

日本のデータサイエンスの過去・現在・ 未来とOR

—データサイエンスを支える最適化という概念—

上田 雅夫

わが国において、データサイエンティストが不足しているという報告を時折目にするが、不足の原因について、また、今後の方向性について言及している指摘はほとんどない。本稿では、わが国の企業の意思決定に、データ分析が活用されてきたことを、事例をもとに示し、日本における、データサイエンティストが少ない理由を、産業構成が主因であることを指摘した。加えて、データサイエンスを活用するためには、何をすべきかについて検討を行い、その中で、「最適化」という考えが、不可欠であることを指摘した。

キーワード：データサイエンス、データ分析、最適化

1. はじめに

インターネットの普及は、社会・産業に大きな変化をもたらした。特に、ソーシャル・メディア、IoTの普及によりインターネットを経由して、膨大な種類と量のデータが蓄積されるようになった。これまで、ビジネスに活用するデータは、調査や実験を通して「集めて」いたが、現在では、データが自然と「集まる」時代となった[1]。集まるデータは、日々のビジネス活動を通じて蓄積され、自社の状況を正確に反映した記録であり、行った施策に対する結果である。この結果のデータを分析し、伸ばすべき点を伸ばし、改めるべき点を改めることができれば、企業は、厳しい市場環境においても確実に成長することが可能である。

データ分析から得られた結果を活用するためには、日々蓄積される膨大なデータを分析する人材、いわゆる、データサイエンティストと呼ばれる人材が必要である。データサイエンティストを雇用することは、企業の競争力の維持・向上という点において喫緊の課題であり、大学などの教育機関は、企業からこのデータサイエンティストの育成に大きな期待が寄せられている。

2010年のマッキンゼーの調査によると、データサイエンス関連の教育を受けた学生は、アメリカの24,730人に対し、日本は3,400人と約8倍の開きがあった[2]。この差を埋めるべく、わが国では、2017年に滋賀大学に初めてデータサイエンス専門の学部が誕生し、翌年には、横浜市立大学でもデータサイエンス学部が設置

された。また、大学間の連携を通して、データサイエンス教育の向上を目的に、数理・データサイエンス教育拠点コンソーシアムが、2016年に立ち上げられた。データサイエンス教育の支援のために国立情報学研究所では、2015年に設置したデータセット共同利用研究開発センターにおいて、研究用のデータを提供している。さらに、学生だけではなく社会人向けの教育のために、総務省は2015年にオンライン講座「社会人のためのデータサイエンス入門」を立ち上げた。

しかしながら、2020年4月28日付の日本経済新聞では、企業の6割がデータサイエンティストの採用が目標に達していないと回答している。

2. データサイエンスを受け入れる土壌

日頃から意思決定にデータを利用することが組織の文化として根付いているか否かにより、先にあげた施策が実を結ぶまでの時間は異なる。過去にデータを活用した経験がなければ、教育や活用に関するノウハウの蓄積から始める必要がある。一方、過去にデータを活用した経験があれば、それらの遺産を活用できるため、施策が効果をあげるまで、さほど時間が掛からない。

わが国の企業は、長きにわたりビジネスにデータを活用してきた。たとえば、品質管理や製造の部門では、実験計画法に基づいて実験を行い、データを収集し、そのデータを分析することで、品質の改善に努めている。よく引き合いに出される例として、伊奈製陶（現LIXIL）において1953年に実施されたタイルの焼成実験の例がある[3]。タイルを連続的に焼成するにあたり、焼成の不揃いを解消することを目的に実験を行い、データを分析し、連続生産の製造法を確立した。また、

