



データエンジニアへの期待

加納 学

加納 学
京都大学大学院 情報学研究科
システム科学専攻 教授

自分が扱うデータがビッグかどうかを気にする必要はない。データから価値を生み出すことが重要であり、それができるスモールデータはとても魅力的である。わざわざ広大な砂漠に砂金を探しに行かなくても、目の前にダイヤモンドの原石があるなら、それを手放さずに入念に研磨すべきである。

膨大なデータから価値ある情報を抽出する試みはデータマイニングと呼ばれ、その分野では随分と前からビッグデータ (Big Data) という用語が用いられてきた。巨大なデータの山から金塊を掘り出す採掘者は、データサイエンティストと呼ばれる。多くの経済人がビッグデータに目を向ける契機になったであろう The Economist (2010年2月25日)の記事“Data, data everywhere”には、新しく誕生した専門職として、データサイエンティストが次のように紹介されている。

“a new kind of professional has emerged, the data scientist, who combines the skills of software programmer, statistician and storyteller/artist to extract the nuggets of gold hidden under mountains of data.”

かつて米国カリフォルニア州がゴールドラッシュに沸いていた頃、採掘に必要なシャベルや鶴嘴 (つるはし)、篩 (ふるい) を売りさばいたのが Samuel Brannan であり、キャンパス地で丈夫な作業ズボンを作ったのが Levi Strauss であり、採掘者が掘り当てた金を換金し、手紙や荷物の輸送、預金や送金といったサービスを提供したのは Henry Wells と William Fargo であった。財を成したのは、採掘者ではなく、彼らである。ビッグデータに埋もれているはずの宝を掘り出そうと、データサイエンティストという採掘者に期待が寄せられる昨今であるが、果たして富を築くのは誰であろうか。

これまでデータ解析にそれほど理解を示してこなかった人達が、突然、ビッグデータで利益を出せないのかと迫ってくるという話も聞こえてくる。もちろん、先駆的な取り組みの実績を聞き及んでのことであろう。インターネット企業やソーシャルネットワーク企業のみならず、

ビッグデータ解析をマーケティングに活用しているところは多く、犯罪予防に役立っている事例もよく知られている。製造業も例外ではない。Industrie 4.0 や Internet of Things (IoT) といった概念が提示され、モノづくりのさらなる飛躍が目指されているが、そこでは、モノ同士がインターネットを介して接続されたときに行き交う膨大なデータの処理が成功の鍵を握る。例えば General Electric (GE) は、世界中で使われているタービンや発電機に取り付けられたセンサから得られるデータを解析することで、機器の予知保全を実現している。この仕組みによる損失回避効果は年間7千万ドルに達すると推計されている。

多数の機器からデータを吸い上げて解析することで予知保全が可能になることは、多数のヒトからデータを吸い上げて解析することで疾病予防が可能になることに似ている。高度なヘルスケアの実現に向けて、医療保険者が保有する健診データやレセプトデータを集約・解析する試みが進められている。さらに、遺伝子情報やウェアラブルデバイスで取得可能な心拍や活動量などのデータも解析対象となっている。まさに医療ビッグデータ時代の到来と言えよう。

では、プラントを対象として、ヘルスケアで期待されているような成果をあげることはできるだろうか。あるいは、GE が実施しているような予知保全ができるだろうか。この問いに答えるためには、前提の違いを認識する必要がある。健診受診者の数は数千、数万に及ぶ。先述の予知保全の場合、監視対象となる機器の数は1,500以上とされている。このように多数のヒトや機器のデータが様々な環境下で収集されているため、そのデータを解析することで、疾病や故障の予兆を検知するモデルが構築できる。ところが、プラントの場合、その多くは唯一無二であり、様々な異常発生時のデータを取得することも、異常状態を表現できるモデルを構築することも困難である。多変量統計的プロセス管理 (Multivariate Statistical Process Control : MSPC) のように、正常状態

を定義するという立場の異常検出方法が活用されてきたのも、このような事情のためである。データ量が多ければ十分というわけではない。加速度的にデータ量が増え、そのデータに十分な情報が含まれていなければ、そこから得られる知識も増えることはない。このようなデータの不十分さを補うために必要なのは、対象プラントについての専門知識 (domain knowledge) である。専門知識とデータ解析の統合利用が成功の鍵を握る。

