

あらゆるモノがネットにつながるIoTの普及に伴い、膨大なデータが世界各地で毎日生み出されている。企業や組織の活動はもとより、一人ひとりの生活や行動に至るまでビッグデータとして記録・分析され、使い次第では生きとし生けるものの幸せに大きく貢献する。有限な資源の「石油」に対して、無限に近い資源の「データ」は正にデジタル世紀が創り出す「新たな資源」である。そのような背景から、データから社会やビジネスのニーズに対応した課題を発見し、問題解決や価値創造に関与できる人材の育成が喫緊の課題となっている。世界からは遅れているが、日本の大学でもデータサイエンス教育への取り組みが始まった。産学連携による教育イノベーションが課題と言われているが、どのような教育プログラムでチャレンジしていくのか、たずねてみた。

## 超スマート社会に向けた データサイエンス人材育成



情報・システム研究機構理事 統計数理研究所長・**樋口 知之**  
データサイエンティスト協会顧問

### 1. 茹でガエルと総取りゲーム

身の回りに起こっている変化について行けなくなっている自分に気がついたとき、私たちはよく「時代の転換点」とか「時代の激変期」と言って、あたかも他人事のように話題にすることが良くあります。話題にすることで、相互にそのことを認識し、茹でガエルのようなことにならぬよう、社会全体で改善を行ってきました。さて、この10年間に起きつつある変化にも、このような対応で十分なのでしょうか？ シニアの方々は、ビッグデータや人工知能（AI: Artificial Intelligence）の本格的到来を熱く語る若者たちを、どこかオオカミ少年として冷やかな見方をしていませんか？

現代生活に欠かせなくなったスマートフォンの登場は、たった11年前です。AIの代表格である深層学習の原始的な形が提案されたのも12年前です。深層学習とは、脳神経のつながりを極めて簡単な数理モデル（非線形の関数）で表し、それらを組み合わせて構成したネットワークのことです。昨今メディアで話題となっている、コンピュータ囲碁ソフト、自動運転、AIスピーカー（グーグル・ホームやアマゾン・エコーの類い）は、この深層学習を基盤に作られています。この10年間は、後世、人間の生き方が変わったメルクマールと間違いなく言われるでしょう。

前述したように深層学習は、数理の観点から言うと非線形関数にすぎません。ただし、膨大な数（億単位）の未定のパラメータを含むため、それらを決定するためにはパラメータ数よりも格段に多いデータ（サンプル）が必要となります。この

からくりは、初等数学で習う、変数と条件の数の関係と同じです。そうすると、いかに多くのデータを集めるかが勝負を決めるため、効率よくデータを集める仕組み作りに企業は注力するわけです。今や世界の時価総額10大企業のうち、モノにかかわる業態は2社のみで、トップ5（ビッグ5と通常呼ばれる）は、アップル、グーグル、マイクロソフト、アマゾン、フェイスブックです。グーグルとフェイスブックにいたっては、その収益のほとんど（ほぼ9割）をオンライン広告で得ています。つまり、ビッグデータを自動的に集め、人々が求める情報を提供できるトップ企業のみが生き残れる、ビジネス的には総取りゲームの時代となっています。

### 2. データサイエンス人材の奪い合い

ビッグ5の研究開発への投資額も“半端ない”です（若者言葉を使ってみました）。少し古い情報ですが、アップルの2013年度の研究開発額は9,000億円ほどで、国立大学全体の運営交付金が1兆1,000億円弱と比べると、その巨額さが分かります。また2017年度、アマゾンはなんと2.5兆円を研究開発に投じていたことも報じられました。この結果、優れたAI技術を持ちつつあるベンチャー企業の、ビッグ5による根こそぎ青田刈り現象も頻発しています。当然、優秀な人材の奪い合いも、米国や中国では激烈となっています。

前述したように、今のAI技術は、深層学習を中心とする統計的機械学習が基盤となっています。統計的と言葉が頭についているように、パラメータの決定や数理モデルの選択は、データにもとづ