うえだ まさお
横浜市立大学
m_ueda@yokohama-cu.ac.jp

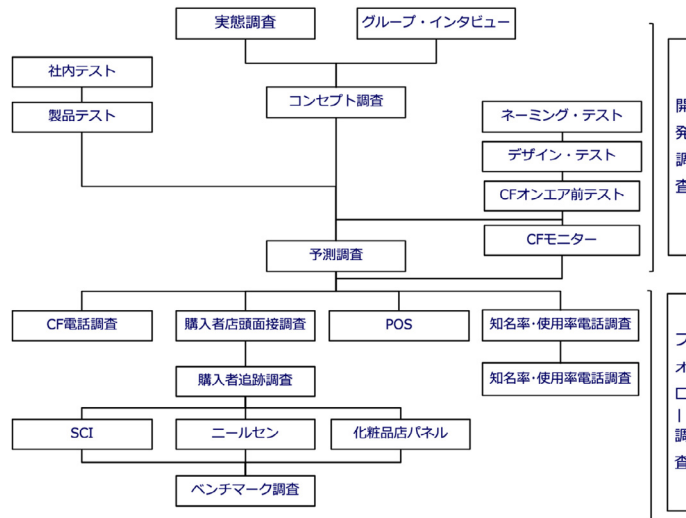


図1 花王のマーケティング・リサーチ体系（文献 [4] より転載）

マーケティング部門では、消費者調査のデータや小売店の POS データを分析し、マーケティングの意思決定にデータ分析の結果を活用してきた。

図1は花王の新商品の開発から上市後までのマーケティング・リサーチ全体の流れである [4]。この図が掲載された書籍は1988年に刊行されており、30年以上前から、商品開発の各ステージにおいて必要なデータが、開発のフローにあわせて活用されていることが理解できる。また、この当時から、花王では、データを分析するための端末を社内に設置しており、必要なときにいつでもデータを分析し、その結果を基に意思決定できる体制が整えられていた。

しかしながら、すべての企業が花王のような完全な活用体系を持っていたとは考えにくい。1982年には、セブン-イレブン・ジャパンにおいて、世界で初めてPOSデータをマーケティングやマーチャンダイジングに活用していた点 [5]、法政大学産業情報センター・小川によるデータ活用についてまとめられた書籍が刊行されている点 [6]などを考慮すると、程度の差はあるにしても、少なくとも1990年代から日本企業はデータを分析し、得られた結果をもとにビジネスに活用していたと考えられる。

加えて、もう一つ特筆すべき点として、データサイエンスという言葉や概念も1990年代にすでに存在したという点である。データサイエンスという言葉が統計学の研究者の間で用いられるようになったのは、1996年に神戸で行われた国際学会からであるという指摘がある [7]。また、データサイエンスという概念も、ほぼ同時期に新村により論文としてまとめられている [8]。そ

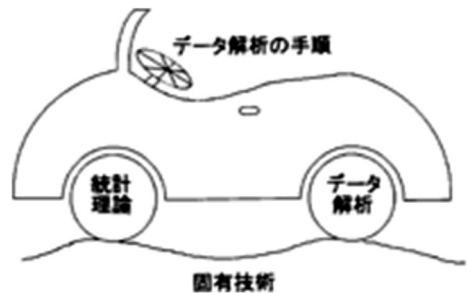


図2 データサイエンスの概念図（文献 [8] より転載）

の論文では、データを分析することを、図2にあるように、自動車の運転にたとえ、前輪に統計理論、後輪にデータ解析のためのソフトウェア、分析者はそれらを駆使して分析を行い、分析の目的に達するというものであると説明している。

図2について、前輪に機械学習を加味し、後輪に、分析用のフリーソフト、分散処理技術、クラウドコンピューターを追加すれば、現在のデータサイエンスを十分に表現することができる。

また、統計理論、データ解析については、大学の各学部や学科（たとえば、経営工学科や商学部のマーケティング・サイエンス、経済学部の計量経済学など）において授業が行われ、そのような知識とスキルを有する人材の育成が行われてきた。

3. 問題の所在

データサイエンスを受け入れる土壌があり、既存の大学でも統計学などの教育を行っていたにもかかわらず、データサイエンティストの需要を満たせない理由

を考えたい。その理由を明らかにすることができれば、現在の施策の成否がある程度判断することができる。

過去にもデータ・マイニングという言葉でデータ分析に注目が集まった時期があったが、そのときと現在では、利用するデータが大きく異なる。データ・マイニングの時代は主に購買履歴データなどの顧客とのトランザクション・データを扱ってきたが、現在は、ソーシャル・メディアの投稿、ウェブのアクセスログなどを分析するようになった。これらの新しいデータはインターネットを經由して蓄積されるため、IT系企業中心に活用されてきた。一方、日本では、インターネットが普及した現在でも、企業といえば、トヨタ自動車、日立製作所といった「モノ」作りの企業が想起されることがほとんどである。GoogleやFacebookが登場した当時、これらの企業と比肩するような複数の巨大なIT企業が、日本に誕生することはなかった。

