

## 神戸大学におけるデータサイエンス教育

### Data Science Education in Kobe University

大川 剛直 (神戸大学 システム情報学研究科 教授)

#### 要旨

膨大なデータの生成・蓄積に伴い、データに裏付けされた各種活動が、あらゆる分野で求められている。そのような中で、専門分野の枠を超えた数理・データサイエンスに関する教育の重要性が強く叫ばれている。神戸大学では、平成 30 年度から、国際人間科学部、経済学部、経営学部、理学部、工学部、農学部、海事科学部の 7 学部の入学生を対象に、「数理・データサイエンス標準カリキュラムコース」を開設し、文系・理系の枠にとらわれず、数理・データサイエンスの基礎を身につけさせるための教育を展開している。ここでは、数理・データサイエンス教育を支える柱として、「数理」、「統計」、「情報」、「データサイエンス」の 4 つの科目区分が設定され、それぞれに対応する科目が配置されている。本稿では、これらの科目区分の中でも、特に「データサイエンス」分野における教育に焦点を当て、カリキュラム編成に対する検討の経緯を交えながら、狙いや具体的な科目構成について概説する。

#### 1. はじめに

多種多様なデータが大量に生成され、比較的容易に入手できるようになっている現代社会において、これらのデータを活用して新たな価値創出につなげることが求められている。企業活動の様々な場面において、いわゆるビッグデータの存在を前提とした問題解決が重要視されているが、残念なことに、我が国は、世界の主要国の水準と比較すると、やや遅れをとっている現状にある。そこで、数理的な思考能力やデータの分析・活用能力を備え、これをもとに価値創出や問題解決を行うことができる人材の育成が急務とされ、それぞれの分野における専門的教育に加えて、数理・データサイエンスに関する教育が強く求められるようになってきた。

このような背景のもと、神戸大学においても、専門分野の枠を超えた全学的な数理・データサイエンス教育の必要性が広く認識され、その提供を主要な目的の 1 つとして、数理・データサイエンスセンターが平成 29 年 12 月に設置され、現在に至っている。

数理・データサイエンスセンターの設置に際し、全学に提供する数理・データサイエンス教育のあり方について検討することを目的として、同センターの設置準備ワーキンググループの下に、カリキュラム・産学地域連携サブワーキンググループ (以下、「カリキュラムサブ WG」と略す) が設けられた。その中では、数理・データサイエンス教育において

重要な役割を果たす「数理」、「統計」、「情報科学」、「データサイエンス」の4つの柱を対象に、必要な科目構成や教育内容などについての検討が行われたが、このうち、筆者は「データサイエンス」に関するカリキュラムについての取りまとめを担当した。

そこで、本稿では、現在、神戸大学において全学的に実施されている数理・データサイエンス教育の中でも特にデータサイエンス分野における教育に焦点を当て、その狙いや成り立ちについて述べる。

## 2. データサイエンスとは

データサイエンスという用語が、一般的に用いられるようになって久しく、また、世の中では、データサイエンティストが広く求められているなどと言われているが、そもそも、データサイエンスとはどのような学問であろうか。これについては、各所で様々な定義が与えられているが、「データから、より汎用的な知識を抽出するための学問」(Dhar, 2013: 64-73) とか「我々を取り巻く世界を理解し、予測することを目的としたデータ活用に関する様々な理論、アルゴリズム、方法論など」(UCI Data Science Initiative) といった説明が代表的であろう。

