

[AI 人材教育]

2 滋賀大学における AI 人材教育



竹村彰通 | 滋賀大学データサイエンス学部

データの時代の到来

ビッグデータの時代の到来とともに、世界的に GAF A に代表されるビッグデータを収集・活用する企業が大きく成長し、プラットフォームとして支配力を増している。日本ではデータを分析するデータサイエンティストの育成が進まず、多くの企業でデータは蓄積されているもののそれを活用できていない状況が続いており、日本の競争力の低下につながっている。ここではこのような状況の背景や滋賀大学データサイエンス学部におけるデータサイエンティストの育成方針について論じる。なお本稿のタイトルは AI 人材としているが、データサイエンスと AI は重なる部分が多いため、ここでは典型的な AI 人材としてのデータサイエンティストの育成を扱う。ただし、教育内容の紹介において、最近の AI 手法の教育の例を示す。

データサイエンスとは

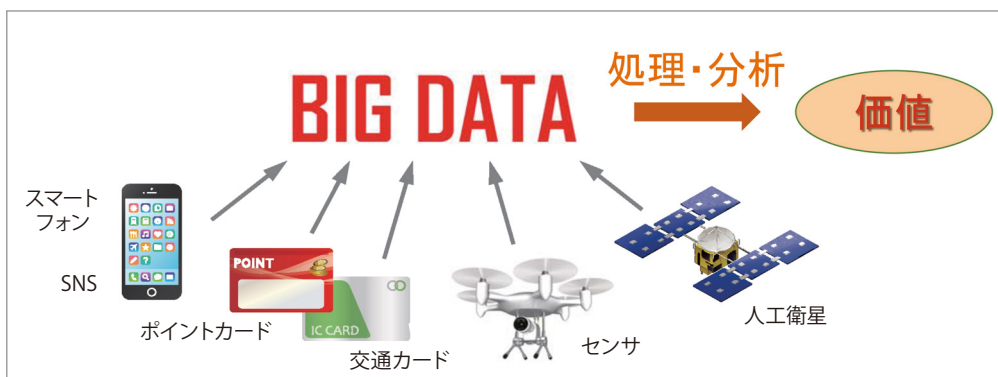
データサイエンスはビッグデータを扱い、そこから価値を生み出すための方法論である。ビッグデータは経済的な資源と考えられており、最近では「21 世紀の原油」

と呼ばれることが多い。

ビッグデータが集まるようになったのはスマートフォンなどのデバイスの発達によるところが大きい。スマートフォンのほかにもコンビニで用いるポイントカードや電車に乗るときの交通カードからも大量のデータが得られる。人工衛星などにより環境に関するビッグデータも蓄積されるようになってきている (図-1)。

特に 10 数年前にはじまったスマートフォンの普及は象徴的であり、今では若い人々の間のスマートフォンの普及率は 100% に近い。地下鉄の中などを含め、我々の通常の生活範囲にはすべて携帯の電波が届くようになり、人々は通勤途中でも電子メールの読み書きや情報検索をしている。スマートフォンの普及以前は、いわゆるガラケーの時代でも画面のサイズや通信容量の制約から、人々は電車の中で新聞や文庫本を読んでいたが、最近ほとんど見ない。

新聞や書籍とスマートフォンの違いは、スマートフォンでは人々の検索等の履歴がネット上に直接蓄積されるという点である。行動履歴を直接取得可能となったことが大きなきっかけとなり、個人を対象とするさまざまなインターネットサービスが提供されるようになった。つまりビッグデータのうちビジネスに直結するデータは



■ 図-1 ビッグデータの取得

人々の行動に関するデータであり、この意味でデータサイエンスの応用分野は文系の分野であることが多い。

スマートフォンなどのデバイスが普及する前は、人々の行動に関するデータを得るには、アンケート調査が主要な手法であった。ただしアンケート調査が不要になったわけではない。人々が何を考えているか、どのような意見を持っているかは、デバイスによって直接観測できるわけではないため、アンケート調査は依然として必要である。実際インターネットによるアンケート調査はしばしば行われている。最近も厚生労働省によるLINEを用いた新型コロナウイルスに関する大規模なアンケート調査が行われ、送信数は8,000万以上、回答数は2,000万以上となり、SNS経由のアンケート調査の有効性が示された。

