

巻頭

データサイエンス教育の重要性 ～データとICTを中心とした問題解決力の育成～

慶應義塾大学大学院 健康マネジメント研究科教授 渡辺 美智子

1. データ中心による研究・産業基盤技術の変化

Googleが開発したコンピュータソフトのアルファ碁が囲碁の世界では最強とも言われる名人イ・セドル九段に勝利したことから、人工知能への関心が広まっている。アルファ碁以外にも、会話するロボット、自動で運転する自動車、モノとモノとをインターネットで繋ぐIoT (Internet of Things) など、これまでSFの世界とされていたことが、急速に現実化してきている。この背景には、機械学習 (マシーンラーニング) という、状況を示す膨大なデータから次々とルールを認識し、最適な予測や判断を行う高度な統計数理モデリング技法の進歩とそれを演算可能にしているコ

ンピュータの能力拡大がある。

この技術は、すでに、スパム検知、クレジットカードの不正検知、数字や顔画像認識、商品レコメンデーション、医療診断、信用リスク予測、自然言語処理など、ビジネス用途で応用され、その成功を要因に爆発的に普及が進み、社会実装が本格化してきている。このことで、現在は、18世紀後半の工業化の黎明期を語る第1次産業革命 (蒸気機関による自動化)、19世紀後半の大量生産と文明化を語る第2次産業革命 (電気による自動化)、20世紀後半の電子化による製品・生産設備システムの進化を語る第3次産業革命 (コンピュータによる自動化) に続く、第4次産業革命 (データ駆動型サービスによる自動化) の時代に突入

CONTENTS

巻頭

データサイエンス教育の重要性 …………… 1
～データとICTを中心とした問題解決力の育成～

解説

情報教育における「情報の科学的な理解」を深める教材… 6

解説

情報科を担当する教員の採用・配置・研修について… 10

授業実践

情報オリンピック問題を活用した
プログラミング学習教材の開発とその実践…………… 14

報告

日本情報科教育学会 第9回全国大会…………… 18

紹介

1万人の子供たちがプログラミングを体験…………… 21

CPSによるデータ駆動型社会の実現のための分野別取組



図1 データ駆動型社会実現に向けた分野別取組 (経済産業省HPより)

したとまで言われている。図1は、平成27年5月に、経済産業省で取りまとめられたデータ駆動型社会実現に向けた分野別取組図である (経済産業省, 2015: Cyber Physical System (CPS))。

筆者においても、独立行政法人情報通信研究機構が助成する「ソーシャル・ビッグデータ利活用アプリケーションの研究開発」で、慶應義塾大学と株式会社タニタヘルスリンクの共同研究として採択された研究課題「ヘルスリテラシー向上のための生体ログデータ分析に基づく健康情報フィードバック」プロジェクトの中で、生体ログデータ分析班の責任者として、近年、個人で身近に使用される歩数計や活動量計から得られる大量で長期間に渡る記録データ (健康ビッグデータ) を対象に、日内×週の人間の活動パターンの類型化と体組成情報との連関を統計分析によって明らかにし、健康を工学的に作り上げていく上で必要な科学的エビデンスの生成に従事している。このようなデータは、いままで計測されてこなかった全く新しい研究素材で、仮説に沿って適切に統計分析されるだけで、漠と感じていた人間行動が多様な観点から科学的に解明され、健康社会構築の大きな一助となる材料になっている。

ビッグデータを基盤とする変革は産業界だけで

はなく、科学研究の方法論に関しても、知識・経験の蓄積を背景とした古くからの「理論科学」と「実験科学」、20世紀半ば頃からのコンピュータの発達に伴う「計算科学」に加えて、膨大なデータから直接、社会・自然・経済・人間行動等の法則を確立する「データ中心科学」が、第4の科学 (The Forth Paradigm) として台頭し、医学、健康科学、生物学、物理学、地学、経営学、経済学、社会学、教育学、スポーツ科学等の多くの領域で、領域固有のビッグデータを活用した創造的研究成果が新しく生まれている (Tony Hey, 2009)。