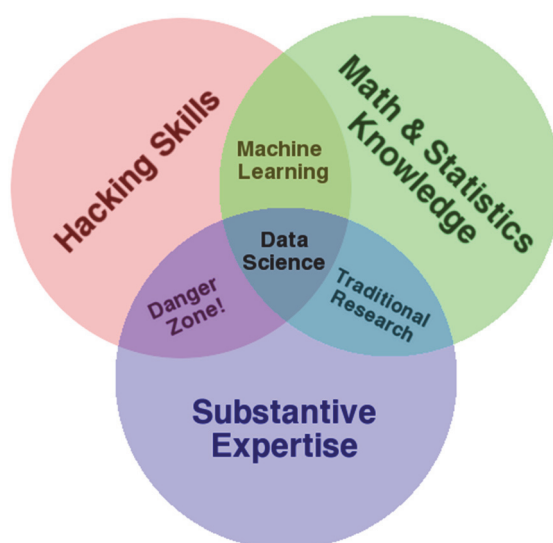


図1 データサイエンス・ベン図

出典：<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

そのような中で、米国のデータサイエンティストである Drew Conway が作成したデータサイエンス・ベン図(図1)は、比較的、その位置づけを明確に表現している。すなわち、数学・統計学の知識(Math & Statistics Knowledge)とITスキル(Hacking Skills)を備えるとともに、対象としている領域の専門性に長じる(Substantive Expertise)ための学問がデータサイエンスであるという考え方である。さらに、この図には、数理的・統計的な裏付けが

ないまま、単に対象とする専門領域において IT を活用して結果を導くことの危険性や、数学・統計学に基づいた計算機処理やアルゴリズムに関する学問は、いわゆる機械学習であって、それに加えて、対象領域における深い知識や洞察（その領域における本質的な課題は何であり、その解決によって、どのような価値創出が見込めるか）に裏打ちされて初めてデータサイエンスと見なすことができるといったことが示されている。データサイエンスやそれを実践する専門家であるデータサイエンティストについては、他にも色々な定義が与えられている<sup>1</sup>が、これら 3 つをデータサイエンティストが備えるべき重要な要素とする考え方は、比較的、一般的に認められているものと言えるだろう。

### 3. データサイエンスのカリキュラム編成

Conway の枠組みを借りれば、上述の通り、データサイエンスは

- ・ 数学・統計学
- ・ IT スキル
- ・ 専門知識

から構成され、まずは、これに準拠したカリキュラム編成を想定することが 1 つの出発点となる。一方で、カリキュラムサブ WG においては、「数理」、「統計」、「情報」、「データサイエンス」の 4 つを数理・データサイエンス教育における柱とし、それぞれの区分においてカリキュラムの具体化を進めるという方針が示された。単純に両者を照らし合わせると、少なくとも「専門知識」が後者に欠落していることがわかる。しかしながら、今回の数理・データサイエンス教育が、専門教育とは別に、その枠を超えた全学的教育として位置づけられることを考慮すると、「専門知識」に関しては、それぞれの専門領域（所属する学部）において教育されると考えるのが自然である。では、数学・統計学、情報技術、専門知識のそれぞれについて個別に学修すれば自動的にデータサイエンスが修得できるのであろうか。この答えに対する回答は明らかに否であり、それぞれの分野において養成した知識や素養を統合的に活用し、価値創造へとつなげていく思考やプロセスに関する教育が必要不可欠と言える。そこで、この部分を担うべきものとして、4 番目の柱である「データサイエンス」を位置づけることとし、その具体化のための検討を行った。

また、このようなカリキュラムの検討と並行して、各専門領域（学部）において、どのような人材像が望まれているか、さらに、それを実現するためにどのような教育が必要とされるかについての調査が実施された。学部によって、若干、取り上げ方に強弱はあるものの、多くの学部においてほぼ共通して、データに対する深い理解のもと、これを科学的に分析できる人材が望まれていること、そして、そのために、データを活用する能力の修

---

<sup>1</sup> 真面目な定義だけでなく、「データサイエンティストとは、いかなるソフトウェアエンジニアよりも統計学に精通し、いかなる統計学者よりもソフトウェア工学に通じている者である」などのジョークも有名である。

得を目的とした教育の充実を求める意見が見られた。

これらをもとに、「データサイエンス」区分におけるカリキュラムの構成として、

- (1) 全学の多くの学生を対象としたデータサイエンスの導入的教育
- (2) 全学の一部の学生を対象としたデータ活用に係る発展的教育
- (3) 問題解決・価値創造を実践する PBL 教育

