学官懇談会、「ビッグデータの利活用のための専門人材育成について」、2015 年 7 月。
- 文部科学省高等教育局 数理及びデータサイエンス教育の強化に関する懇談会、「大学の数理及びデータサイエンス教育の強化方策について」、2016 年 12 月。