以上のように、人々の行動に関するデータが得られるようになったことが大きな変化であるが、人々の行動のみならず自然やモノからのデータも蓄積されている。キーワードはIoT (Internet of Things, モノのインターネット) である。通信機能を持ったさまざまなセンサの低価格化に伴い、製造工程の機械や環境をモニタする機器にもセンサをとりつけてデータを吸い上げている。典型的な例は「コネクテッドカー」であり、自動車にさまざまなセンサと通信機器を搭載し、自車や周囲の状況に関する大量の情報をリアルタイムにやりとりする。安全な自動運転を実現するにはこのような通信機能は不可欠である。このように今後はモノからも常時大量にデータを取得可能になり、その利用が進んでいく。ビッグデータは一時的な現象ではなく、社会の大きな変化を表す言葉である。人々の行動に関するデータの利用は文系的であるが、モノからのデータの利用は理系的であることが多い。しかしながら、たとえばスマートウォッチから得られる人々の健康状態に関するデータは、文系とも理系とも分類しにくい。データ自体に文系と理系の区別があるわけでもない。このようにデータサイエンスは本質的に文理融合的な分野である。

この意味で、データサイエンティストを含むAI人材の育成には、従前の文系理系の区別にとらわれない教

育が必要である。

データサイエンスの要素技術と価値創造

データサイエンスはビッグデータから価値を引き出すための方法論であるが、その要素技術はデータを処理・加工するための情報学とデータを分析するための統計学である。さらに、それらの技術を用いて、データから価値を引き出す価値創造もデータサイエンスの要素である。情報学、統計学、および価値創造は、しばしばデータサイエンスの3要素と呼ばれる。

このうち情報学については、ビッグデータを手計算で扱うことは不可能であり、ビッグデータの処理・加工においてコンピュータを駆使することは前提であるから、情報学は必須である。基礎的なプログラミングの知識に加えて、データベース、機械学習や深層学習などのライブラリを活用する必要がある。さらに実際のデータ処理作業の場面では、データの形式を揃えたり欠測値を処理したりなど、「前処理」のノウハウが意外に重要である。

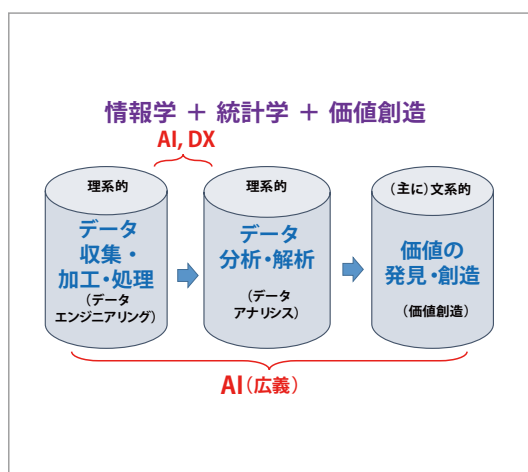
データの前処理が終わると、データ分析のステージとなる。このための方法論が統計学である。統計学は伝統的にはサイズの小さなデータから有用な情報を得るための方法論として発展してきたが、統計学の中で生み出されてきたさまざまな手法や概念は、ビッグデータを見る際にも基本的である。手法としては、基本的な記述統計・推測統計の手法に加えて、回帰分析、多変量解析、時系列解析、シミュレーションなどを学ぶことが重要である。データの種類については、最近ではテキストデータや画像データがビッグデータの大きな部分を占めるようになってきたため、これらのデータの扱いも学ぶ必要がある。

