

データサイエンス

成蹊大学理工学部情報科学科准教授/統計数理研究所客員准教授 小森 理



筆者が大学の学部生だった2000年初めの頃にはデータサイエンスの重要性は日本では全く認知されていなかった。そのような状況下で柴田里程先生のデータサイエンス研究室に修士まで所属し、指導を仰ぐ機会に恵まれたことは幸運だった。「これからはデータサイエンスの時代だ」とよくおっしゃっていたが、その本当の意味が今になって理解できるようになった。やはり先見の明がある先生であった。

データサイエンスが注目されるようになったのは1997年のJeff Wu教授のミシガン大学就任講演(Statistics=Data Science?)の頃からである。当初は統計学(Statistics)とほぼ同義であったが、近年のIT技術の発達やビジネス界でのビックデータの利活用と相まって、今ではその定義は様々である。よく用いられるものに米国のデータサイエンティスト Drew Conway のベン図がある[1](図1)。主に3つの領域から構成され、数理統計(Math & Statistics Knowledge)、IT技術(Hacking Skills)、専門知識(Substantive Expertise)に区分される。数理統計とIT技術の重なった領域が機械学習と呼ばれる領域で、AIの基幹技術を担っている[2]。囲碁のAlphaGoなどで用いられている深層学習もこの領域に含まれる。また数理統計と専門知識との重なりは従来型の研究領域であり、専門分野ごとに生物統計、医療統計、計量経済、マーケティングなど多岐にわたる。ここでIT技術と専門知識の重なる部分が「危険領域」(Danger Zone!)になっていることに注意したい。つまり数理統計の知識が下支えとなり、初めてデータサイエンスを実践できるということである。

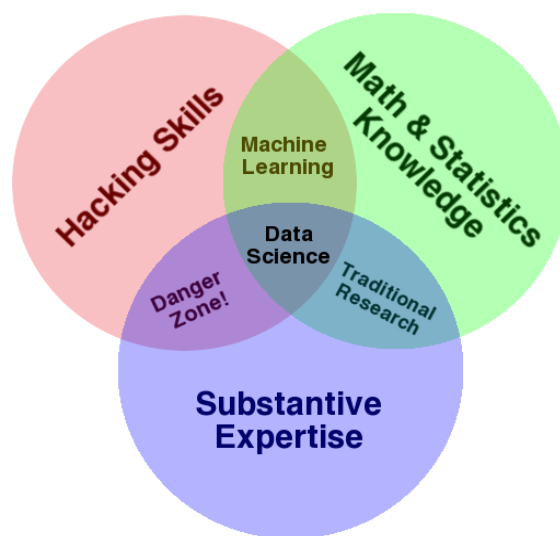


図1：データサイエンスの構成図

上記のことを踏まえ日本でも2015年頃からようやくデータサイエンスの重要性が認識され始め[3]、世界最先端IT国家創造宣言では「IoT、データサイエンス等、常に世界最先端の技術や知識の習得を積極的に支援する学習環境を整備」などの文言が明記された。このことを受け2017年には滋賀大学に日本で初めてのデータサイエンス学部が創設され、その後全国各地でデータサイエンスに関する学部、学科、センターなどの設置が相次いだ。今日では全国的な組織である「数理・データサイエンス教育強化拠点コンソーシアム」がデータサイエンスの教育拡充の中心的な役割を担っている[4]。この組織の頭文字に「数理」と付けられている意味も、図1から明らかであろう。また国連が定める持続可能な開発目標(SDGs)でもこのデータサイエンスが注目されている[5]。

成蹊大学でも2022年度からデータサイエンス副専攻を開始し、本格的にデータサイエンス教育に取り組むことになった。そのため今年上記の数理・データサイエンス教育強化拠点コンソーシアムにも連携校として参画することになった。米国のデータサイエンス教育でもPython、Rなどのソフトを使ったデータ解析、確率統計の授業、機械学習、それにビックデータ解析の授業などが行われており[6]、成蹊大学でもそれに即した形で副専攻のカリキュラムが作成されている。またデータサイエンスは実学の要素が色濃く、産業界との結びつきは大切である。データサイエンスがなぜ社会で重要視されているのか、またデータサイエンスが如何に実社会で応用されているのかを学生に実感してもらうことは非常に重要であり、PBL(Project Based Learning)の重要性が指摘されている[7]。成蹊大学では丸の内ビジネス研修などでもともと産業界との連携も強く、またサステナビリティ教育研究センターでは外部の研究者との交流も盛んである。このような成蹊大学の強みをうまく活用したデータサイエンス教育の推進が今後期待される。

<参考資料>

- [1] Drew Conway Data Consulting, LLC. 2015 「The Data Science Venn Diagram」 2020 年 11 月 8 日
<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>
- [2] AI 戦略 2019(統合イノベーション戦略推進会議決定)
- [3] 「我が国を支えるデータサイエンス力の高い人材育成」(総務省), 2015 統計関連学会連合大会
- [4] 数理・データサイエンス教育強化拠点コンソーシアム「数理・データサイエンス教育が未来社会を拓く」 2020 年 11 月 8 日 <http://www.mi.u-tokyo.ac.jp/consortium/>
- [5] 佐々木 史織, 「データサイエンスから挑む SDGs の実現 UN-ESCAP との取り組み」, CUC View & Vision No47 pp. 40-45 (2019)
- [6] 櫻井 尚子, 「データサイエンス概観— 教育を中心にして —」, 東京情報大学研究論集 Vol. 21 No. 1 pp. 51-59 (2017)
- [7] 大川 剛直, 「神戸大学におけるデータサイエンス教育」, 神戸大学 大学教育推進機構 『大学教育研究』 Vol. 27, pp. 1-7 (2019)

筆者のプロフィール

小森 理 (こもり おさむ)

成蹊大学理工学部情報科学科准教授/統計数理研究所客員准教授

2005 年慶應義塾大学理工学部数理科学科卒業. 2007 年同大学院基礎理工学科修了. 2010 年総合研究大学院大学複合科学研究科修了(博士 統計科学). その後 2015 年 8 月まで統計数理研究所に勤務. 同年 9 月から 2018 年 3 月まで福井大学で勤務. 同年 4 月から現職の成蹊大学で勤務. 専門は生物統計, 医療統計, 機械学習, データサイエンス. 日本統計学会, 日本計量生物学会, 応用統計学会に所属. 現在は日本計量生物学会誌編集委員, Japanese Journal of Statistics and Data Science, associate editor, 応用統計学会の理事を担当. 2016 年には日本計量生物学会から奨励賞を授与される.

<https://www.ci.seikei.ac.jp/komori/>