

数理・データサイエンス・AI 教育の紹介

筑波大学における データサイエンスリテラシー教育

筑波大学共通科目「情報」推進室

岡 瑞起・佐久間 淳・津川 翔・福地 一斗・平田 祥人



1. はじめに

近年の情報技術の発展に伴い、情報を適切に収集・管理し、その分析結果を有効に役立てる技術があらゆる分野において必須な時代となりました。学問分野では、大量のデータを収集・分析することにより研究を推進させる第4のパラダイムとしてのデータサイエンスの重要性が広く認識されるようになってきています。また産業分野では、データの効果的な集積と活用による価値の創出が、産業競争力の強化に大きな役割を果たすようになりました。さらに急速に発展する人工知能技術は、一部の分野では人間の認識・判断能力を凌駕するレベルに到達しつつあります。このような技術的背景に基づいた社会環境の変化は、大学が輩出すべき人材像にも変革を促しています。

本学では1973年の開学以来今日に至るまで、基礎的な情報リテラシーとコンピュータの利用技術の修得を目的とした教養教育としての科目「情報」を全ての初年次学生の必修科目として開設してきました。「情報」は、社会の情報環境の進展・変化に対応して、その学修内容を継続的に見直しつつ実施してきましたが、データサイエンス教育も情報リテラシーの一つとして位置付けることとし、全国の大学にさきがけて2019年度から全学の初年次生の必修科目として2単位の「データサイエンス」を開講しました。

本学の学生が所属する学部の専門分野は、人文・文化、社会・国際、人間、生命環境、理工、情報、医学、体育、芸術と多岐にわたり、学生が

持つ興味や知識、スキルは学部によって大きく異なります。この点を考慮して「データサイエンス」教育は以下の3点を考慮して設計されています。

① 学生の多様な背景への対応

データサイエンスを活用するさまざまな専門分野の教員から直接ビデオ講義を通じて語りかけてもらい、学問に対する興味や動機が多様に異なる学生が、データサイエンスを学ぶ必要性を理解し、データサイエンスを学ぶ動機を高められるようにしています。その実現に中心的な役割を担うのが筑波大学オープンコースウェア（OCW）で提供されているビデオ講義です。

2022年現在、本学の人文・生命・理工・情報・医学・体育・芸術などさまざまな分野におけるデータサイエンスに関わる研究者が、先端的なデータサイエンスの事例に触れ、社会におけるデータの具体的な活用に関する知識を提供する演習付きビデオ講義が公開されています。

② データサイエンスを「自分ごと」に

現実世界のデータや自分たち自身を対象として収集されたデータを扱い、そのデータ処理結果に触れる機会を持つことで、データサイエンスの結果がもたらす効果を実感し、自分ごととして捉えることができるようにしています。具体的には、履修者全員にSDGsやキャリア観など自分と関連の深い項目についてのアンケート調査を行い、そのデータを用いたデータ分析演習を授業に取り入れています。

③ データに基づくデータサイエンス教育の改善

本学では学生のデータサイエンスに対する意欲や知識を主観的および客観的に調査する授業評価アンケートおよび教育効果測定を実施しています。これらの評価結果に基づき、さまざまな分野を専攻する学生のデータサイエンスの学習動機や主観的・客観的理解度の調査やデータサイエンスの客観的理解度の増加に貢献する要因の絞り込み等を通じて、授業内容や演習内容、実施方式の適正化に取り組んでいます。

本稿ではこれらの観点から、OCW、学生データ演習、授業評価アンケート、教育効果測定等の取組みとその結果について報告します。

2. 筑波大学オープンコースウェア

本学では2019年度より全ての一年次学生（約2,100名）を対象として、データサイエンスを必修科目としています。データに基づく判断や意思決定を適切に行う能力は、理工系のみならず全ての学生にとっていまや必須のスキルと言えます。その一方でこのような多様な専門領域を志す多くの学生を対象として、一律にデータサイエンスを必修科目として課すことは、日本のみならず世界的にみても稀な、極めてチャレンジングな試みです。

データサイエンスは統計学やデータ工学など理工系の学問を背景に持つことから、理工系以外の学生にとって取り組みにくい印象を与えてしまう可能性があります。このため、データサイエンスの教材作成にあたっては、理工系以外の学生にとっても平易であることを心がけるのはもちろんのこと、理工系以外の学生がデータサイエンスを学ぶ動機を講義で与えることが重要です。

