

巻頭

新しいデータサイエンスの展開

滋賀大学データサイエンス学部長 竹村 彰通

1. データの時代の到来とデータサイエンス

インターネットやスマートフォンの普及とともに、大量のデジタルデータがネットワーク上に蓄積されるビッグデータの時代となった。インターネット上のウェブサービスが開始されたのは約25年前であり、その後まもなくグーグルやアマゾンといったインターネットを基盤とする企業が生まれた。また約10年前にスマートフォンが登場し、スマートフォンによって人々は常時インターネットに接続することとなった。グーグルやアマゾンなどの企業は、ビッグデータを収集しそのデータを分析することによって、人々にとって魅力的なサービスを提供し、インターネット上の支配的な

地位を占めるようになった。日本は1980年代まで、高性能かつ高品質な工業製品の生産により世界をリードしていたが、その後のインターネットやデータをめぐる競争ではあまり存在感を示すことができず、アメリカに大きく後れをとることとなった。また最近よく報道されているように、キャッシュレス決済などインターネットサービスの分野では中国の発展が著しい。

このようなデータの時代を迎え、データを処理し分析する分野としてのデータサイエンス、そしてデータサイエンスを担う人材であるデータサイエンティストという職業が注目されるようになってきた。そしてインターネットにおける日本の後れの原因の一つが、データを分析する人材が不足

CONTENTS

巻頭

新しいデータサイエンスの展開	1
解説	
音などの新しいタイプの商標登録	7
授業実践	
全日制普通科「情報コース」の授業実践	11

授業実践

機械学習についての知識・技能を活用する授業実践	15
授業実践	
生徒の興味を喚起する実習を用いた授業実践	19
報告	
日本情報科教育学会第11回全国大会	23

していることであると考えられるようになった。日本におけるデータサイエンティストの不足は、実際の人材の雇用状況にも反映されており、多くの企業がデータサイエンティストを採用しようとしているが、ここ2、3年は人材がなかなか得られない状況が続いている。このように日本においてデータサイエンス分野での後れの認識が広まり、データサイエンティストに対する需要が顕在化したのは最近のことであり、以下でみるように人材育成対策の議論もやや急ごしらえの側面が強い。

なお、データサイエンスが対象とするのは必ずしも莫大な量のデータとは限らない。「ビッグデータ」という用語は時代の変化を表すキーワードの意味合いが強い。

2. 大学におけるデータサイエンス関連の人材育成に関する最近の議論の動向

まずは大学における人材育成に関する最近の議論の動向をみてみよう。2018年12月4日に、日本の経済界を代表する経団連が「今後の採用と大学教育に関する提案」を発表し、以下のような提案を行った：

多様な価値観が融合するSociety 5.0時代の人材には、リベラルアーツといわれる、倫理・哲学や文学、歴史などの幅広い教養や、文系・理系を問わず、文章や情報を正確に読み解く力、外部に対し自らの考えや意思を的確に表現し、論理的に説明する力が求められる。さらに、ビッグデータやAIなどを使いこなすために情報科学や数学・統計の基礎知識も必要不可欠となる。

そのため大学は、例えば、情報科学や数学、歴史、哲学などの基礎科目を全学生の必修科目とするなど、文系・理系の枠を越えて、すべての学生がこれらをリテラシーとして身につけられる教育を行うべきである。理系とされる学部でも語学教育を高度化する必要があるし、文系とされる学部でも基礎的なプログラミングや統計学の学修が求められる。さら

に、近い将来には、文理融合をさらに進め、法学部、経済学部、理学部、工学部といったこれまでの学部のあり方や学位のあり方、カリキュラムのあり方などを根本から見直すことが必要になるとと思われる。

経団連の提案には「データサイエンス」という用語は出てこないが、「ビッグデータやAIなどを使いこなすために情報科学や数学・統計の基礎知識」がデータサイエンスの基礎的な内容に対応すると考えられる。

この提案にあるSociety 5.0（ソサエティー5.0）とは、平成28年1月に日本政府が定めた「第5期科学技術基本計画」で提唱された未来社会像である。これまでの人類の社会の発展段階を、1. 狩猟社会、2. 農耕社会、3. 工業社会、4. 情報社会と分類し、これらに続く新たな社会であるSociety 5.0ではサイバー空間（仮想空間）とフィジカル空間（現実空間）を高度に融合させたシステムにより、経済発展と社会的課題の解決を両立する人間中心の社会が実現するとしている。

