

[AI 人材教育]

4 早稲田大学における AI 人材教育

応 一般 —学生から社会人まで—



松嶋敏泰 | 早稲田大学

AI・データサイエンス人材

早稲田大学におけるデータサイエンス人材教育を述べるにあたって、そもそもデータサイエンスとは何か、データサイエンス人材とは何かを、少々遠回りになるがまず考えてみたい。

AI とデータサイエンス

まず AI (Artificial Intelligence : 人工知能) の出自は明確で、有名な 1956 年のダートマス会議で、情報を入力として高度な判断を出力する人間の学習等の知能をコンピュータにより模倣することで、知能とは何かを探求する重要な研究テーマを、4 名の著名な科学者が提案した。しかし知能の模倣は結果として高度な情報処理機能を実現することになる。そのため興味がさまざまな応用へと移っていき、人間についての科学研究から工学的研究へと目的がシフトしていった。

1959 年に A. Samuel によって Machine Learning (機械学習) という用語もつくられ、コンピュータで学習機能を実現するという意味では知能全体を目指す AI の一分野と捉えられるが、情報を与え続けて学習させることを主眼とするより、情報から認識や判断の意思決定を行う高度な情報処理全体を表す意味合いが強くなっていった。

しかし、最近の第 3 次 AI ブームでは、AI や機械学習の 1 つの手法である深層ニューラルネットが注目を浴び、AI というこの深層ニューラルネットを指すと思う方もいて、AI は機械学習に含まれるのかその逆かというような問いが出てきてしまうようだ。

それでは、データサイエンスまたはデータ科学の出

自については、1974 年に P. Naur の著書の中に Data Science という言葉が登場するが、ほぼコンピュータサイエンスと同意語のように使われていた。それでは現在の概念はどのあたりから登場してきたのであろうか。まず、統計学の分野では 1960 年ぐらいから、Data Analysis (データ解析) が使われていた。1990 年前後から Data Mining (データマイニング) も登場する。これは、データから知識を抽出するという帰納 (事実から法則を導出する推論) を強調した用語ではあるが、学習とは帰納を行っているといえることもでき、扱われる問題や手法・アルゴリズムも拡張解釈された機械学習とほぼ同じであった。データサイエンスの定義はさまざまであるが、上に挙げた AI、機械学習、統計学、データ解析、データマイニング等を合わせた概念、学際的学問領域であるという定義はおおむね承されているようだ。

しかし、実はデータサイエンスの意味はこれだけではない重要な別の側面がある。それは J. Gray の “Data science as a fourth paradigm of science” という言葉に象徴される、いわゆるデータ駆動型の知の創造パラダイムとも呼ばれる理念である。Data Deluge (データ洪水) とも呼ばれる、数値のみならず、文字、音声、画像などの多様なデータが容易に世界中から収集可能になった、情報・通信インフラの発展による社会環境の劇的変化と、先に述べたデータを分析する理論や技術の進歩が融合することで、研究や仕事を含む人間の知的活動にパラダイム・シフトが起こりつつある。

データサイエンスは 1 つの科学領域と捉えるより、科学的思考や意思決定の方法の科学であり、メタサイエンスであると捉えられる。AI が人間の知的活動は何かを問うものであったのに対し、データサイエンスは人間の

知的活動を飛躍、変容させるものと言えるかもしれない。

データサイエンティストとデータサイエンス人材

データサイエンティストという言葉が有名になったのは2012年のHarvard Business Reviewの“Data Scientist: The Sexiest Job of the 21st Century”で、おそらくビジネスサイドの方はこの言葉からデータサイエンスを知ることになり、buzzword的なビジネス用語として使用されているケースが多くなったとも言われている。

前節でも述べた社会の変革期を迎え、データサイエンス人材の不足を懸念する声が多岐な方面から上がっている。ビジネス面を強調したデータサイエンティストに限らず、これからはすべての分野においてデータサイエンスが必要であり、ある専門分野に特化して、あるいは分野を選ばず、データを活用し問題解決ができるスペシャリストが必要となろう。しかし、必要なデータサイエンス人材とはそのようなスペシャリストだけなのだろうか。