そこで、本学では多くの学生にスムーズにデータサイエンスの学習に取り組んでもらうための教材としてオンライン動画の作成を進めています。

オンライン動画教材では、本学においてさまざまな専門領域においてデータを活用した研究を実施する教員が自身の研究活動を初学者向けに紹介しています。これをデータサイエンスの第一回目の講義において教材とすることで、今後自分が専門としようとしている分野ではデータの活用がどのような成果につながっているかを学生に理解し

てもらい、データサイエンスを学ぶ動機づけを行うことを目的としています。

各動画は25分程度で、研究の背景にあるデータの収集・管理・解析がどのように行われているかに触れ、データサイエンスを学ぶモチベーションを高めてもらうような構成になっています。これまでに作成した動画は本大学オープンコースウェアサイト (<https://ocw.tsukuba.ac.jp/data-science>) にて公開しています（図1）。

これまでに作成した動画タイトルは次のとおりです。高度なデータの管理と活用に関して「ビッグデータとIoT/CPS」と「人工知能と機械学習」の2動画を公開しています。また、「仮説検定入門」を発展版として公開しています。そして次の20個の動画を導入動画として公開しています。

- ビジネスとデータサイエンス
- 都市計画とデータサイエンス
- 生命科学とデータサイエンス
- データサイエンスと社会科学
- 臨床医学・社会医学とデータサイエンス
- 生物科学研究におけるデータサイエンス
- 植物標本データが伝える生物の進化と危機
- デジタル・ヒューマニティーズ
- 言語教育にデータはどのように活用されるのか
- 「心」のデータ：行動デザインの視点から
- ヒューマンインタラクション
- 障害科学におけるデータに基づく意思決定
- 芸術とデータサイエンス
- サッカーの上達にデータを生かす
- 現代サッカーボールの空力特性
- SNS活用の分析におけるデータサイエンス



図1 筑波大学オープンコースウェアサイト

の活用

- AIと制御の融合によるロボットの高速汎用物体操作
- 人工知能における 倫理的, 法的, 社会的問題
- データ駆動型社会における津波高即時予測
- セシウム137全球データベースおよび環境放射能データの検索と公開サイト

これらの動画教材はクリエイティブ・コモンズライセンス「表示-非営利-継承 2.1 日本」の下、どなたでも利用できる教材として提供しています。

3. 学生アンケートデータを用いた演習

本学DSL（データサイエンス・リテラシープログラム）では、受講者がデータサイエンスを主体的に学ぶ動機を持たせるための取組みとして、データサイエンスの受講者に対して実施したアンケートの結果を、受講者自らが解析する演習を実施しています。データサイエンスを学ぶ意義を理解するためには、演習のために作られたデータでExcelの操作や統計解析の手法を学ぶだけではなく、受講者にとって身近な生のデータを解析することが重要であると考え、2020年度からこのような演習を取り入れました。本章では、学生アンケートデータを用いた演習の概要について紹介します。

本演習を実施するにあたって、データサイエンス全10週のうち2週目に、データサイエンスの全受講者に、いくつかのアンケートへの回答を依頼します。学生はアンケート回答に先立って、人を対象とする研究で個人情報を収集する際の一般的な手続きにしたがい、このアンケートの目的、

個人情報保護のための対策、などの説明を受けます。

この説明を受け、アンケート結果の収集に同意した場合のみ、学生は同意書を提出し、manabaと呼ばれるLearning Management System（LMS）上でいくつかの質問に回答します（図2）。なお、本アンケートの収集は、筑波大学システム情報系研究倫理委員会の承認の下で実施しています。

アンケートの質問項目は、Sustainable Development Goals（SDGs）に対する認知度、コンピュータおよびメディアの利用状況、将来のキャリアに対する意識、を問う質問で構成されます。質問項目は、総務省や内閣府で実際に実施された調査の質問項目から抜粋しており、それらの調査結果と、本演習で取得したアンケート結果を比較することも可能です。また、回答は、名義尺度、順序尺度、間隔尺度、比例尺度を含むように設計されており、データサイエンスで学ぶさまざまな解析手法や可視化手法をこのアンケートデータに適用することが可能になっています。

