

生命科学部における情報科学およびデータサイエンス教育について

野口瑠¹ 森河良太² 西田洋平³ 野口航⁴ 小島正樹⁵ 高須昌子¹

1. 社会的潮流

人工知能に関する社会の動向について説明する。

1. 1. 人工知能

人工知能(artificial intelligence; AI)は、1950年にアラン・チューリングによる「機械は思考できるか?」という疑問を発端とし[1]、1956年のダートマス会議においてジョン・マッカーシーらによって定義された。当時、機械学習[2]やニューラルネットワーク[3]もすでに誕生していた。機械学習とは人間が自然に行なっている学習という行為をプログラム(機械)に実装するための手法である。しかし、計算機器の性能不足や学習するための膨大なデータが存在しないなど、環境面が整わないことから日の目を見ず、AIは冬の時代を迎えることとなった。

2000年代以降、インターネットの普及やビッグデータ・オープンデータの形成、CPU・GPUの価格低下などの条件が揃い、機械学習の適用事例が増加した。この2000年初頭から現代までの盛況は第三次AIブームと言われているが、大きな趨勢となった理由に深層学習の存在がある。深層学習とは、多層化したニューラルネットワークを用いた機械学習の1つである。この躍進は凄まじく2012年に画像識別のコンペティションであるILSVRCにおいて深層学習が既存の手法を凌駕した[4]。また、2016年にはDeepMind社が作成したAlphaGo[5]が、通常では困難とされた実在棋士を早々に破り、深層学習の進化が人々の想像以上であることを世界に発信した。深層学習が既存の機械学習手法よりも優れた成果をあげている理由として特徴量の取り扱いにある。特徴量とは機械学習による予測を行うために入力される数値で、データの性質を十分に表現する値である。この設計は、既存の機械学習手法では人の経験や試行錯誤によって行われてきたが、深層学習では設計を含めての学習(特徴量表現学習)が可能となり、より簡便で高精度な予測が可能となった。

1. 2. 生命科学分野における適用例

生命科学分野における深層学習の適用について、最も衝撃的であったことはタンパク質の立体構造予測であると思われる。タンパク質の立体構造予測精度を競うCASP14においてDeepMind社が作成した深層学習を用いた予測プログラムAlphaFold2が他を圧倒した[6]。本ソフトウェアは公開され、タンパク質の立体構造予測だけではなく、タンパク質複合体の形状予測、ペプチドのドッキングなどの新しい利用方法が模索されている[7]。また、CREST「データ駆動・AI駆動を中心としたデジタルトランスフォーメーションによる生命科学研究の革新(バイオDX)」が2021年より開始した。本プロジェクトはデータ駆動型科学の生命科学分野における適用である。例えば、自動化・画一化

¹生命科学部 生命物理科学研究室 ²情報教育研究センター ³東海大学 教職資格センター

⁴生命科学部 応用生態学研究室 ⁵生命科学部 生物情報科学研究室

された実験により、複数の拠点で24時間365日データを取得し続ける。こうして生み出されたビッグデータを統計解析することによって、創薬やタンパク質設計に留まらず、進化や細胞増殖メカニズムの解明などの理解を明らかにしようとしている[8]。

1. 3. デジタル社会への課題

このような技術革新が起こっている中で、それに従事できるIT系人材の養成は伴っていない。2019年の試算においてIT人材の不足は2022年で約47万人、2030年では78万人まで拡大される可能性が指摘されている[9]。また、日本ではIT人材の殆どがIT企業に就職しており、その偏りが問題視されている。デジタルトランスフォーメーション(DX)を行うためには組織内部にIT人材を配置し、迅速な体制変革を行うことが望ましい。しかし、実際にはIT人材のほとんどがIT企業に在籍していることから、DXをするにあたり各企業はIT企業へ外注せざるを得ない。このようにシステム開発を内製できない歪んだ社会構造がDXの推進を阻んでいることが指摘されている[10]。