もちろんそのようなスペシャリストは今後ますます珍重されるであろうが、発想を変え、すべての分野において、自身の専門性をベースにデータを活用して意思決定ができる人材が育てば、スペシャリストへの依存度は下がっていくと考えられる。すべての専門領域でそのような人材を育てていくことが、本当のデータサイエンス人材の育成と考えてもよいのではないだろうか。

早稲田大学の データサイエンス教育の概要

データ科学センターの設置

早稲田大学では、ここまで述べてきたように人類の意思決定や知の創造のプロセスが大きく変わるこの変革期に、最も変わらなければならないのは、知の創造と継承の拠点である大学の研究・教育ではないかという視点から、早稲田大学データ科学センター^{☆1}を2017年12月に全学的な組織として設置した。

☆1 <https://www.waseda.jp/inst/cds/>

早稲田大学がデータサイエンス学部や研究科ではなく、全学的なデータ科学センターを設置した狙いは、前章で述べたようにどの専門分野においてもメタサイエンスとしてのデータサイエンスが必要であり、また逆にデータサイエンスは各専門領域の知識と融合してはじめて威力を発揮するという考えからであった。つまり、総合大学の強みである人文社会系や理工系のさまざまな分野の専門知識とデータサイエンスの融合を軸として、研究面と教育面の両面での効果を目指している。

研究面ではデータを基にした知の創造プロセスの進化・深化と、さまざまな専門分野との融合による革新的な研究、たとえば複雑でグローバルな社会問題の解決等を狙っている。教育面ではそれぞれの専門性の上にデータを活用する能力を持った社会で有用な人材の育成を目指している(図-1)。

全学共通データ科学教育プログラム

早稲田大学ではデータ科学センターを中心に、どの学部、研究科の学生も自分の興味や、将来の希望に合わせて、後で詳しく述べる初級(データサイエンス全般を把握している)、中級(自身の専門にデータサイエンスを活用できる)、上級(データサイエンスのスペシャリストに近い)の各レベルを目指して、オーダメイドに教育が受けられる全学共通データ科学教育プログラムを展開している。

先に述べたように早稲田大学では、それぞれの専門性の上にデータを活用する能力を持った人材の育成を目指しているが、身につけさせるデータサイエンス能力のレベルは、各学生の必要性に応じて柔軟に設定されればよいと考えている。このようにそれぞれが目指す社会での活躍の場に必要データ活用能力を持ったデータサイエンス人材を、卒業生として毎年一万人輩出できればと思っている。

この総合的なデータサイエンス人材の育成プログラムは図-2の(I)~(V)の5つのブロックで構成されていて、授業だけでなく多面的に専門性とデータサイエンス力を兼ね備えた学生を育成する全学共通の総合教育プロ