このような将来のビジョンを持つことは重要であるが、技術の変化を含め将来を予測することは困難であり、また大学には長い時間をかけて確立した学問的知識を継承・発展させる教育研究機能も求められるから、既存の学部の抜本的な見直しの具体的な内容は明らかではない。その意味では経団連の提案は、議論のための問題提起という側面が強いように思われる。

日本政府のさまざまな会議でも最近では「AI・数理・データサイエンス」教育が大きく取り上げられている。2018年9月の「統合イノベーション戦略推進会議」や「AI戦略実行会議」の会議資料（首相官邸のウェブサイトで公開）では、学校教育改革・大学改革と連動したAI・数理・データサイエンス教育の拡充（民間活用含む）が議論され、より具体的には次のような案が検討されている。

●文理を問わず普通高校、専門高校、高等専門学校校のAI・数理・データサイエンス教

育の抜本的充実，理数系教員の拡充，高等学校教育全般におけるSTEAM教育の充実による文理分断からの脱却

- 大学入試改革（大学全学部に数学，情報Ⅰ科目の採用）
- AI・数理・データサイエンス教育を3年以内に大学全学部学生に必修化
- 各産業分野においてAI・数理・データサイエンスの知見を活用できる人材を輩出する，産学が連携した，大学・大学院の仕組み／体制の整備
- リカレント教育による社会人へのAI・数理・データサイエンス教育の充実

なお，高等学校等での教育に関する第1点にあげられているSTEAM教育とは，Science（科学），Technology（技術），Engineering（工学），Mathematics（数学）を統合的に学習する「STEM教育」にArt（芸術）を4番目の要素として加えたものである。

これらの会議で議論されている案も経団連の提案も，AI・数理・データサイエンス分野における日本の後れを回復すべく早急な改革を目指しているが，すぐには実現の難しい提案が多い。政府の会議資料の第2点では，大学入試改革において大学全学部に数学，情報Ⅰ科目を採用するとしているが，それぞれの大学の入試は，大学の教育理念や経営方針にも左右されるものであり，また受験生への十分な周知も必要とされることから，急激な大学入試改革は難しい。このことは大学入試センターの新テストをめぐる動きからも明らかである。また第3点の「3年以内に大学全学部学生に必修化」も，とても実現できるとは思えない。まず，AI関連技術の進展は最近ほぼ毎日のように報道されているが，その学問的な内容はまだ流動的であり，標準的な内容として何を大学で教えるべきかは難しい問題である。例えばニューラルネットワークで隠れ層（中間層）の数を多くした「ディープラーニング（深層学習）」の手法は，画像認識等で従来の手法を大幅に上回る性能を示し

ているが，なぜそのような性能向上が得られるのかについて，教科書に書けるような十分定着した理論的な理解が得られているとは思われない。また，より伝統的な分野である情報科学や統計学についても，日本ではこれらの分野があまり重視されてこなかったこともあり，教える教員自体が不足しており，全学部学生必修化に対応できるような状況ではない。

第4点の産学連携教育や第5点の社会人のリカレント教育については，筆者も賛成であり，企業から大学への講師の派遣，大学院における社会人のリカレント教育，データサイエンス分野における長期の実質的なインターンシップ，博士人材の積極的な採用などは，現状の大学の枠組みにおいても積極的に進めるべきと考えられる。日本の企業は，これまで終身雇用を前提として企業内での人材育成を重視し，必ずしも大学での教育に期待してこなかった側面があるが，専門性を有する人材が転職することが当たり前のようになってきており，企業が大学に期待する教育のあり方も変化していると思われる。

以上のように，経団連の提案や日本政府の会議で議論されている案に示されている改革はいわば「掛け声」であり，早急な実現は難しいと思われるが，同時に，データ時代における日本の後れを取り戻すための着実な実践ももちろん必要とされている。いくつかの具体的な実践例については，以下の滋賀大学データサイエンス学部の教育の紹介で示す。

3. 高等学校及び中学校の次期学習指導要領におけるデータサイエンスの扱い

データ時代を迎え，高等学校や中学校でもより積極的にデータサイエンスの内容を学習することが望ましいが，そのような変更は多くの議論を経て，次期学習指導要領に取り入れられている。

