

ビッグデータを活用できる人材の育成

情報・システム研究機構 北川 源四郎

情報通信技術の飛躍的発展によって、学術研究分野や社会においてデータが爆発的に増加し、その有効活用が今後の学術研究や産業発展の鍵となっている。このビッグデータの登場は、産業革命に次ぐ大きな社会変化をもたらすものと言われており、研究の領域でも科学的方法論、研究基盤および研究人材育成のすべてについて抜本的な変革が必要になっている。

実際、ビッグデータが内包する価値は膨大とはいえ、その価値密度は非常に低く、データ量の増加に比例して人間の知や経済価値が増加するわけではない。これは膨大なデータから重要な情報や知識を取り出すことができ始めてビッグデータの恩恵を享受できることを意味している。この問題を解決するためには、科学的方法論に関しては、従来の実験科学と理論科学さらには近年急速に発達したシミュレーションを中心とする計算科学に加えて、大規模データからの知識獲得のための第4の科学（データ中心科学）が不可欠となる。一方、研究の基盤としては、データと知識の共有・統合・活用による知識発展を目指すオープンサイエンスを視野に入れた基盤構築が必要である。そして、ビッグデータを活用するために重要となる人材がデータサイエンティストである。

本報告では、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による、ビッグデータ時代に対応できる人材の育成に関する提言の内容を中心に説明する。ビッグデータから有益な情報を取り出すためには、従来の統計学に加え、機械学習と最適化のスキルが必要である。また、現実の課題解決で直面するビッグデータ解析のためには、従来の統計ソフトや言語の利用能力を超えて、並列計算ツール等のビッグデータ工学に関するある程度の知識も必要になる。さらに、従来の統計研究者と同様、現実の課題解決のためには、当該分野の知識も必要となることから、データサイエンティストの育成においては、T型・II型研究者を目指すことが必要となる。

データサイエンティストには、ビッグデータ解析の専門家に止まらず、学術研究やビジネスにおいて現実の課題やその背景をよく理解した上で課題解決のための全プロセスを企画し、遂行できる高度な能力を持った人材であることが期待されている。日本学術会議の提言においては、そのために必要な要件として、戦略立案能力、異分野研究者・事業者との連携能力、データ分析結果の事業等への実装能力、研究倫理など9つの条件をあげている。このような要素技術と異なる能力を如何に身につけるかも重要な課題である。

現在行われている人材育成に関する産官学懇談会においては、データサイエンティストをいくつかの階層に分け、それぞれの階層で必要なスキルやその育成方法の検討も行っている。学会では、日本学術会議の提言内容とともにこのような議論の内容も報告する予定である。

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提言「ビッグデータ時代に対応する人材の育成」2014年9月