特集
Special Feature

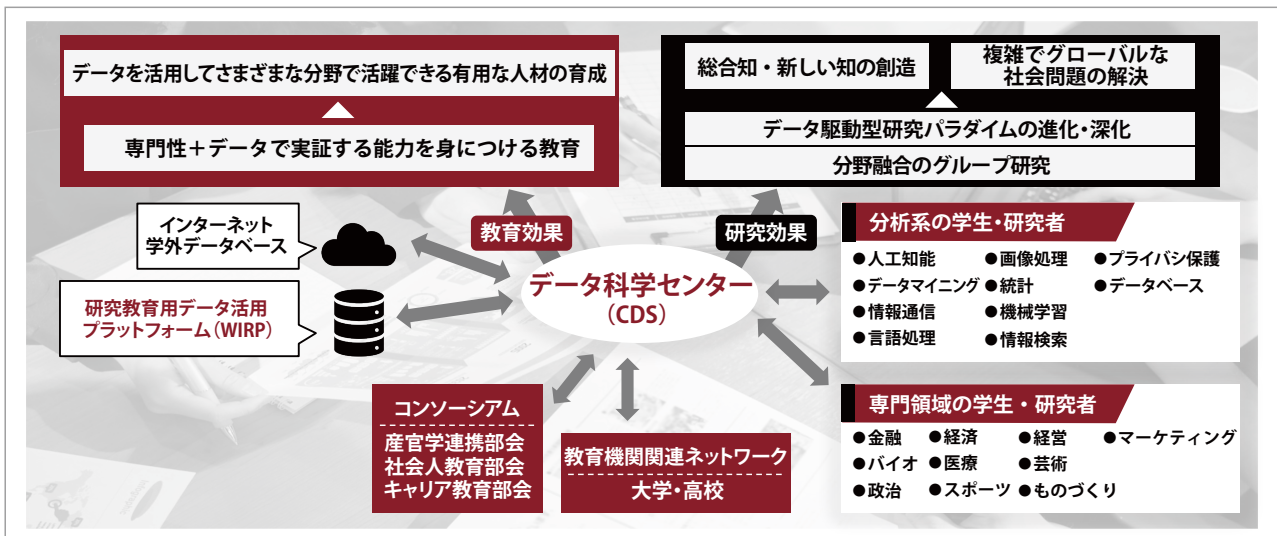
ラムとなっている。この育成プログラムの中心となるのは (I) のブロックで、データ科学センターが、全学への教育を担うグローバルエデュケーションセンター (GEC) を通じて展開している共通科目群である。この共通科目群はどの学部、研究科の学生も履修することが可能で、早稲田独自の特徴的カリキュラムになっている。

以下の章では、まずこの (I) について詳しく述べ、次に (I) の共通カリキュラム以外の (II) ~ (V) の総合的で多面的な育成プログラムについて説明する。また、早稲田大学では人材育成を大学内に限って行おうとは考えておらず、社会人に対してこのデータ科学センター

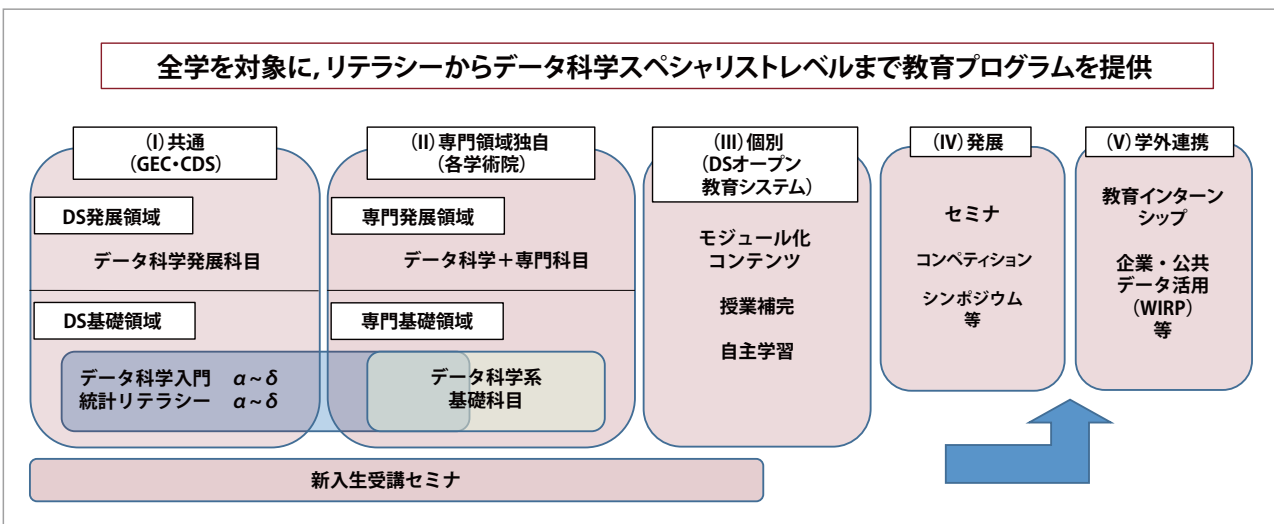
の教育プログラムを提供することも視野に入れており、それについても最後の章で述べてみたい。

