

データサイエンスに求められるもの

○ 林 篤裕

名古屋工業大学 大学院工学研究科

1 取り巻く状況

データサイエンスという言葉が何時ごろから使われるようになったのか、定かには判らないが、近年のビックデータや AI ブームを背景に現在では市民権を得たように思われる。これが何を指すかは人によっても多少異なるのであろうが、今年 4 月に滋賀大学に開設されたデータサイエンス学部のホームページに依ると「社会に溢れているデータから<価値>を引き出す学問」と記されている。従来の言葉で言えば中核をなすのは統計学であろうが、これにとらわれずデータの入り口から加工・分析・出口までを包括的に示す概念を示しているように思われる。

しかし、1980 年代に統計ソフトが普及し始めた際にその誤用に警鐘が鳴らされたようにデータサイエンスについても、ややもするとその言葉と技術が無秩序に利用され理論的背景やデータの持つ意味に注意を払われずに利用されている側面があるのではないか。つまり、統計学では綿密な計画を練ってデータを採取し分析に供してきたものを、近年のセンサー技術等の発展からシステムティックに収集されたものに限らずに結果的に集まってしまったデータを調査対象としているような場合も散見される。また、この種のデータは体積自体は大きいものの、その中に占める測定値の量は疎であることも多い。

2 求められるもの

このような状況下でデータサイエンスに求められるものとしていくつか挙げられるが中でも、調査対象となるデータを丁寧に扱うことの再認識ではないかと考える。データの採取が以前よりは容易に行える時代になったとは言え、導き出される分析結果が素データの質に強く依存していることに変わりはない。調査目的に即したデータが採取できているかを注意深く精査することはデータ分析の基本であることを知ってもらうことであろう。また、採取されたデータの背景情報を熟知することも、測定時に混入するノイズの理解や分析結果の解釈の際にも大いに理解を助けるものになると思われる。他にも統計データに接する際の見方・考え方の教育も重要であろう。加えて大量に生成されるデータを考慮した解析技術の開発も望まれる。

このようにデータの取り扱いを中心に据えた分析姿勢がより質の高い調査結果を得る鍵となり、今後のデータサイエンスの浸透に寄与すると考える。

参考文献

データサイエンスって何?, 滋賀大学データサイエンス学部ホームページ (2017年6月15日現在), <https://www.ds.shiga-u.ac.jp/about/>.

(連絡先: hayashi.atsumuro@nitech.ac.jp)