IT人材の不足の問題は、バイオテクノロジー分野では特に深刻である。文部科学省の2020年の調査では医薬品製造業の企業に勤務する20928人の研究者のうち、情報科学を専門とする研究者は56名しかいない[11]。同様にバイオインフォマティクス人材の不足も明らかで、取り分けて生物学と情報科学、ウェットとドライ双方の理解を持った人材の不足が顕著である[12]。

IT人材の中でAIなど高度な技術を取り扱う人材をデジタル人材と称する。IT系人材の不足から容易に想像できるが、デジタル人材の不足が深刻化することが指摘されている。そのため、日本は、デジタル人材養成として、高等教育における情報の必修化や文理を問わず数理・データサイエンス・AIを習得するなどの目標を設定した[13]。そして、2021年には文部科学省主導による数理・データサイエンス・AI教育プログラムの認定が開始した。本プログラムは年間50万人の大学生・高専生が数理・データサイエンス・AIの初級レベル(リテラシーレベル)を習得することを目標としている。

2. 生命科学部のこれまでの取組

このような社会的な潮流を受け、生命科学部では2020年度からの新カリキュラムにおいて3つの特別プログラムが発足し、そのうちの1つとしてデータサイエンスプログラムが設定された。本プログラムは数理・データサイエンス・AI教育プログラムのリテラシーレベル取得を目標としており、その取得のために、カリキュラムなどが一部変更となった。ただし、生命科学部ではデータサイエンスが社会的潮流となる以前より、情報科学教育に力を入れてきた[14, 15]。そのため、本プログラムの導入は比較的にスムーズに行われたと思われる。

また、本プログラムの前身として2019年度から旧カリキュラムにおいて「バイオ・データサイエンス入門プログラム認定証」の制度が始まった。この認定証は、旧カリキュラム(2019年度までに入学)の学生を対象に、規定の8科目全てを履修した者に発行される。ここで8科目とは、旧カリキュラムの数学I・II、情報科学I・II、統計学、応用数学、プログラミング基礎、バイオ

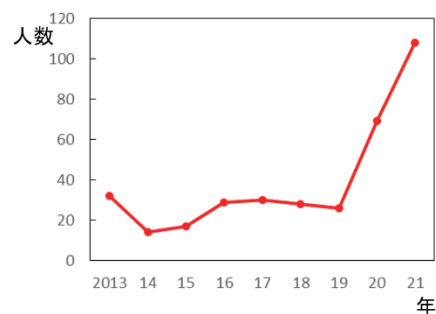


図1 「バイオ情報科学」(旧カリキュラム)における履修者数の変化

情報科学である。本認定証の発行は、2019年度2名、2020年度7名、2021年度24名（2022年1月6日現在）である。また、図1に旧カリキュラムのバイオ情報科学の履修者数を示す。データサイエンスプログラムが始まった2020年度から急激に履修者が増加しており、データサイエンスに対する関心の高さがわかる。

本学部が情報教育に力を入れている理由は、生命科学と情報科学が密接に関連した学問であるからに他ならない。例えば、ニューラルネットワークは神経細胞を源流とする数理モデルである。また人工知能や機械学習は、人が自然と行っている知的行動を機械(プログラム)として実装することである。さらに生命現象も最終的には情報科学の問題に行き着く。例えば、コドン暗号が3塩基からなるという生命科学では定番の問いは、4つの塩基で20種のアミノ酸を記述するために必要最低限のビット数問題として捉えることができる。また、DNAをデータストレージとする研究も進んでおり[16]、DNAに画像ファイルを保存するなどの技術が開発されている。このように、情報科学の研究は生命へ、生命科学の研究は情報へ相互に向かい進んでいる。

3. データサイエンスプログラム認定について

データサイエンスプログラム認定は生命科学部独自の認定制度である。本節では、この認定制度の内容について説明する。

3. 1. 認定制度について

本プログラムは表1に示す9つの必修科目と2つの選択必修科目のうち1科目の合計10科目を履修することによって認定証が発行される[17]。

表1: 2021年度のデータサイエンスプログラム認定証発行のための履修条件[17]