データサイエンス教育の基本方針と共通カリキュラム

データ科学センターが GEC を通じて展開している全学の学生が履修可能な (I) のデータサイエンスの共通カリキュラムについてまず説明する。この共通カリキュラムにはデータ科学センターのデータサイエンス人材育成教育の基本方針と工夫が凝縮されている。



■ 図-1 早稲田大学データ科学センターの概要



■ 図-2 早稲田大学 全学共通データ科学教育プログラム

さまざまな専門分野があることが総合大学の強みであると述べたが、さまざまな学術院（学部と大学院をあわせた組織）の専門分野と融合するようにデータサイエンスをいかに教えるかは、実はそう簡単ではない。各学術院独自の専門カリキュラムの中に新たにデータサイエンスの共通カリキュラムを組み込むためには、各学術院のカリキュラムの時間割に干渉しないように、学習時間を圧迫しないようにと工夫が必要となる。もっと大きな問題は、私立大学ではセンター試験を通過しない学生の数理系知識のレベルに差がある上、それぞれの学術院の学生が身につけている専門知識や興味の対象も多様である。

これらの問題をクリアするためには、まず育てたい人材に身につけさせたい内容を整理し、データサイエンスで最も重要な本質的内容を絞り込むことが必要であった。それは結局データサイエンスとは何かという前章の間に戻ることとなり、我々としてのデータサイエンスの定義を明確にすることからはじめた。AI、機械学習、統計学、データ解析、データマイニング等における共通した考え方をまとめ、データサイエンスとは「データから意思決定を行う明確で合理的なプロセスを扱う学問体系」とシンプルに定義した。

明確なという意味は、意思決定の問題や目的やプロセスが曖昧性なく誰に対しても説明できるということで、そのために問題等の記述言語として数学が用いられる。合理的な意思決定の部分は、論理的な演繹法を用いて結論が導出されるということで、ここにも数学やコンピュータによる演繹が用いられる。

これだけ数学的要素があると数理系に不得意意識のある文系の学生には、やはりデータサイエンスを教えるのは難しいのではと思えるが、重要なのはデータを用いて決定したい問題と目的は何かを明確に整理し記述できたかという点である。たとえばこの共通カリキュラムでは、データを入力とし意思決定の結果を出力とする写像により、問題を数理的に記述することの大切さをまず教えている。

このような論理的考え方の記述言語として最低限の数

学の利用は避けられないが、体系的であるが数学を極力抑えた、文系の学生にも優しい構成を心がけた。どの学部の学生にも興味が持てるように、平易な表現や専門に近い例題を用いる等さまざまな部分で工夫をしている。

また、データサイエンスの考え方のプロセスを再整理することで、オーソドックスなデータサイエンスの教育カリキュラムのように、数学や最適化理論の基礎知識を教えてから、統計学、機械学習、AI等と順番に教えていくような構造はとらず、すべてを合わせて教えてしまふまったく新しい発想のカリキュラムとした。この利点はそれぞれの考え方の違いがより鮮明になり本質的な考え方が身につく、また個別に教えるより重なりを少なくし短時間で教えることが可能となる。

具体的には、まず、各学術院の専門カリキュラムを圧迫しないように1年間という時間制約の中で、どの専門分野の学生もデータサイエンスの全体像が学べるよう、初級レベルのデータ科学の入門シリーズとして、「データ科学入門 α , β , γ , δ 」の4つのクォータ科目を開講している。この授業は、動画、スライド、小テスト、レポート、試験で構成されるフルオンデマンド型で、各学術院の時間割と干渉せず、学生は空き時間にPCやモバイル端末でも学習できるようにしている。

このオリジナルなデータ科学入門シリーズとは別に、オーソドックスな統計学を中心とした入門シリーズの「統計リテラシー α , β , γ , δ 」がある。これも4つのクォータのフルオンデマンド授業で、すでに数年前から共通カリキュラムとして開講されていて、この統計シリーズを開発したノウハウがデータ科学入門シリーズでも活かされている。この2つのデータサイエンス系の入門シリーズだけで全学の受講者数は昨年度の場合、年間延べ8,000名以上となっている。