以上のように、情報学と統計学がデータサイエンスの基礎的スキルであるが、それらだけでは「価値創造」、すなわちデータから有用な価値を引き出すことに繋がるとは限らない。情報学と統計学のスキルを持った上で、企業や自治体の実際のデータを分析し、さらに分析結

果を実装に結びつけることが重要であり、これがデータサイエンスの出口となる。価値創造の能力を育てるには、講義では不十分であり、演習の形で実際のデータを分析し、それに基づいて具体的な提案を考えるような経験が不可欠となる。

以上の3要素の関係を図-2のようにまとめることができる。なお図-2ではデータサイエンスとAIの関係も示している。手法としてのAIは、深層学習などのデータ処理から分析の部分の自動化や高度化を表す。しかしより広い意味ではAI手法の結果の解釈における倫理的・社会的な側面等も考慮する必要がある。政府のAI戦略2019におけるAIの扱いも、広義のAIである。

以上では価値創造の部分を含めたデータサイエンス教育の考え方を強調したが、実はアメリカにおいては、コンピュータと統計に強ければそれだけでデータサイエンティストとしての専門性が評価され、就職が有利になる状況がある。これはキャリアとしてのデータサイエンティストや統計家の役割が確立しているからである。これに対して日本では、データサイエンティストは理系のスキルを有するだけでは十分に機能できず、文理融合的な能力が求められる。理系のスキルだけだと日本では「数字だけ分かっている現場では役に立たない」といような反応を受けることが多い。この意味で日本におけるデータサイエンス教育には文理融合的な視点が非常に重要である。



■ 図-2 データサイエンスの3要素

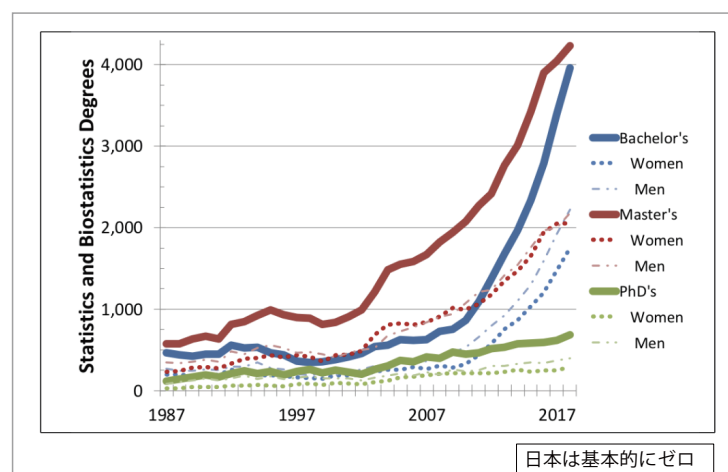
データサイエンティスト育成の海外動向と日本の対応

データサイエンティストの育成において我が国は大きく遅れている。私の所属する滋賀大学データサイエンス学部は、この遅れを取り戻すべく2017年4月に開設された日本初のデータサイエンス学部である。滋賀大学データサイエンス学部では、データサイエンスの要素技術として統計学を重視しており、統計学に関する専門学部としても日本初である。以下ではまず私の専門である統計学の状況について述べる。

日本の大学には統計学を専攻する学部や学科はこれまで存在していなかった。これに対して、アメリカには統計学部・学科は150近くの大学にあり、韓国には50程度、中国には300以上の大学にある。その数も増加傾向である。統計学に関しては、日本はアメリカと中国・韓国の間で挟まれてこれまで統計学部・学科が存在しないというきわめて特異的な状況にあった。

次図はアメリカ統計学会のニューズレター(2019年12月号)に掲載された統計学あるいは生物統計学を学位名に含む学位数(学士、修士、博士)の推移を表したグラフである。