学部の必修科目	数学 I、情報科学 I、生物統計学
プログラムの必修科目	数学 II、情報科学 II、情報科学 III、生命科学 IV(基礎情報学)、 生命科学 V (計算機の論理とデータ構造)、 分子生命科学 VI(バイオインフォマティクス)
プログラムの選択必修科目	生命医科学 IX (ゲノム医科学)、応用生命科学 IX(ゲノム進化学) 上記のうち1科目以上取得

科目構成は、数学や統計学のような数理科目、情報科学や計算機科学のような情報系科目、これらを他分野に応用した学際科目からなる。旧カリキュラムにおけるプログラミング基礎(全14回)は情報科学 III(全7回)に変更となった。また、バイオ情報科学(全14回)は生命科学 V(計算機の論理とデータ構造)および分子生命科学 IV(バイオインフォマティクス)のそれぞれ全7回の講義に分かれた。

3. 2. 各科目の概要

本項ではデータサイエンスプログラム認定に必要な科目について各科目の内容について簡単に説明する。データサイエンスプログラムの認定に伴いカリキュラム内容が大幅に変更された情報科学 I、II に関しては次節で説明する。

3. 2. 1. 数学 I、II

数学 I は 1 年前期の必修科目であり、数理・データサイエンスの基盤である線形代数および微積分について学ぶ。数学 II は 1 年後期の選択科目であり、微分方程式など数学 I の内容をさらに発展させた内容について学び、さらにプログラミング言語の 1 つである Python を用いて線形代数を計算することにも取り組んでいる。

3. 2. 2. 生物統計学

本科目は 2 年前期の必修科目であり、研究や学会発表などの際に重要になる統計学的解析について、基礎的な用語や手法の理解を目標としている。講義前半 7 回に統計基本量や確率、母集団と標本といった統計の基本を習得する。後半 6 回では、区間推定や分散分析など推定や検定の基本を学習する。また本科目は、統計検定 2 級相当の内容であるため、本試験の受験にも有用である。

3. 2. 3. 生命科学 IV(基礎情報学)

本科目は、2 年後期の選択科目であり、データサイエンスの倫理や情報倫理の項目と関連が深く、現代情報社会を支える高度な情報テクノロジーの本質と、その背後にある人間的、生命的問題を考える力を養うことを目的としている。現在の AI やロボットの起源となったサイバネティクスとその後期の哲学的転回を軸に、記号学や記号論、構成主義哲学などの人文知も交えながら、情報の意味や価値が前提となる生命現象とテクノロジーとの関係を解説し、情報テクノロジーを無自覚に受容するのではなく、自らの力で批判的に捉えるための知的トレーニングとなるよう授業を構成している。

3. 2. 4. 情報科学 III

本科目は、2 年後期の選択科目であり、プログラミングにおけるアルゴリズムとその処理について、プログラミング言語の 1 つである C 言語を例にその概略を理解し、実装することができるようになることを目標としている。そのために学生は自分の所持するノートパソコンに構築したプログラミング環境を用いて、指定された処理を行うための C 言語のソースコードをグループワークの中で解説、編集し、コンパイルした後に実行する演習を行っている。

3. 2. 5. 生命科学 V (計算機の論理とデータ構造)

本科目は、3 年前期の選択科目であり、2022 年度より開講予定である。前述のように、旧カリキュラムのバイオ情報科学の前半部分の内容が別科目となった。本科目では、コンピュータの基礎である論理回路やデータベース、アルゴリズムについて学び、C 言語による実習も行う。また、IT パ

スポーツ試験や基本情報技術者試験に関するガイダンスを行い、就職活動に役立つ資格取得を奨励している。プログラミングを実験系の卒業研究に用いた例を紹介し、4年の卒業研究や大学院の研究に役立てることを奨励している。

3. 2. 6. 分子生命科学 VI(バイオインフォマティクス)

本科目も、生命科学 V (計算機の論理とデータ構造)と同様に、3年前期の選択科目であり、2022年度より開講予定である。配列解析からは情報理論、進化系統解析からはグラフ理論というように生命科学と情報科学の学際領域であるバイオインフォマティクス(生物情報科学)の概念について学習する。また本科目は、バイオインフォマティクス技術者認定試験の内容と重複部分があり、本試験の対策にも役立つ。