の3段階の想定のもと、特に(1)と(2)を中心に科目の具体化について検討した。

(1)の導入的教育については、まず、数学・統計学、情報技術、専門知識の3要素を活用するデータサイエンスのイメージを理解させることを目的として、ビッグデータの活用を軸とするデータサイエンスが、社会の問題解決に何をもたらしているかについて、(必要に応じて事例紹介も交えて)講述する科目として「データサイエンス入門1・2」を準備した。加えて、そのようなデータサイエンスを実践するにあたって、何を学ぶ必要があるのかについて概観する科目として「データサイエンス概論1・2」を設定することとした。検討当時のシラバス案(抜粋)を表1に示す。これら2種類の科目は、数理・データサイエンス教育を実践する上で、必要最低限の科目設定と位置づけられる。すなわち、これらの科目の受講によってデータサイエンスを十分に学んだとは言えないが、少なくともデータサイエンスの必要性や意義について理解させることを目的としている。

表1 データサイエンス導入的科目のシラバス

	データサイエンス 入門1	データサイエンス 入門2	データサイエンス 概論1	データサイエンス 概論2
授業の テーマ	ビッグデータと社会との関わりについて学ぶ		データサイエンスを実践する際に必要となる技術の概要を学ぶ	
授業の 概要	<ul style="list-style-type: none"> <li>・ データサイエンスとは</li> <li>・ IoTとビッグデータ</li> <li>・ オープンデータ</li> <li>・ 各種応用事例</li> </ul>	<ul style="list-style-type: none"> <li>・ 人文科学分野におけるデータサイエンス</li> <li>・ 社会科学分野におけるデータサイエンス</li> <li>・ 生命科学分野におけるデータサイエンス</li> <li>・ 自然科学分野におけるデータサイエンス</li> </ul>	<ul style="list-style-type: none"> <li>・ データ収集・蓄積(センシング、データベース、データリテラシー)</li> <li>・ データ処理(信号処理(音声・画像)、データ構造・アルゴリズム、クラウドコンピューティング)</li> <li>・ データ解析(統計、多変量解析、データマイニング)</li> </ul>	<ul style="list-style-type: none"> <li>・ データ理解(モデリング、機械学習、人工知能)</li> <li>・ データ予測(シミュレーション、最適化、意思決定)</li> <li>・ データ可視化</li> </ul>
単位数	1	1	1	1

一方、(2)の発展的教育に関しては、機械学習や深層学習（ディープラーニング）などに代表される AI 技術や、データを活用するための様々な技法の修得を、主要な目的として設定するとともに、主たるターゲットを文系学生と理系学生の2つのカテゴリに分類し、前者に対しては、パッケージやツールなどの利用方法の修得を中心に、後者に対しては、基礎理論や原理など、より深い内容の理解を意図した科目を設定することとした。すなわち、各部局から寄せられたデータサイエンス教育に対する期待や要求を反映したカリキュラム構想となっている。主要な科目の概要について表2に示すが、検討の段階においては、学内の教育リソース（教員、教室、設備など）については基本的に考慮せずに、理想のカリキュラムの作成を目指した構成となっている。これらの科目は平成31年度後期以降に学部2年生から3年生を対象に開講することを想定しており、今後、既設の関連授業科目等の活用も視野に入れながら、実現可能性についての検討を踏まえて、具体的な実装が進められていくものと思われる。

表2 データサイエンス発展的科目の概要

主に文系学生が対象	データベース	データ解析演習	機械学習演習1	機械学習演習2
	データとデータベースの扱い方について学ぶ	統計解析ソフトRを用いたデータ解析の方法を学ぶ	Python、TensorFlow などの利用により、実際のデータを用いて機械学習を行う方法を学ぶ	
	単位数：1	単位数：1	単位数：1	単位数：1
主に理系学生が対象	データ工学1	データ工学2	機械学習1	機械学習2
	データを管理し、分析するための理論と技術を学ぶ		ニューラルネットワークを中心とした機械学習理論を学ぶ	統計的機械学習理論とその応用について学ぶ
	単位数：1	単位数：1	単位数：1	単位数：1