もし、GoogleのようなIT企業が、Googleが誕生した時期に日本でも、複数社が誕生していれば、その時点からデータサイエンティストのニーズが顕在化し、そのニーズに対応するためにさまざまな施策（現在行われているような施策）が実施されたと考えられる。データサイエンスを支える土壌がすでにあつたことを考えると、その時期に実施された施策の効果により、データを分析するスキル・知識を身に付けた人材が社会に供給され、現在とは異なった状況となつていただろう。しかしながら、日本ではGoogleのような企業が登場せず、従来のキープレイヤーの企業がそのまま活躍したため、従来の仕組みの中でビジネスを継続してきた。その結果、データサイエンティストに対する需要が、インターネット普及初期にアメリカほど大きくなることはなく、データサイエンティストに対するニーズの顕在化が遅れた。

インターネットの普及は、データ環境の変化だけではなく、新しい消費者ニーズや新しいビジネスのコンセプトの出現を促した。「モノ」を所有するよりも、体験する「コト」を重要視する消費者が登場し、市場で生き残るために、このような消費者に対し、モノ作りの企業も新たなビジネスの創出、および、ビジネス環境への適応に迫られている。たとえば、トヨタ自動車では、カーシェアリング事業 [9] や「KINTO」という新会社を設置し、サブスクリプション・サービスを事業化している [10]。これらの新しい事業は、「モノ」の販売ではなく、「コト」の販売である。「コト」の販売で継続的に収益を上げるには、消費者と継続的な関係を構築し、その関係から得られる利用状況のデータを

分析し、提供する価値の向上を図る必要がある。

さらにIndustry 4.0という新しい製造プロセスの概念が提唱され、モノ作り企業全体において、データサイエンスの必要性が生じた。加えて、わが国にも、メルカリ、Gunosy、freeeなどのIT企業が次々に誕生した。これらの企業では、自社が提供する商品・サービスの改善にはデータ分析が欠かせないため、データサイエンティストに対するニーズが高い。

このように、既存の企業と新規の企業、両方の需要が、ほぼ同時期に生じたことでデータサイエンティストの需要が供給を上回り、既存の教育システムでは対応できず、企業側で十分な採用ができないという結果に陥った。このことが、データサイエンティストが不足しているという原因と考えられる。

4. さらなる飛躍を目指して

先に、データサイエンティストが不足している問題は、日本とアメリカにおける主力となる業種の違い、および、IT企業が誕生した時期による差に起因すると説明した。加えて、日本にはデータサイエンスを受け入れる土壌があることを考えると、今、行われているデータサイエンティスト育成のための施策が、将来、成果を生み出すことは間違いない。ただし、データサイエンスが企業の競争力の源泉となる時代になったことを考慮すると、今の施策の効果をさらに向上させることを考えるべきである。

これまでは、データを分析し、分析した結果を用いて意思決定する部署・人は限られていた。これからは、企業のあらゆる活動のデータが収集されるため（集まってくるため）、あらゆる部署においてデータ分析した結果を用いて意思決定するようになる。この変化に対し、これまで一部の人が有していたデータサイエンスに関する知識・スキルを拡大する必要がある。そのためには、Citizen Data Scientistと言われるビジネスでデータを利用する人と先進的な知識・技術を有するデータサイエンティストとの間を橋渡しする人材の育成 [11] や、データを利用する人が分析結果を誤認しないように、データサイエンスに関するリテラシーの向上も行う必要がある。

このような流れの中で、教育機関である大学は、一部の学生に対する専門的な教育を提供するだけでなく、幅広い学生にデータサイエンスに関する基礎的な知識を提供する必要がある。データサイエンスの手法は、さまざまな研究者が精力的に研究を進めた結果、膨大な手法が提案されている。それらの内容をすべて理

解することは専門の研究者ではない限り難しい。そのため、手法の誤用並びに結果を正しく解釈できるような入門的な教材を作成するべきである。

データサイエンスは、データに依存しながら発展してきた側面をもつ。たとえば、自己組織化マップという手法は、変数の数が多い多次元のデータを理解しやすいように縮約すると同時にセグメンテーションを行う目的で開発された。データとは、人や機械のさまざまな行動の結果、蓄積されるものであり、国や地域によってその内容が異なる。

データサイエンスは、社会での活用を目的とする応用研究的な側面を有する。そのため、国内のデータだけではなく広く海外における活用状況を学ぶことで、新しい方法論、課題の解決方法を学ぶことができる。また、企業の海外進出を考えると、企業内のデータサイエンティストも、国内のデータだけではなく、海外のデータを分析することもあるため、大学においては海外の大学と提携し、データの種類や分析事例についての情報を収集し、適宜、授業やシンポジウムなどを通じてフィードバックする必要がある。