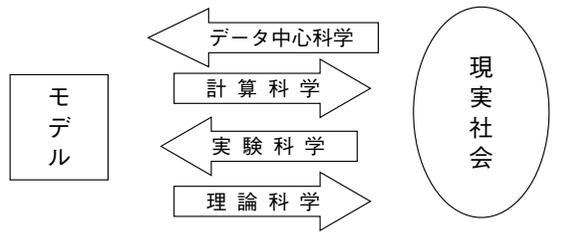


図2 ビッグデータを活用した「データ中心科学」

2. 社会で必要とされるデータ分析能力

データ中心による産業および研究における基盤技術の変革が進む中で、その変革を担う人材不足は深刻であり、そのための人材育成や教育システ

ムの構築は、各国政府の喫緊の課題となっている。すでに1990年代より日本以外の国際社会は、身近な具体的課題に対する問題解決をデータに基づいて主体的に、かつプロジェクトベースラーニングによって経験する方式に、統計教育の改革を初等中等教育から大学初年次段階まで体系的に進めており、ICTを活用した協働的・統計的問題解決力の育成に関しては、カリキュラムや教材、教授法や評価の枠組みなどの標準化が行われ、今日のデータ分析系人材育成の礎を築いている。

その上でさらに、ビッグデータ・オープンデータ時代が本格化した2010年以降、コンピュータサイエンスと応用統計科学を融合したデータサイエンス学部（学科）や研究科、学位の創設、初等中等教育における共通コアカリキュラムでの統計内容の強化、高校におけるデータサイエンスコースの新設など、データ分析系人材の質と量の双方の拡大に向けた取組みを急ピッチで進めている。

一方で、従来、統計を専門とする学部や研究科を大学に有しておらず、また、統計の授業といえは数理統計的な理論が主であった日本においては、データ分析系人材の欠如は欧米・中国等の海外諸国とは比べものにならないほど大きく、産業界からも指摘されているところである。

この状況を受け、科学技術イノベーション戦略（平成27年6月閣議決定）において、「我が国は欧米等と比較し、データ分析のスキルを有する人材や統計科学を専攻する人材が極めて少なく、我が国の多くの民間企業が情報通信分野の人材不足を感じており、危機的な状況にある」が示され、続く第5期科学技術基本計画（平成28年1月閣議決定）において、『データ解析やプログラミング等の基本的知識を持ちつつビッグデータやAI等の基盤技術を新しい課題の発見・解決に活用できる人材などの強化を図る。』方針が明記されている。つまり、各府省は、今後この方針の具体化に向けた政策を進めていくことになる。

文部科学省でも、『第4次産業革命に向けた人材育成総合イニシアティブ～未来社会を創造するAI/IoT/ビッグデータ等を牽引する人材育成総合



図3 「第4次産業革命に向けた人材育成総合イニシアティブ」（文部科学省HPより）

プログラム～』（平成28年4月）を公表し、数理・情報が第4次産業革命の鍵として（図3）、高等教育だけではなく初等中等教育も含めて、「次代を拓くために必要な情報を活用して新たな価値を創造していくために必要な力や課題の発見・解決にICTを活用できる力を発達の段階に応じて育成」することを示した。ここで言う「数理統計」は、統計領域の中の狭義の数理統計ではなく、現実の社会課題解決に向け、コンピュータサイエンスと融合した、数理科学における広義の統計領域（データサイエンス）を意味している。

ただし、データサイエンス教育自身は、必ずしも、先述のような大規模な社会問題の解決に向けた高度なデータアナリティクス系人材の育成だけを目指すためのものではなく、基本的な日常の意思決定や判断において、データを客観的なエビデンスとして科学的な思考および統計的な思考で対峙する能力自身を指し、国民全体に涵養すべきキーコンピテンシーとしても位置づけられている。