まず「データサイエンス」が直接の内容として含まれる高等学校の情報Ⅱについてみてみよう。平成30年3月に発表された次期学習指導要領の情報Ⅱの内容は次の4つの項目からなっている。

- (1) 情報社会の進展と情報技術
- (2) コミュニケーションとコンテンツ
- (3) 情報とデータサイエンス
- (4) 情報システムとプログラミング

このうち(3)の情報とデータサイエンスでは「多様かつ大量のデータを活用することの有用性に着目し、データサイエンスの手法によりデータを分析し、その結果を読み取り解釈する活動を通して、次の事項を身に付けることができる」としている。まず知識及び技能として、

- (ア)多様かつ大量のデータの存在やデータ活用の有用性、データサイエンスが社会に果たす役割、目的に応じた適切なデータの収集や整理、整形
- (イ)データに基づく現象のモデル化やデータの処理を行い解釈・表現する方法
- (ウ)データ処理の結果を基にモデルを評価することの意義とその方法

をあげている。

また思考力、判断力、表現力等として、

- (ア)目的に応じて、適切なデータを収集し、整理し、整形
- (イ)将来の現象を予測したり、複数の現象間の関連を明らかにしたりするために、適切なモデル化や処理、解釈・表現
- (ウ)モデルやデータ処理の結果を評価し、モデル化や処理、解釈・表現の方法を改善

をあげている。これらはデータサイエンスの内容としてはかなり高度である。例えばデータ処理に関する解説を「指導要領解説情報編」(平成30年7月公表)で見ると、以下のように、回帰や分類などの手法を適切なソフトウェアを活用して理解するとしている。

データの処理に関しては、回帰、分類、クラスタリング及びそれらがどのような場面で活用されているか、これらを応用して人間が判断や意思決定を行う代わりにデータを基にどのような仕組みでコンピュータが判断を行っているかを理解するようにする。回帰に関しては、重回帰分析などについて扱い、そのモ

デルを変更することによって結果がどのように変化するか、分類に関しては、条件付確率、近傍法、木構造などを用いた予測について扱い、これらの手法や技術がどのような場面に活用されているか、それぞれ適切なソフトウェアの活用を通して理解するようにする。全体を共通の特徴を持ったいくつかの集団に分割するクラスタリングに関しては、似たものを集団にしていく階層的方法と、集団の数を決めてから要素を所属させていく非階層的方法などについて扱い、適切なソフトウェアの活用を通して理解するようにする。

このような内容は大学のデータサイエンスの講義でも当然扱うものであり、高大連携の観点からの教育も必要と思われる。

数学では、統計学の内容が強化されている。数学Iの(4)のデータの分析において、具体的な事象において仮説検定の考え方を理解することがあげられている。また数学Bの内容が、(1)数列、(2)統計的な推測、(3)数学と社会生活、となり、統計的な推測の比重が増している。

その内容は、知識及び技能としては、

- (ア)標本調査の考え方
- (イ)確率変数と確率分布
- (ウ)二項分布と正規分布の性質や特徴
- (エ)正規分布を用いた区間推定及び仮説検定の方法

があげられている。

また思考力、判断力、表現力として(ア)確率変数の平均、分散、標準偏差、(イ)標本調査の設計、があげられている。なお、数学Aの「場合の数と確率」についての指導要領解説ではベイズ論的な確率も言及されている。

平成29年3月に改訂された中学校数学では、これまで高校の数学Iの内容であった四分位範囲や箱ひげ図の内容が中学2年生の「Dデータの活用」に追加され、(ア)四分位範囲や箱ひげ図の必要性和意味を理解する、(イ)コンピュータなどの情報手段を用いるなどしてデータを整理し箱ひげ

図で表す、とされている。

このような情報や数学の指導要領改訂に対応して、日本統計学会が公認する統計学に関する標準的な試験である「統計検定」では、2020年4月より3級、4級の出題範囲を拡大することとし、出題範囲表の改訂版を公表している。筆者は2011年の開始当時より統計検定の運営にかかわっているが、統計検定3級及び4級での出題は、具体的な文脈の中でデータの読み方を問うものとなっており、高等学校や中学校での教育利用にも適したものである。