いて実行されます。したがって、統計的機械学習の習得に必要な素養は、統計学や最適化が中心となります。統計的機械学習の知識と、その計算機への実装能力（プログラミング能力）を備えた人を、米国ではデータサイエンティストと呼んでいます。また、データにもとづいて意思決定を行うことが基本である米国では、薬や治療法の許認可や政策の決定の場などに統計家と呼ばれる専門職が多数配置されています。統計家は、伝統的な統計学の概念や方法論を尊重しつつ、現代的なデータ環境の中で活躍する人材になります。日本にはこの統計家も非常に少ないという、諸外国からすると驚くべき状況が放置されてきました。

USジョブランキングという、給与、労働環境、ストレスなど総合的観点から人気の仕事を調査するアンケート報告があります。2016年は、データサイエンティストと統計家が1位、2位を占めました。2017年は統計家が1位、データサイエンティストは5位でした。このようにデータサイエンティストや統計家への期待とニーズは世界中で爆増しています。ちなみにこのランキングには不人気ベスト10もあり、2017年のワースト1は新聞記者、2番は放送関係者でした。これらからも、日本が井の中の蛙に近いことが理解できると思います。

### 3. データサイエンス教育プログラムの充実：現場力をつけさせる

学生は、学部や学科を選択するとき、人気職の動向に非常に敏感です。世界的にこの数年、データサイエンスに関する教育プログラムが急増しています。特徴として2点あげられます。一つ目は、教育レベルは修士を対象とするものが中心であること、二つ目は2016年あたりから増加の勢いが落ち着いてきた点です。まず後者については、世界的に見れば、データサイエンスの教育プログラムは、ほぼ社会からのニーズに応える規模になって、今は成熟期に入っていると言えます。他方日本は、ようやくスタート地点に立ったばかりです。

前者は、データサイエンスの特性を語る上で大変興味深い点です。産業界で活躍するデータサイエンティストが備えるべき資質として、データサイエンス力（統計学、機械学習、最適化など）、データエンジニアリング力（プログラミング、データベース、コンピュータなどに関する知識とスキル）、そしてビジネス力の3つがよくあげられ

ます。最後のビジネス力は、現場を知り、現場を感じる、言わば「現場力」に相当し、学部教育で考えると、様々な専門分野固有の知識やスキルにあたると思います。よって、専門分野を習得した学生に、追加的にデータサイエンスを教える方が、学生が興味をもって熱心に取り組んでもらう観点から効果的であることが示唆されます。また、学部レベルでデータサイエンスのカリキュラムを組む際には、この点を念頭に置いて、PBL（Project Based Learning）などの演習の内容を丁寧に設計しなくてはならないでしょう。日本でこれからさらに増えていくであろうデータサイエンス学部、学科が成功する鍵はここにありそうです。

この「現場力」が今後重要性を増す兆候はデータにも見て取れます。IPA（情報処理推進機構）の「グローバル化を支えるIT人材確保・育成施策に関する調査」に、IT人材が、ベンダー企業（ITをビジネスとしている企業）側か、それともユーザー企業側にいるのかをまとめたグラフがあります。それを見ると、中国・インド・日本を含めほとんどの国ではベンダー企業側にいますが、米国は7割がユーザー企業側にいます。このことは、米国以外の国では、ITを活用した「情報サービス」が、ユーザー企業からベンダー企業にアウトソーシングされている一方、米国ではそのプロセスがユーザー企業に内製化されていることを物語っています。つまり、ITのフレームワークがコモディティ化しつつあり、エンドユーザーや消費者に近い、下流にIT技術者がその活躍の場をシフトしつつあると考えられます。米国のこのビジネスの状況は、Society5.0を考える上でも、また日本の将来の姿を予想する上でも大変参考になります。

### 4. 「究める」から「活用する」へ移行する時代に何を学ぶか？

インターネットによってすべてがつながる時代には、あらゆるモノやシステムが、シェアリングとエコシステムの観点で、再検討・再構築されていきます。その動きは先行してビジネス界で顕著で、さらに私たちの生活の奥深くにまで浸透しつつあります。これからの若者には、これまでの記憶力に代わって、人を共感させるアイデアとセンスが最も大切な素養になるでしょう。そして、データサイエンスの知識とスキルが、その若者の夢の実現に欠かせないことも明らかです。