とくに、データに基づく科学的意思決定力の強化は、21世紀型スキル教育やグローバルリーダーシップ教育の中でも最重要視されている内容である。

日常の現象を複数の具体的な事象の関わり合いとして捉え、各事象にデータを対応させ、事象間の関連性の法則を数学モデルで評価するというデータ思考力 (Think with Data) を十分に鍛えておかないと、今後、私たちの身の回りで実現され、提供されてくる様々なデータ駆動・人工知能型のサービスの基本技術が、国民にとって全くのブラックボックスとなる危険性がある。データサイエンス教育は、国民の基本リテラシーであり、その上で、将来の専門職能の育成も見据えた教育、この双方の視点で、早急に初等中等教育から高等教育・職業教育を繋ぎ体系的に実現されるべき内容となっている。

3. データサイエンスのフレーム

データサイエンスは、

- ① 統計・数理学
- ② コンピュータサイエンス
- ③ 背景となる固有領域のサイエンス

の融合領域とされる。①と③の融合では、実証分析を伴った計量経済学、計量生物学、計量心理学、計量ファイナンス、計量文献学、計量政治学、環境計量学、宇宙計量学など、今日の“計量”を冠する多くの研究領域がある。そこでは、従来型の実験や調査の設計とデータ分析が必須となる。②と③の融合で、固有領域の工学的アプローチとそのソフトウェアが生成される。①と②の融合で、データマネジメント (データ加工・データクリーニング) 処理も含めた統計分析ソフトウェアが誕生し、多くの実証分析を支えてきた。さらに、近年の機械学習も含めた予測・推論アルゴリズムの開発と実装化が進歩し、現実社会の適用範囲を飛躍的に拡大させるという意味で③が融合し、今日のデータサイエンス領域が生み出されている。

データサイエンススキルとは、領域固有の現象の枠組みの中で、以下を実践する総合的なスキルである。

- ① 解くべき課題をつかむ
- ② 関連する事象を洗い出し構造化する
- ③ データで解ける問題に落とし込む

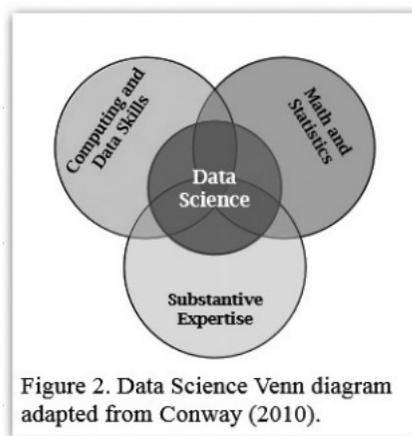


図4 データサイエンス・ベン図
(<http://www.dataists.com/>より)

- ④ 具体的な探究仮説を明確にする
- ⑤ 統計数理モデルに基づきICTを活用し、データ分析による探究を行う
- ⑥ 分析結果の数理解釈を行う
- ⑦ もとの領域の文脈で考察し、問題解決に繋がる提案をする

問題解決の成否は、データ分析技術が高度であればよいということではなく、分析の前段階である、現象への理解と仮説の形成および分析の後段階である、現象の文脈に沿った結果の解釈とそのエビデンスに基づく現象への意思決定の適切性で決まる。そのため、個人の専門知識や経験価値での対処より、情報を他者と共有しながら議論することを通じて、多様な考えを統合し協働することで、効果的に達成される。したがって、データサイエンススキルの育成に関しては、ICTのスキルや数理的・統計的分析手法などの個別の方法論の習得と並行して、科学的 (数理的・統計的) および協働的作業を含めた問題解決の一連のプロセスへの理解を経験的に積上げることが重要であり、生徒自らが主体的に取組むグループでの探究活動を行うアクティブ・ラーニングの機会を発達段階に応じて繰り返し設ける必要がある。