4. 滋賀大学のデータサイエンス学部の設立と提供するコンテンツ

数年前から日本におけるデータサイエンス教育の後れが急速に認識されるようになった中で、筆者の所属する滋賀大学ではデータサイエンスを専門に学ぶ学部を設立すべく名乗りをあげ、3年ほどの準備の後に2017年4月に日本初のデータサイエンス学部を開設した。定員は1学年100名である。準備はおおむね順調に進んだが、学内調整や大学設置審議会への申請も手間のかかるものであった。この動きは各方面から注目され、翌年の2018年4月には横浜市立大学にもデータサイエンス学部が設立されるなど、後に続く動きも活発である。

2節でも述べたように、データサイエンス関連の人材育成に関する経団連の提案においても、急速な大学教育改革が提案されているが、新学部の設立など実際の改革には時間がかかるのが実態である。滋賀大学から卒業生が初めて出るのも2021年の3月となる。

滋賀大学データサイエンス学部では、データサイエンスを、(1) 情報学、(2) 統計学、(3) 応用領域での価値創造、の3つの要素からなるものととらえている。入学する学生にはノートパソコンの購入を必須とし、1年生からコンピュータによるデータ処理や、統計パッケージの使い方を学ばせている。またデータサイエンスに必要な数学も基礎から教え、統計的手法についてもモデルの数

学的な仮定などをきちんと理解させるようにしている。このように、早めに情報学と統計学の基礎を身に付けさせ、高学年に進むにつれて、演習やゼミにおいて実際のデータ分析を行う。このことによって、応用領域での価値創造の成功体験を積むことができる。

企業や自治体からのデータサイエンス学部卒業生に対する期待は非常に高く、例えば昨年夏にはすでに2年生10名以上が長期の企業のインターンに行き、企業内でのデータ分析の課題に取り組んだ。

滋賀大学データサイエンス学部は、文部科学省の数理及びデータサイエンスに係る教育強化の拠点校として、北海道大学、東京大学、京都大学、大阪大学、九州大学とともに選定され、全国の大学へのデータサイエンスのコンテンツの提供も期待されている。

経団連の提案にあるような多くの学生へのデータサイエンス教育については、オンライン学習サービスであるMOOC (Massive Open Online Courses) の利用が有用である。滋賀大学では2つのMOOCコースを作成し、ドコモgacco社のプラットフォームgaccoを通じて無料で公開している。

一つは「高校生のためのデータサイエンス入門」であり、高校生が夏休みに視聴し、課題レポートを提出することにより、本学部のAO入試に活用している。このコースは全20回の講義で構成されており、高校生は2週間に渡って、RESASやe-Statを用いたデータの取得方法から、統計学の基礎、また分析の初歩が学べるようになってい



図1 「高校生のためのデータサイエンス入門」講座

表1 「高校生のためのデータサイエンス入門」の内容一覧

担当回	教員	タイトル
第1週第1回	伊達平和	この講義の概要とねらい
第1週第2回		データサイエンスとは
第1週第3回		公的データを入手する
第1週第4回		地域経済分析システム (RESAS) の利用
第1週第5回		政府統計の総合窓口 (e-Statの利用)
第1週第6回	松井秀俊	代表値
第1週第7回		標準偏差
第1週第8回		標準化
第1週第9回		ヒストグラム
第1週第10回		箱ひげ図
第2週第1回	姫野哲人	2変数データと散布図
第2週第2回		層別データの扱い
第2週第3回		相関係数とは
第2週第4回		相関係数の特徴
第2週第5回		相関係数と因果関係
第2週第6回	笛田薫	クロス集計
第2週第7回		層別クロス集計
第2週第8回		時系列と指数化
第2週第9回		時系列の移動平均
第2週第10回		時系列の季節調整

る。内容の一覧は表1にある。

またMOOCの第2弾として「大学生のためのデータサイエンスI」を公開している。このコースは4週全38回の講義で構成されており、大学生は1ヶ月に渡って、データ分析の初歩、Excel、



図2 「大学生のためのデータサイエンスI」講座

R, Pythonといったデータサイエンスでは必須となる統計処理ソフトやプログラミングの技術を学ぶことができる。さらに、画像処理や音声処理といったデータサイエンスの応用事例、さらには保険、金融、マーケットリサーチ、生命科学や品質管理といった現場での使用例によって、データサイエンスがカバーする多様な領域についても知ることができる。

MOOCは、インターネット環境があれば誰でも、またどこでも学べることから、本MOOCに加えて、2018年度中に「大学生のためのデータサイエンスII」も作成して、データサイエンスの諸手法をより詳しく解説し、2019年度に公開する予定である。