このフルオンデマンド型のコンテンツにはもう1つ工夫がある。各コンテンツは、より細かいモジュールで構成され、それぞれのモジュールはある程度独立性を持たせた内容となっている。このようにモジュール化されていることで、コンテンツ内のモジュールを組み替えることで、学生の専門分野や興味の多様性やレベルに対応

する柔軟なカリキュラムが構成可能となる。また、図-2の(III)の学生の自主学習や授業補完の目的で独立した利用も可能となり、これについては次章で詳しく述べる。

この初級レベルの次は、早稲田大学が育てたいデータサイエンス人材の中心レベルである、データサイエンスを自身の専門の研究や仕事に活用できる中級レベルの共通カリキュラムとなる。これらも初級レベルと同様なフルオンデマンド型の授業で、各学院の専門とデータサイエンスをつなぐ科目となる。たとえば、それぞれの専門分野で必要となる回帰構造、時系列構造、潜在構造等を有するデータを扱う科目となり、順次開講を進めているところである。

そして最後は上級レベルで、自身の専門領域外のデータに対しても問題を数理的に抽象化し、データサイエンスを活用することができるレベルを目指す。もちろんデータサイエンティストやデータサイエンス研究者は理工学術院等で育成されているが、他の学術院の学生でもこの全学の共通カリキュラムだけで、データサイエンスのある程度のスペシャリストにまで到達可能なカリキュラム構成になっている。

また、数学は最低限にとしたが、やはり数学は必要であり、文系でも数学をある程度押さえておきたいと思う学生もおり、フルオンデマンド授業の「データ科学のための数学」も用意している。また、ソフトの利用については、やはりフルオンデマンド授業の「Rによる統計解析」があるが、ソフトの使用法だけの正規科目は設置していない。これは、科目として教えなくともさまざまな使いやすいアプリがあるということと、使用法だけ教えることで、逆になぜその結果が出てくるのかの基本的考え方や理論を知らず、盲目的に使うことの危険を回避したいためでもある。しかし、実際に手を動かすことを重視したハンズオンセミナーやワークショップ等もセンターでは定期的に開催している。

このような対面でないオンデマンドの授業を補完するため、常設の相談窓口も設置し、授業に関する質問に限らず、高度なデータサイエンス全般の質問にも答えられるLA (Learning Assistant) が対応している。さら

に高度な研究活動のサポートとしては、全学の大学院生、ポスドク、研究員、教員向けにデータ科学研究相談窓口も設置されている。

総合的・多面的データサイエンス人材育成プログラムとシステム

前章で解説した共通カリキュラムは、図-2の早稲田大学の総合的なデータサイエンス人材育成を目指した全学共通データ科学教育プログラムの(I)にあたる。この節ではそれ以外の(II)～(V)の4つのブロックで構成される、専門性とデータサイエンス力を兼ね備えた学生を、授業だけでなく総合的に多面的に育成するしくみについて以下に説明していく。

(II)は各学院の学部や研究科に設置されているデータサイエンス関連の基礎科目や、それぞれの専門分野にデータサイエンスを応用した科目である。(I)(II)の連携の科目として、GECに設置されている共通科目をそのまま学部の必修科目として利用している学部もある。

(III)はデータ科学オープン教育システムで、学生個人がスキルアップのためや研究での必要性から、自主的にコンテンツの視聴ができるようになっている。また、教員が担当授業の中で、コンテンツを部分的に反転授業のような授業の補完目的で利用することも可能となっている。

(IV)は各種シンポジウムやデータサイエンスコンペティションである。たとえば昨年コンペティションでは参議院選の当落予測を課題として、全学から約200名、61チームが参加し、5社の協賛もいただき開催した。活気ある大会となって、最優秀チームは、政治経済学術院のチームとなり、専門性の上にデータサイエンスが活用されるという人材育成の狙いが期せずして現れる結果となった。コンペティションの狙いは、競わせることでのモチベーションの高まりもあるが、教育的効果が真の目的で、的中率よりデータからデータサイエンスの適切な考え方のプロセスに従い予測を行ったかが重要な評価基準となっている。