上にあげた二つの項目に関する対応を取る際、学部単体で教育を行うのではなく、全学的な横断組織が主体となり、教育を行うことが望ましい。もし、データサイエンス関連の底上げをするための教育を、各学部が主体となり教育を行うと、その学部で主に使う手法に偏りがちになる可能性がある。また、海外の大学との連携のような場合、窓口を一つにして対応した方が連携を円滑に行える。データサイエンスのための学部横断的な組織を、すでに設置している大学は複数あり、たとえば、早稲田大学では、「データ科学センター」という全学横断的な組織を作り、データサイエンス・統計学の学習コンテンツを作成し、大学全体に展開している。今後はこのような横断的な組織と各学部との連携によりお互いの長所を生かした教育が行われることが期待される。

5. データサイエンティストが活躍するには

データサイエンティストを育成するために、専門の学部の設置、大学間の連携さらに学内横断的な組織が設置されるなど、さまざまな施策が実施されている。これらの施策により、データサイエンティストを確保できないという問題は早晩に解決されると考えられる。ただし、企業にとっては、データサイエンティストを雇用することが目的ではなく、データを分析した結果を用いて合理的な意思決定を行うことが目的である。

したがって、データサイエンティストを雇用しただけでは不十分であり、データサイエンティストが活躍できる環境を整備する必要がある。そのためには、どのようにデータを活用することが企業活動全体にとって望ましいか検討し、必要に応じて、業務のフローを修正する必要がある。業務のフローを考え直す際は、自社の業務の全体最適を目的とし、再設計する必要がある。同時に、データを利用した意思決定の意義を社内に浸透させることも重要である。

企業活動とは、顧客へ提供する価値を最終的な目的とし、それに付随する制約条件を考え、その条件下において提供する価値の最大化を図ることである。その際は、価値の提供フロー全体を最適化する必要がある。生産工程を最適化し、品質の高い商品を作ったとしても、物流に問題があり、消費者が必要としているときに提供できなければビジネスとしての意味はない。データが手許にあったとしても、全体最適の概念・考え方が企業全体に十分に浸透していないと、企業全体の活動を最適化することは容易ではない。企業の経営が複雑になった現在では、そのリスクを低下させるためにデータ分析した結果を用いて意思決定を行う。データは意思決定を行ううえで不可欠な材料である。ただし、「集まる」データには、「集める」データとは異なり、真に必要なデータが手許にない場合がある。その際は、コストをかけてデータを収集するのか、必要なデータと相関の高い「集まる」データで代替するのかを考える必要がある。データがないからデータを用いずに実行するのではなく、データを参照して実行するにはどのような手段を採用すべきか考えることが重要である。

先にあげた考えは、オペレーションズ・リサーチ (Operations Research; OR) という学問の考えと一致する。後者の考えについても、後藤により代替となるデータの重要性はすでに指摘されている [12]。経営工学科出身の人々はカリキュラムに OR が含まれており、全体を最適化することの重要性は理解しているが、他学部、他学科では、OR が教科として含まれていることが少ないため、企業人のすべての人が OR の知識や考え方を理解しているとは考え難い。現在はデータ環境の変化に合わせた業務の進め方、社内システムの再考する時期である。OR 的な考えを基に、データを用いた意思決定の意義を理解し、現在の業務の進め方を見直すことができ初めて、データサイエンティストを雇用した目的が達せられる。

6. まとめ

2022年のデータのトラフィック量が2017年の3倍になるという予測がある [13]。今以上にデータのトラフィック量が増加することは、データサイエンティストが活躍する機会が今以上に増えることである。その機会に対応するためにも人材の育成は重大な課題である。

日本において、データサイエンティストが少ない理由は、産業構成が主因であり、データサイエンスの歴史、データサイエンスを受け入れ発展させる土壌がないというわけではない。データサイエンティストの需要を満たせていないという現状の課題はあるが、その課題を克服するためのさまざまな施策が行われており、データサイエンスに関するこれまで蓄積された資産を用いれば、今後については悲観することはない。むしろどのように活用すれば、結果が出せるのかという点を考えるべきである。

企業では将来を見据え、データサイエンティストの採用を進めようとしているが、データサイエンティストを雇用するだけでは、業務においてデータサイエンスを活用することはできない。データ分析から得られた何らかの知見を活用するには、自社のビジネスのどこに組み込むのか、また、データを活用することで、ビジネス全体にどのような効果があるのかといった、全社の視点で見る必要がある。