対象プラントについての知識の在処として、熟練運転員を挙げることができる。ところが、その知識の多くがノウハウとも呼ばれる暗黙知であることから、製造業では2007年問題や2012年問題が技術伝承の観点から議論された。暗黙知を形式知に変換するために、Exapilotのような運転支援システムの活用が進められ、スタートアップの自動化などで顕著な成果をあげている。いつ何をすればよいか、どのような条件が満たされれば次の手順に進めるか、といったことが明らかにされていることが、フローチャートに落とし込むためには必要である。だからこそ、運転支援システムは定型作業の自動化に威力を発揮する。では、IF-THEN ルールになっておらず、運転員の経験に基づいて行われている作業はどうしたらよいだろうか。ここにデータを活用する余地がある。熟練運転員を関数、それもプラントに関して知りうるすべての情報を適切な操作に変換する関数だと考えれば、データから入力と出力の対応関係を抽出することで、熟練運転員のコピーを作り出すことができる。運転員による手動介入が行われた場合に、その介入操作と生じた結果、そして介入に至った状況を記憶することで、将来、同様の状況が生じたときに、適切な介入操作を提案できる。今までは、人力で構築しなければならなかった運転支援システムが、データから自動的に構築される。

このような運転支援システム自動生成システムが構築できたとして、果たして、事故に繋がりにかからない状態で、データからブラックボックスを経て生成された運転手順にすべてを任せられるだろうか。スマホに広告を表示するのは訳が違う。さらに、未経験の状況にどう対応するかという深刻な問題も残る。ここで、やはり対象プラントについての専門知識が必要になる。物理や化学の法則も含め、これまでに蓄積され、体系化されてきた膨大な知識がある。大量のデータと大量の知識。これらを統合させた先に、未来のプラントオペレーションの姿がある。その統合を実現させる技術を人工知能と呼ぶなら、それもいだろう。問うべきは、人工知能やビッグデー

タをいかに使うかではない。為すべきことは何か。そのために必要な技術は何かである。

実は、プラント運転と同様に、データ解析も暗黙知の塊である。冒頭に紹介した The Economist の記事にも、データサイエンティストを評するのに artist という語が用いられているが、実際、データ解析にはアートな部分が多いと感じる。つまり、どのような解析をするかが解析者の経験や閃きによって決められ、解析結果は誰が解析をしたかに依存する。この属人性はデータ解析の課題の1つである。属人的になる原因として、解析対象の変数やデータを選ぶにも、前処理の方法を選ぶにも、モデル構築の方法を選ぶにも、とにかく試してみないとわからないことが挙げられる。そこで、これまでの経験に基づいて、データ解析手順を標準化している組織もある。その経験が大規模になれば、データ解析、特に時間と労力を要するデータ前処理の自動化も進むだろう。さらに異常検出や仮想計測など目的を限定すれば、データ解析の自動化は比較的容易に実現できるかもしれない。熟練したデータ解析担当者がいなくても、自動データ解析システムが必要な処理を適切に実行してくれるようになれば、生産性は大きく向上するだろう。

人工知能ブームが沸き起こり、人工知能に奪われる職業のリストが注目されたりもするが、かつて無声映画の衰退とともに活動弁士が職を失ったように、これまでも、これからも、技術の発展と共に必要とされる職業は変化する。自分の能力を高めるのは自分である。ただ、データ解析技術が急速に進歩しているものの、しばらくは人手が、それも高度な技能を持つ人材が必要である。このため、データ解析分野での人材育成は喫緊の課題と言えよう。ここで私が思い描いている人材とは、いわゆるデータサイエンティストではない。プログラミングができて、統計学の素養があり、芸術的な解析をこなせることは十分条件ではない。対象プラントについての専門知識を有し、データと知識を統合して、みずから道を切り開いていけるデータエンジニアである。

そのようなデータエンジニアが育つためには、そしてデータ解析技術が組織に根付くためには、まず、小さな課題で成功することが大切である。その対象はスモールデータで構わない。その先に、大きな成功はついてくるだろう。多くのスモールデータエンジニアが誕生することを期待している。

* Exapilot は、横河電機株式会社の登録商標です。