4. 生命科学部の情報科学 I、II

本節では、従来の情報科学 I、II について説明した後、データサイエンスプログラムによる変更した箇所を説明する。

4. 1. 従来の情報科学 I、II の概要

生命科学部の情報教育は学部創立当初より、全学生がノート型 Mac を 1 台所有し、それらの環境設定を自身で行う教育を推進してきた。共有の端末ではなく個人の端末を所有することにより、自主性や主体性を高めると共に、コンピュータをブラックボックス化させない狙いがある。本学が OS に Mac を採用している理由は、設立当時の教員の多くが Mac ユーザであった歴史的背景に加え、MacOS のコンピュータが、情報科学を理解するにあたり必要不可欠である Unix 系 OS の利用が WindowsOS よりも簡便であることもある。

情報科学 I は 1 年生前期の必修科目であり、情報科学 II は 1 年後期の選択科目である。以降、情報科学 I、II で主として取り組んできた「PC の基本操作」、「アプリケーションの取り扱い」、「ネットワークに関する理解」、「Unix 系 OS によるファイル操作」の 4 つの内容について説明する。

4. 1. 1. PC の基本操作

入学時、PC に初めて触れるという学生は多くいる。そのような学生に PC の起動や終了、キーボードによる入力、コピー&ペーストなどのショートカットを指導する。ほとんどの学生が Mac という同一環境を使用することにより、指導内容を具体的に指示できる。そのため、全学年対象の授業であるが、初学者に対してもきめ細やかな指導が可能となる。

4. 1. 2. アプリケーションの取り扱い

Excel や PowerPoint のような Microsoft Office や ChemDraw のような講義及び実習、研究で必須となるアプリケーションについて、その利用法を授業で取り扱っている。Office の学習においては、アプリケーションの使用法と共にロジカルライティングや参考文献の付け方、分かりやすいプレゼンテーション資料とは何か、データ解析の基本などについても合わせて学習する。

4. 1. 3. ネットワークに関する理解

第三次 AI ブームの起因がインターネットの普及、オープンデータの拡充であったことから分かる通り、ネットワークに関する理解は情報科学の基本であるに留まらず、データサイエンスの根幹と繋がっている。また、本学の学内ネットワークは教職員によって創設・維持されてきた歴史的背景[18]から、本内容を教えるにあたり恵まれた環境がある。主に情報科学 I において、インターネットの基本概念や無線ネットワーク、メールなどの仕組みに触れ、本学のネットワークの構成について扱っている。現在では、後述の理由で削減されたが、HTML ファイルを自身で作成し、自身の自己紹介をする Web ページを作成する授業も行なっていた。SNS が普及する前は、学生間の交友を深めるに役立つ面もあった。

4. 1. 4. Unix 系 OS によるファイル操作

一般的な PC の操作は Graphical User Interface (GUI) と言われ、アイコンとマウス、一部キーボードの利用で成立する。一方、Character-based User Interface (CUI) と呼ばれる、人間の入力情報とコンピュータの出力情報の対話によって操作方法がある。この操作方法の習得がファイルシステムの理解は極めて重要である。これは直感的には不便なように感じるが、実際には自由度や操作性が高く、習得すれば GUI の何十倍もの効率で作業を行うことができる。また、プログラミングや高度な分析装置や計算機器の操作に関しては CUI の使用は必須であり、IT 人材の養成のみならず研究者や社会人にも必須のスキルとなっている。本内容は情報科学 I でさわりのみを扱い、高等的な内容は情報科学 II で習得する。

4. 2. データサイエンスプログラムによる変更

数理・データサイエンス・AI 教育プログラムのリテラシーレベルに該当する内容として、人工知能や機械学習の内容不足していたため、情報科学 I では人工知能や情報リテラシー、情報科学 II ではプログラミングによるデータの分析や数値解析、機械学習の実装などの内容を拡充した。授業で利用しているプログラミング言語は Python と呼ばれる言語である。本言語は、実装が比較的容易であり、豊富なモジュールを利用し簡便に機械学習や深層学習が実装できる。プログラミング言語の人気指標である TIOBE インデックスでは、1 位が C 言語、2 位は Python であり、その人気度が伺える。Python には Jupyter のような極めて簡便にプログラミング実装する統合開発環境が存在するが、前述した CUI を習得するため本授業では利用していない。