現在、OECDが21世紀型スキルとして、ICTを活用した協働的・科学的問題解決力の育成を重視しているが、データサイエンススキルは、そのた

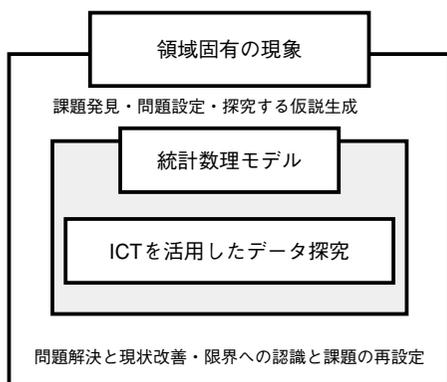


図5 データサイエンススキルの育成

めの主要なキーコンピテンシーである。

4. 米高等学校でのデータサイエンス入門コース

米国や中国では、学部や研究科の新設によるデータサイエンス教育を急拡大させているが、これは高等教育に限ったことではなく、高等学校でもデータサイエンスコースの開設が進んでいる。

米国ロサンゼルス市統合教育区教育委員会(LAUSD)は、全米科学財団(NSF)の助成も併せて、12.3百万ドルで、カリフォルニア大学ロサンゼルス校(UCLA)と共同で、データサイエンス入門コース(IDS)を開発し、各高校での授業に供している。このコースでは、2011年、初等中等教育において統計内容が大幅に強化された数学全米共通コアカリキュラム(CCSSM)および全米学術研究協議会(NRC)による次世代科学教育スタンダード(NGSS)に沿い、実社会の現象理解を目的に、実際のデータ取得、統計分析を行う計算コード(ソフト)、データ可視化ツールも含めて提供し、よりダイナミックな計算機ベースの統計と確率を扱うことで、ビッグデータ社会の高度化された情報と科学技術を理解し、批判的な数理思考ができる生徒の能力の育成を図っている。

具体的には、

- データの構造とタイプ、表現と共有手段
- Computational thinking (the problem-solving, logical thinking at the heart of CS)

○データ探究サイクル・分析アルゴリズム (the rules governing data collection and strategies for analysis)

○Problem-solving skills, Innovation, inventiveness, and interdisciplinary collaboration

等の実践力を①データの解釈；グラフによる要約、シミュレーションと推測（基本グラフ、並列箱ひげ図、散布図）、統計モデル（線形、k-means、平滑化、学習・決定木）、②統計的推測（推定と検定）と判断の妥当性；データソース、ランダムサンプリング、ランダム割り付け、A/Bテスト、シミュレーションによる推測、データ取得の実践、③確率；computers and randomness (Web services, 乱数)、統計的確率、確率計算、④R-Studioによる演算；ベクトル、アルゴリズム、関数、モデルとデータの適合性評価、高次元データの可視化等の内容を踏まえて、具体的なプロジェクトベースで身につける構成になっている。詳細はMobilize (2016) 参照。

今後も実世界のデジタル化とデータ流通は急速に進んでいく。その中で、データの処理と解析結果を実世界へフィードバックする能力育成をいかに図るのか、教育内容の充実が喫緊の課題である。Glassdoor (2016) が発表した米国のthe 25 Best Jobs in America for 2016の第1位は、「データサイエンティスト」となった。日本でもデータサイエンス学部や学科、副専攻やコースの設置などが急速に大学で進められている。高大接続の観点からも、データサイエンスの内容が高校で充実されることが望まれる。

参考文献

Conway, D. (2010). The data science venn diagram. Dataists [Web page]. Retrieved February 9, 2012, from the <http://www.dataists.com/2010/09/the-data-science-venn-diagram/> database.

Mobilize (2016) : <http://www.mobilizingcs.org/curriculum>
産業構造審議会 (2016) : CPSによるデータ駆動型社会の到来を見据えた変革