図-3より、2018年には修士号が4,000名以上、学士号も4,000名程度、博士が700名程度授与されていることが分かる。



■ 図-3 米国の統計学学位授与数

それよりも顕著なのは、2008年あたりからの学位授与数の増加のスピードである。当時カリフォルニア大学バークレー校の著名な経済学者で、その後グーグル社のチーフ・エコノミストとなった Hal Varian が「私がいつも言っているのは、今後10年間で魅力的 (sexy) な仕事は統計家 (statistician) だということだ」と発言し話題になったが、実際にそのことが学位授与数に表れている。このように統計学の学位授与数が増えているのは、給与面を含め職業としての魅力が大きいためである。中国でも状況は似たようなものであり、BAT^{☆1}とよばれる巨大なインターネットサービス企業等が統計学部や統計学科の卒業生を採用している。

これに対して、日本ではこれまで統計学部や学科が存在しなかったため、データサイエンティストの組織的育成ができていない。数年前の資料となるが、このような状況についての詳しいデータは、日本学術会議情報学委員会 E-サイエンス・データ中心科学分科会提言『ビッグデータ時代に対応する人材の育成』(2014年9月)や同数理科学委員会数理統計学分科会提言『ビッグデータ時代における統計科学教育・研究の推進につ

☆1 「Baidu (百度, バイドゥ), Alibaba (阿里巴巴集団, アリババ), Tencent (騰訊, テンセント)」である。

いて』を参照されたい。

日本でも、最近になって多くの企業がデータサイエンティストの中途採用を進めており、人材の取り合いの状況が生じている。このようなデータサイエンティストの不足については、政府の文書でも2015年頃より頻繁に取り上げられるようになった。これはGAFAの躍進により、日本がこの分野で大きく遅れていることが政府においても明確に認識されるようになったためだと考えられる。2015年6月閣議決定の『科学技術イノベーション総合戦略2015』では「我が国では欧米等と比較し、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きわめて少なく、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」としている。また2017年6月閣議決定の『未来投資戦略2017—Society 5.0の実現に向けた改革』では「数理・データサイエンス教育の重要性・必要性は分野を超えて高まっているが、理系の一部の学生しか学んでおらず、文系理系を問わず、学ぶ機会が乏しい」としている。

さらに2019年7月に政府が発表した「AI戦略2019～人・産業・地域・政府すべてにAI～」では、教育改革の主な取り組みとして、デジタル社会の「読み・書き・そろばん」である「数理・データサイエンス・AI」の基礎などの必要な力をすべての国民が育むとしており、特に大学・高専卒業生全員50万人/年に数理・データサイエンス・AIのリテラシー教育を行うとしている。図-4は2019年7月9日の内閣府政策統括官(科学技術・イノベーション担当)によるAI戦略2019概要に示された図である。