一方、プログラミングの内容を拡大した結果、既存では重点的に行われた Microsoft Office、特に Excel の内容は大幅に削減された。加えて、前述の HTML ファイルの編集および Web ページの作成などの内容も削除されてしまったが、これらの一部は教職の新科目である「情報機器の活用に関する理論及び方法」で取り扱う予定である。

5. データサイエンスプログラムの課題と今後

2021年度の情報科学 I の講義において、アンケートを実施し、216名の回答を得た(図2)。その結果、生命科学分野においてデータサイエンス教育が必要であると考えている学生は全体の91.2%(197名)に及び、57.4%(124名)がデータサイエンスに興味があると回答した。一方、現時点でデータサイエンスプログラムの履修を検討している学生は全体の42.1%(91名)であった。データサイエンスプログラムの必須科目である情報科学 II の履修者が103名であり、大きく反した結果ではないと考えられ、現段階での興味関心は強いと思われる。

一方、データサイエンスプログラムに興味がある学生があるものの履修を断念した学生が約28%いる。そこで、履修を断念した理由について解析することとした。本内容を掲載できないため、その要点をワードクラウドにしたものを図3に示す。これは出現頻度が高い用語を大きく記載し図示したものである。

「課題」「内容」「理解」は、情報科学 I の授業および課題内容が難しいと感じている意見である。加えて「対面」は情報科学 II が対面授業となるなら履修を検討するというものであった。本プログラムは情報科学だけで構成されているものではないが、情報科学 I でのアンケートという性質上、情報系科目に偏ったと思われる。従来、情報科学 I は学科ごとの対面授業であり、適宜質問対応を行ってきた。また、コンピュータの操作が苦手な学生は質問すること自体が困難であるが、教員やTA側でその意図を汲み取り学習支援に取り組んできた。しかし、2020年度および2021年度はオンライン授業となり、学生が質問すること、教員やTA側で躓いている点を汲み取ることが困難となってしまった。また、3学科同時開催となり、質問対応できる量に限界があった。本年度は昨年度の反省を踏まえ、対面補習などを行なったが、出席者が多かったとは言えず、その効果は限定的であった。また、アントレプレナープログラムとの両立、履修した先輩について知りたい、キャリアプランが見えないなどプログラムの方向性を明らかにすることを求める声も多かった。

データサイエンスプログラムの履修者を増やすためには、選択科目である情報科学 II の履修者を増やす必要がある。上記の結果を踏まえると、以下のことが対策としてあげられる。