(V)は企業との連携で、データサイエンス人材の育

成を目指す教育プログラムである。後で詳しく述べるが、データ科学センターが運営するデータ科学研究教育コンソーシアムの3つの部会の1つにキャリア教育部会があり、その活動の一環として教育的インターンシップを行っている。データサイエンスの教育において、実務の現場でデータサイエンスがどのように活用されているかを学生が体験することは大変有意義であり、このインターンシップではコンソーシアム会員企業のご協力により、たとえば5日間以上の長期日程でデータを活用する実務の現場で実習を体験させていただいている。

また、データサイエンス教育において実データを分析することは大変重要である。早稲田大学は多くの企業と学術交流協定を結んで、銀行、広告代理店等の貴重なデータを無償で提供していただき研究教育に利用させていただいている。これらのデータを利用するためには、個人情報の保護、秘密保持の問題があるが、センターではWaseda Integrated Research Platform (WIRP) というデータ利活用のプラットフォームにより、秘密保持の契約面と情報システム面でセキュアな利用を推進している。たとえば、情報システム面では、データストレージとデータ解析用サーバはデータ科学センター内にあり、各研究室からの利用者はリモートデスクトップでこのシステムに入り、データ自体をダウンロードせず研究が行えるセキュアなプラットフォームとなっている。

学外への教育プログラムの展開

この章では、データ科学センターの教育プログラムの学外特に社会人への展開について述べる。前章で、企業側から大学への教育協力プログラムとして、教育的インターンシップを述べたが、逆に大学内で展開している教育プログラムの企業への展開も視野に入れている。専門性とデータを活用できる能力を持った人材の育成は大学内だけの課題ではなく、先に述べたように社会全体の危急の課題である。センターでは人材育成を大学内に限って行おうとは考えておらず、社会人に対してもデータ科学センターの教育プログラムを提供すること

で、すでに業務で培われた専門性にデータを活用するスキルを持った人材の育成が可能と考えている。

センターの教育プログラムはさまざまな専門や基礎知識のばらつきを吸収する柔軟なカリキュラムとなっているため、社会人にも有効な教育効果を発揮することが期待される。さらに、フルオンデマンドであるため、社会人にも学習しやすく、モジュール化されたコンテンツはニーズに合わせて組み換えも可能となっている。

すでにコンソーシアムの社会人教育部会で教育プログラムの展開がはじまっており、会員企業においてセンターの教育コンテンツを利用した社内教育が試験的にスタートしている。コンソーシアムについての全体説明が後回しになってしまったが、以上のような企業との連携の仕組みとして、以下の3つの部会で構成される早稲田大学データ科学研究教育コンソーシアムを運営している。

- 1) 産官学連携部会：データ科学にかかわる産官学連携による共同研究プロジェクトの推進。
- 2) 社会人教育部会：データ科学分野における社会人教育プログラムの開発・運用。
- 3) キャリア教育部会：データ科学およびキャリア教育による次世代人材育成。

紙面の都合で紹介できなかったが、企業と同様に他大学との連携も重要と考え推進している。データサイエンス人材育成には、学内の教育プログラムだけでなく、ここで挙げた教育インターンシップや社会人教育など企業をはじめ学外との連携が不可欠であり、このような包括的なデータサイエンス人材育成プログラムが日本の将来にも大きな影響を与えていくと考えている。

(2020年11月4日受付)

■松嶋敏泰 (正会員) toshimat@waseda.jp

早稲田大学理工学術院応用数理学科教授、データ科学センター所長。1993年早大理工学部助教授。2001～02年ハワイ州立大・電気工学科客員研究員。2011～12年カリフォルニア州立大・バークレイ校・統計学科客員教員。電子情報通信学会 基礎・境界ソサイエティ会長。情報理論研究専門委員会委員長。情報理論とその応用学会副会長。品質管理学会理事。人工知能学会、電子情報通信学会、品質管理学会論文誌編集委員等。