AI戦略2019で示された方針はその後の文部科学省の政

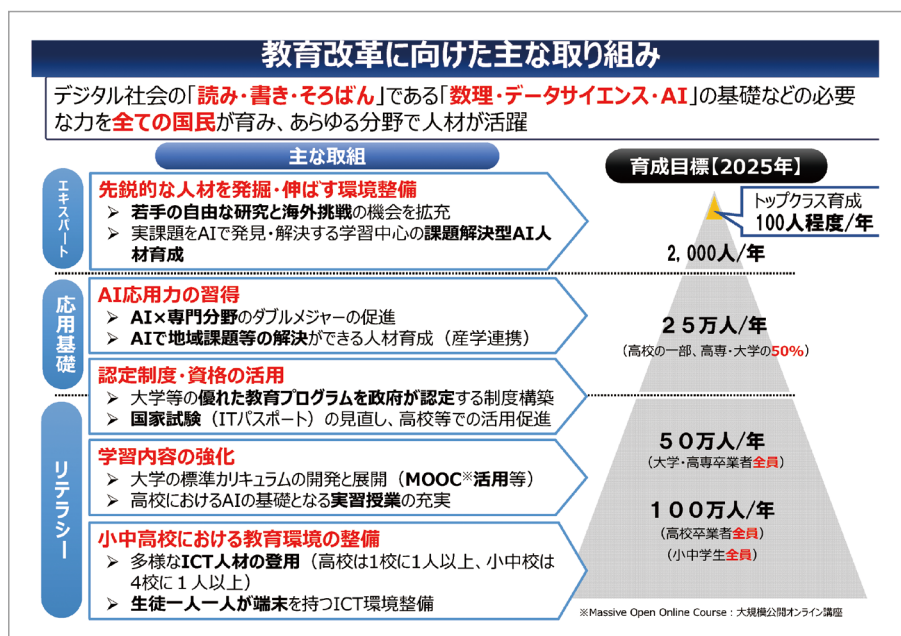


図-4 AI戦略2019の教育改革方針

策にも大きく反映されており、この方針に基づいた「数理・データサイエンス・AI教育プログラム認定制度」が整備されている。

滋賀大学データサイエンス学部の開設とそれに続く動き

以上のような状況の中で、滋賀大学データサイエンス学部は、日本初のデータサイエンス学部として2017年4月に開設された。定員は1学年100名である。滋賀大学データサイエンス学部は、文理融合教育を標榜しているが、入学生の4割程度が文系の学生であり、また女性の比率も2割程度となっている。

続いて2019年4月には大学院データサイエンス研究科修士課程を前倒しで開設した。定員は20名である。学部開設後2年での修士課程設置であり、まだ学部の卒業生がいないときであったため院生の大部分は企業からの派遣社会人である。実際1期生は23名の入学者のうち19名が企業派遣であった。このように前倒し設置に踏み切ったのは、さまざまな企業から企業内のデータサイエンティスト育成の要望が寄せられたためである。19名の企業派遣の院生の派遣元の業種は、金融系5名、製造系4名、調査系4名、IT系3名、政府系3名となっており、広い業種に渡っている。これはデータサイエンティストがさまざまな業種で必要とされていることを示している。

さらに2020年4月には博士課程(定員3名)を開設し、滋賀大学としては学部から博士課程まで一貫したデータサイエンス教育の体制を確立した。

滋賀大学に続いて、2018年には横浜市立大学データサイエンス学部、2019年には武蔵野大学データサイエンス学部、兵庫県立大学社会情報科学部、2020年には長崎大学情報データ科学部が開設された。その後も、立正大学データサイエンス学部、一橋大学ソーシャル・データサイエンス学部の新設が予定されている。このような動きにより、データサイエンティスト不足の一定の改善が図られるものと思われるが、図-3に示したよ

うなアメリカの状況を鑑みれば、差は大きくなる一方であり、引き続き多くの大学でデータサイエンス学部・学科が新設されることが望まれる。

以上で挙げた大学のうち、滋賀大学、長崎大学、一橋大学、兵庫県立大学、立正大学、および総合研究大学院大学複合科学研究科統計科学専攻の6者はこの8月に「データサイエンス系大学教育組織連絡会」を立ち上げた。その趣旨は国公私の垣根を越えた連携により、データサイエンス分野の専門教育の推進や専門教員の養成に戦略的に取り組んでいくことである。データサイエンス系の専門学部設立やデータサイエンスのリテラシー教育をさらに進めていくには、この分野の専門教員の不足が深刻な状況である。専門教員の育成は一朝一夕には進まないため、政策的にも専門教員の育成への支援が強く求められている。

滋賀大学データサイエンス学部のカリキュラムと提供するコンテンツ

滋賀大学データサイエンス学部のカリキュラムは上で説明したデータサイエンス3要素のバランスをとるよう設計している。

学生にはノートパソコンを必携とし、情報系の講義演習により、プログラミングやデータ分析のツールに習熟させる。また統計系の講義演習により、データ分析やモデリングの理論を理解させ、それらを応用できる力をつけさせる。また、統計系のカリキュラムの中で、データサイエンスに必要な数学の講義と演習を必修としており、特に文系で入学した学生に数学の基礎力をつけさせている。