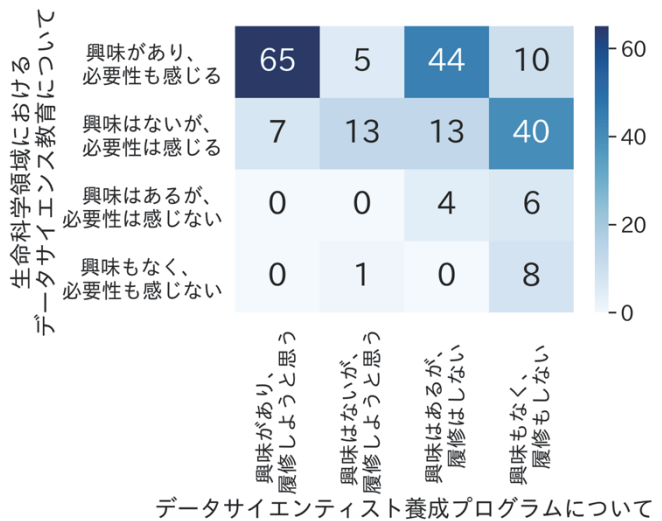


図2: 情報科学 I におけるアンケート結果



図3: 履修断念理由にまつわるワードクラウド

1. 情報科学 I の難易度低下
2. データサイエンスプログラムのキャリアパスおよび履修者に関する広報活動
3. 本プログラムの制度周知と他プログラムとの連携

1.については情報科学 I が次年度から対面授業に戻る見込みであり、既存のフォローアップ体制が取れば改善すると思われる。また、情報科学 I ではプログラミングについて扱っているが、十分な学習時間は確保できず、むしろ学生の混乱を招いている場合もある。そのため、情報科学 I から情報科学 II, III のような選択科目に移行することも検討する必要がある。同様に、情報科学 II ではプログラミングに合わせてデータ解析や機械学習のアルゴリズムに関して学習するが、この時点で学生は本学習に重要な統計学を履修していない。そのため、統計学を学習した2年後期以降に統計的機械学習に関する専門講義を設けることができればより効果的である。このように各科目で学習内容が重複すること、または座学と実習の反復することは学習効率を高めるために重要であると考えられる。2.に関しては2021年度には本学広報課のご尽力によりWebページを掲載いただいた[18, 19]。データサイエンスプログラムを受講している中で最上位学年は2年生のため、まだ就職をした学生などはないが、追跡調査を行いつつ、このような広報活動を継続して行うことが学生のデータサイエンスに関する学習意欲向上につながると思われる。

AI にまつわる世界的潮流より、日本のデジタル人材不足が問題視されている。そのような中で、生命科学部が取り組んでいるデジタル人材養成制度であるデータサイエンスプログラムについて簡単に書きまとめた。本原稿を執筆するにあたり、多くの方々の支援を賜りました。この場を借りて、御礼申し上げます。

参考文献

1. A. M. Turing, *Mind*, **59**(236), 433-460(1950).
2. A. L. Samuel, *IBM Journal of Research and Development*, **3**(3), 210 - 229(1959).
3. F. Rosenblatt, *Psychological Review*, **65**(6), 386-408(1958).
4. A. Krizhevsky, I. Sutskever, G. E. Hinton, *NIPS'12: Proceedings of the 25th International Conference on Neural Information Processing Systems*, **1**, 1097-1105(2012).
5. D. Silver, J. Schrittwieser, K. Simonyan *et al.*, *Nature*, **550**, 354-359 (2017).
6. J. Jumper, R. Evans, A. Pritzel, *et al.*, *Nature* **596**, 583-589 (2021).
7. R. Evans, M. O. Neill, A. Pritzel *et al.*, bioRxiv(2021)
doi: <https://doi.org/10.1101/2021.10.04.463034>
8. 文部科学省, 『バイオ DX』による科学的発見の追究,
URL: https://www.mext.go.jp/content/20210312-mxt_kiso-000013144_5.pdf (2021.12.11 閲覧)
9. みずほ情報総研株式会社, IT 人材需給に関する調査, 2019.3
10. 経済産業省「DX レポート IT システム「2025年の崖」の克服とDXの本格的な展開」2018.9
11. 文部科学省 科学技術・学術政策研究所 科学技術予測・政策基盤調査研究センター, 科学技術指標 2021, 2021.8
12. 佐藤恵子, 白木澤佳子, 高木利久, 藤博幸, 情報管理, **56**(11), 782-789(2013).
13. 統合イノベーション戦略推進会議, AI 戦略 2019 ~人・産業・地域・政府全てに AI~, 2019.6
14. 大島泰郎, 東京薬科大学研究紀要, **1**, 65-67(1998).
15. 森河良太, 西田洋平, 山田寛尚, 宮川毅, 東京薬科大学研究紀要, **20**, 51-58(2017).
16. 井上紫織, *IMES Discussion Paper*, 2020-J-11(2020).
17. 東京薬科大学生命科学部, 2021年度授業計画 2020年度以降入学生用, 2021.
18. 森河良太, 林昌樹, 宮川毅, 土橋朗, 東京薬科大学研究紀要, **1**, 77-84(1998)
19. <https://www.toyaku.ac.jp/lifescience/interview/students/voice-202110194771.html>
20. https://www.toyaku.ac.jp/lifescience/newsttopics/2021/1020_4770.html