



contents

[コラム]

ビッグデータ利活用促進とデータサイエンティスト人材育成
…石井一夫

[解説]

実践的情報教育協働ネットワーク enPIT
…井上克郎 楠本真二 後藤厚宏 鵜林尚靖 北川博之

■ 応 般 Column

ビッグデータ利活用促進と
データサイエンティスト人材育成

新聞、雑誌等のメディアではビッグデータに関する記事がほぼ毎日、毎週のように掲載されている。筆者は必要に駆られて 10 年ほど前から、大規模データ解析を開始し、R や、Perl/Python/Ruby、MySQL などを用いたデータ解析を行ってきたが、このようなビッグデータが注目される時代が来るとは思いもよらなかった。当時、私の属していた分野である、生化学・分子生物学は国内では分析化学を基盤とした実験生物学者が多く、統計学や情報科学を基盤としたデータ解析をやっても邪魔者扱いや無能者扱いされたものである。2010 年頃から次世代シーケンサーや質量分析器などの大量のデータ産出（出力）を行う機器が普及してきようやく風向きが変わってきた。

現在、農学部で大規模データ解析を行いたい大学院生にそのノウハウを教授する職に就きビッグデータ分析を伝授する立場にある。私も含めて多くの方は学生時代に十分な数学、統計学、プログラミングを学ぶ機会のないままに、必要に駆られて、あるいは特に意識することなく、いろいろな経緯を経てデータ解析を学びにきている。このため、非常に歪んだ背景知識やデータ感覚を有している。たとえば、今はやりのノンパラメトリック検定という言葉を知っているのに、統計学を十分に学べていないために中心極限定理を知らず統計学的検定を十分に使いこなせていなかったり、 t 検定は理解できてもフィッシャーの正確確率検定を理解できなかったりというケースがある。そんな状況の中での、人材育成である。データ解析についての基礎固めに関して課題はかなり多い。データの取り扱い、特にセキュリティやプライバシー保護を含む倫理面についても、その配慮の欠如から深刻な事故に至ることをしばしば経験しており、管理の徹底を促すこともある。

一方、巷のビッグデータや統計解析に関する勉強会、講習会は盛況で、参加が非常に困難な状況になっている。勉強会への参加者はきわめて熱心で知識の習得にも貪欲であり、習熟度も高い。私自身も大学の講習会を企画して外部からの聴講者を受け入れる場合もあるが、学外参加者が殺到して学内参加者や学生を圧倒するということもある。

私自身が IT フォーラム “ビッグデータ活用実務フォーラム” を設置するという考えに至ったのは、そのようなデータ解析についての人材育成のアンバランスの解消と、情報共有の機会提供の手段として活用したいという思いからである。その思いが少しでも実現することを祈りたい。

石井一夫（東京農工大学ゲノム科学人材育成プログラム）

ロゴデザイン ● 中田 恵 ページデザイン・イラスト ● 久野 未結