情報学と統計学のスキルの基礎の上で最も重視しているのが演習を通じたデータからの価値創造である。価値創造を体験するには、企業から実際の課題とデータを提供してもらいたい。滋賀大学データサイエンス学部では企業連携を積極的に進めている。2020年4月現在で企業等との連携協定数は50件以上となり、その他共同研究等を含む連携は延べ120

件以上となっている。以下ではこのような観点からの講義および演習の事例を3件挙げる。

1つ目は自主ゼミの形で行った「対話ロボットの作成」である。このゼミでは2年生4名が市川治教授の指導のもと、ソフトバンク社のPepperとIBM社のWatsonを用いて、滋賀大学に関する自由な質問(たとえば「トイレはどこにありますか」など)に対する応答システムを構築し、2018年夏のオープンキャンパスで披露した。システム構築には、音声認識、自然言語処理、音声合成の技術を組み合わせた。また大学に関する質問の教師データは学生が作成した。その様子はYouTubeに公開している(<https://www.youtube.com/watch?v=iF1nYhUvQy8>)。

2つ目は(株)ALBERTの協力のもとに行った特別講義「作って学ぶ機械学習」である。この講義では、深層学習の仕組みや実装方法についてALBERT社のデータサイエンティストが実践的な講義と演習を行った。講義は3日間(計18時間)の集中講義形式で、1日目は「ニューラルネットワークとはなにか」、2日目は「計算グラフと自動微分の実装」、3日目は「ニューラルネットワークの実装」を扱った。学部生や修士院生約20名が参加し、参加者はALBERT社の講師からさまざまな実践的なノウハウを学ぶことができた。

3つ目は(株)帝国データバンクの協力を得て2019年秋に実施した集中講義「データエンジニアリング人材養成演習」である。この集中講義では、オープンデータを用い、実践的なデータ研磨のスキルの習得に重点を

置いた演習が行われた。データ研磨とは、データ分析の前段階で必要とされるデータの前処理を指す。すでに述べたようにデータの前処理は意外に重要であり、そのスキルはデータサイエンティストにとっては必須である。

この章の最後に滋賀大学が公開しているMOOCコンテンツについて紹介する。滋賀大学は2016年12月に「数理・DS教育に係る教育強化」6拠点大学の1つに選定された。拠点大学としてのミッションはデータサイエンス教育の全学・全国への展開である。このことから、滋賀大学ではMOOC形式の「大学生のためのデータサイエンス」のシリーズ(I~III)を開発し、全学の教育に用いるとともに、無料で全国に公開している。そのIでは、統計の基礎、RおよびPythonを用いた分析例、実務での応用事例を紹介している。そのIIでは、機械学習のさまざまな手法を紹介し、深層学習についても詳しい説明を与えている。そのIIIでは、データ分析事例を用いてデータからの価値創造の方法を説明している。図-5にこれらのコンテンツ紹介の図を示す。

データサイエンティスト育成の加速を

ビッグデータの時代となり、多様かつ多量なデータが蓄積されるようになった。日本の多くの企業は蓄積されたデータを活かしきれていない。その主な理由は、データを処理し分析する人材すなわちデータサイエンティストの不足である。滋賀大学データサイエンス学部をはじめとして多くの大学でデータサイエンティストを育成する学部の開設が相次いでいるが、データからの価値創造を見据えたデータサイエンティストの育成をさらに加速することが求められる。

(2020年10月4日受付)



■ 図-5 滋賀大学の提供するMOOC講座

■ 竹村彰通 a-takemura@biwako.shiga-u.ac.jp

1976年東京大学経済学部卒業、1982年米国スタンフォード大学統計学部Ph.D.取得。1984年東京大学経済学部助教授。1997年東京大学経済学部教授。2001年東京大学大学院情報理工学系教授。2016年滋賀大学データサイエンス教育研究センター長。2017年滋賀大学データサイエンス学部長。