#### 4. おわりに

以上、神戸大学におけるデータサイエンス教育のためのカリキュラムについて、検討当時の経緯を踏まえながら概要を述べた。神戸大学では、平成30年度から、国際人間科学部、経済学部、経営学部、理学部、工学部、農学部、海事科学部の7学部の学生を対象に、「数理・データサイエンス標準カリキュラムコース」を開設し、文系・理系を問わず、数理・データサイエンス教育を実施している。上述のデータサイエンス関連科目のうち導入的教育に相当する部分は、数理科目、統計科目、情報科目の各区分を構成する科目とともに、同標準カリキュラムコースを構成する科目として、実装されている。まだ始まったばかりであり、現時点において教育の効果や改善点などについて議論するのは時期尚早と言えるが、入口となる入門科目の一部を実際に担当し、これまでに学んできた内容や興味の

対象が様々な学生が受講する姿を目の当たりにして、これら多様な学生に、いかにデータサイエンスの重要性を認識させ、学びに対するモチベーションを高めさせるかについて、一層の力点を置いた授業展開を行う必要があると改めて感じている。

現時点では、関連科目の一部を自前で既に開講している学部を中心に「数理・データサイエンス標準カリキュラムコース」が開設されているが、今後、他の学部に対しても同コースを拡大することが期待されている。教育リソースの増強が大きくは望めない現状においては、各学部学科で独自に開講されている科目を他学部や他学科の学生が受講できるよう、履修に際しての門戸を開くことが必要となるであろう。しかしながら、これらの科目の多くが積み上げ型の特性を有しているため、単に他学部他学科生に対して「履修可」とするだけでは制度として不完全である。そこで、例えば、その科目を履修して理解するために必要な知識や素養を備えているかどうかについて何らかの形で評価し、これに基づいて履修の可否を決定する仕組みの導入など、専門の異なる他学部他学科の学生（特にデータサイエンスに対して強い学修意欲があり、ある程度は独学で基礎的な勉強ができる学生。このような学生は、学問分野を問わず、必ず少数は存在するはずである）が、実質的かつ効果的な教育を受けることができるような制度の整備が望まれる。これにより、文字通り、数理・データサイエンス教育の全学展開が実現されるものと思われる。

本稿では、データサイエンス教育のためのカリキュラムの概要について述べたが、特に発展的科目に関しては、やや技術的側面に寄っている傾向にあるように思われる。あくまでも個人的な見解であるが、データサイエンスを実践する上で身につけるべき最も重要な素養は、ツールを使いこなす技量やデータの処理手法ではなく、価値創造を思い描く想像力と創造力ではないかと思っている。一般に、データの取得にはコストがかかるため、データを活用することで何ができるようになるのかをアピールし、データを取得・蓄積することの有用性を明確に示すことが重要である。その一方で、具体的なアピールを行うためには、根拠を示すための十分なデータが必要となり、いわゆる「鶏が先か卵が先か」のジレンマを抱えている。そこで、問題解決の対象としている領域において、どのようなデータがどのような形で得られそうか、そして、それを活用することで、どのような価値を生み出すことができるのかについて、きっちりデザインした上で、系統立てて説明できるような能力を養成することが、データサイエンス教育の本質ではないかと考えている。

このような素養は、知識や方法論を植え付けるような教育によって身につくものではなく、いかにして経験を深めるかが鍵となるであろう。これには、ありきたりではあるが、やはり実社会と連携した実践的課題を取り上げた PBL (Project Based Learning) をカリキュラムに取り入れることが効果的と考えられる。しかしながら、特に価値創造の要となる上流工程を対象とした場合、時間的制約に加え、教育する側の力量が大きく問われるなど、実施に向けての課題が山積しているのが実情と言える。

なお、筆者は、現在、数理・データサイエンスセンターの配置（兼担）教員として、同

センターが提供する科目の一部を担当しているが、カリキュラム全体について俯瞰できる立場にはない。本稿は、あくまでも構想段階における検討の過程に基づいた内容となっており、現在、実施されているカリキュラムとは正確には整合していない部分があることを付記しておく。

#### 参考文献

Vasant Dhar (2013) "Data Science and Prediction," *Communications of the ACM*, Vol. 56 No. 12, pp. 64-73.

UCI Data Science Initiative "What is Data Science?"

<http://datascience.uci.edu/about/> (最終アクセス : 2018 年 11 月 25 日)