先にデータサイエンス的な取り組みは、大半の日本企業は行ってきたと述べたが、それは、品質管理部やマーケティング部などの部署単位の活動が中心であった。これまでは、データは「集める」モノであったためにコストが掛かり、必要な部署で必要なデータを収集して活用していた。このことは、データを活用する業務、人材が一部の部署に偏ることを意味する。しかし、現在は企業の活動全体にわたって、データが「集まる」時代となった。営業部はPOSデータ、取引明細のデータ、宣伝部はソーシャル・メディアの投稿データ、広報部は自社の広報サイトのアクセスログ、製造部は工場の機械のセンサーのデータといった具合に、データは企業の活動とともに集まってくる。この集まるデータをもとに、どのように業務を再構築するか検討する時期である。言い換えると、全社の活動に関して、データを用いて円滑な意思決定ができるようになることである。意思決定を行ううえで、最善のデータがなければ、次善の策を考えるべきである。企業の活動にはリスクが伴う。このリスクをコントロールするためにもデータを用いた意思決定は不可欠である。

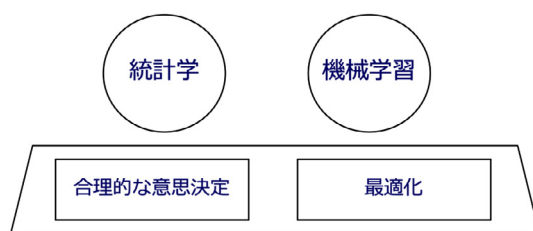


図3 データサイエンス

データサイエンスの時代には、「合理的な意思決定」並びに「最適化」という言葉が、企業が競争を優位に進めるためのキーワードである。企業活動全体でデータを活用できる現在では、企業活動全体を通しての最適化が求められる。どこかの部署という部分最適ではなく、顧客に届ける価値の創造から、その価値を伝達するところまで、全体を通しての最適化である。そのため、最適化という概念の普及が重要である。データサイエンスでは、解決できなかった問題を解決する手法の開発も重要であるが、それと同時に、データを活用するうえで基本となる考えの普及も重要である。合理的な意思決定並びに最適化という考えを普及させるためにも、ORもデータサイエンス教育に加えるべきであろう。図3にあるように、統計学や機械学習を支える根本的な考えが合理性と最適化である。

同時に、そのような考えを広く知らしめるような活動が重要になる。その活動については、さまざまな組織が担うことができるが、社会にORの考え方を浸透させることができる組織として、日本オペレーションズ・リサーチ学会に対する期待は小さなものではなく、データサイエンスの時代に今以上に存在感を発揮することが期待される。

参考文献

- [1] 星野崇宏, 上田雅夫, 『マーケティング・リサーチ入門』, 有斐閣, 2018.
- [2] Mckinsey & Company, https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Big%20data%20The%20next%20frontier%20for%20innovation/MGI_big_data_full_report.pdf (2020年6月30日閲覧).
- [3] 田口玄一, 『品質工学の数理』, 日本規格協会, 1999.
- [4] 陸正, 『マーケティング情報システム』, 成文堂新光社, 1988.
- [5] セブン-イレブン・ジャパン, https://www.sej.co.jp/company/aboutsej/info_03.html (2020年6月30日閲覧).
- [6] 法政大学産業情報センター, 小川孔輔, 『POSとマーケティング戦略』, 有斐閣, 1993.
- [7] 椿広計, “システム科学とデータ科学,” 横幹, 14, pp. 64–69, 2020.

- [8] 新村秀一, “データ解析からデータ・サイエンスへ—情報技術(統計ソフト・WWW・AI)との共生により統計知識を国民の知的共有財に一,” 統計数理, **45**, pp. 23–40, 1997.
- [9] トヨタ自動車, プレスリリース資料より, <https://global.toyota.jp/newsroom/corporate/30300614.html> (2020年6月30日閲覧)
- [10] トヨタ自動車, プレスリリース資料より, https://global.toyota.jp/newsroom/corporate/26466090.html?_ga=2.82175190.1780975151.1587696577-506449421.1562820609 (2020年6月30日閲覧)
- [11] J. Tapadinhas and C. Idoine, Pursue Citizen Data Science to Expand Analytics Use Cases, <https://www.gartner.com/en/documents/3539717/pursue-citizen-data-science-to-expand-analytics-use-case> (2020年6月30日閲覧)
- [12] 後藤正夫, “OR 昔ばなし,” オペレーションズ・リサーチ: 経営の科学, **42**, pp. 435–439, 1997.
- [13] CISCO, Cisco Visual Networking Index (VNI): 予測とトレンド, 2017–2022年ホワイトペーパー, https://www.cisco.com/c/ja_jp/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html (2020年6月30日閲覧)