2週目に収集したアンケート結果は、DSL担当教員の元で匿名化処理され、第6週以降の演習で利用するために、学生および授業担当教員に配布されます。本アンケートデータを用いた演習課題の例としては、インターネット利用時間の分布をヒストグラムでプロットするものや、書籍を読む時間とスマートフォンの利用時間との相関を調査するものなどがあります。また、最後の演習では、学生が自由に目的を設定して、本データを解析します。図3はアンケートデータを実際に解析した例で、1日あたりのインターネット利用時間の分布を示しています。このようなデータから、分布のピークの位置がどこにあるのか、外れ値が存在するかどうかなど、データの読み取り方法を学びます。

受講者に対して実施しているアンケート結果等

図2 LMS上でのアンケートへの回答画面

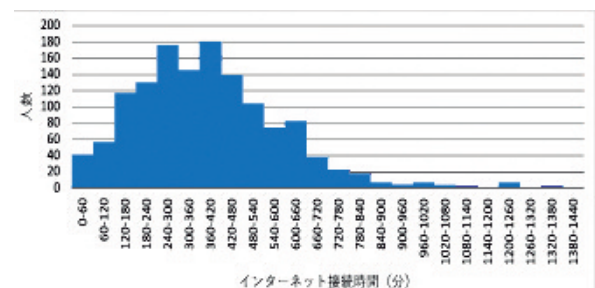


図3 アンケートデータの解析例

からは、本演習のポジティブな効果が示唆されており、このような演習は、データサイエンスを主体的に学ぶ動機付けに有効であろうという当初の予想が裏付けられたと考えています。

4. 学生授業評価アンケート

本学ではFD活動の一環として、全学規模で学生が授業の評価を行うアンケートを実施しています。このアンケートは本学の全ての授業において実施され、FD活動に役立てています。授業評価アンケートは、学生の授業に対する主観的な評価に関する全授業で共通の4つの設問と、各講義独自に設定した設問で構成されています。データサイエンスの授業では15の独自の設問を含む、19設問のアンケートを実施しています。表1に、実際の設問一覧を示します。独自のアンケート項目は、授業の理解度や難易度、授業のわかりやすさなどの基本的な質問に加えて、データサイエンスに対する主観的な習熟度に関する設問を設けています。アンケートは全10週の授業において、第9週の最後にオンラインのシステムを通じて回答します。表2に、2019年度のデータサイエンスの授業開始から2021年度まで実施のアンケートの回答者数を示します。2020年に回答者数が前年と比べて減少していますが、これは、COVID-19の影響でオンライン授業へ移行した影響です。

解析はアンケート項目に加えて、学生の所属学類を加えて学類間の差異の調査も行います。解析の結果を使ったFD活動の例として、モチベーションを高めるためのOCWコンテンツの拡充があげられます。2019年度のアンケートデータの解析の一つとして、興味関心に関する項目が低い学類の割り出しを行いました。このような学類では、第2章で述べたOCWにおいて、その学類に対応した動画コンテンツが十分ではないことがわかりました。実際に、2020年度の授業のために、これらの学類の動画コンテンツの拡充を行いました。2020年のアンケート結果で、実際にこの動画コンテンツ拡充の効果として、動画コンテンツを用意した学類の興味関心に関する項目の結果が2019年度と比べて向上していることが観測できています。このように、アンケート解析結果をもとに授業教材の更新を行い、それをその年のアン

ケート結果で効果の確認を行うことが可能になっています。

表1 授業評価アンケート質問項目

授業の準備は十分にされていたと思いますか。
教員の説明や授業の進め方は適切でしたか。
授業を通じて、この科目に関連する分野への興味や関心が高まりましたか。
総合的に判断して、この授業を受講してよかったと思いますか。
授業の内容をどの程度理解しましたか。
授業で習ったことは、今後、どの程度役に立つと思いますか。
授業の難易度はどの程度でしたか。
理解すべきポイントを分かりやすく説明していたと思いますか。
適切な教材（スライドなど）を準備していたと思いますか。
スライドや板書は見易さに十分配慮していたと思いますか。
声の聞き取りやすさに十分配慮していたと思いますか。
講義開始前、あなたのデータ収集・管理の習熟度は次のどの程度だったと思いますか。
また、講義が終わって現在のあなたのデータ収集・管理の習熟度は次のどの程度だと思いますか。
講義開始前、あなたのデータ可視化・分析の習熟度は次のどのレベルだったと思いますか。
講義終了後、現在のあなたのデータ可視化・分析の習熟度は次のどのレベルだと思いますか。
中学・高校の授業におけるデータ解析ソフトの学習内容と今回の講義内容はどの程度重複していましたか。
データサイエンスについてより詳しく知りたいと思った項目を上げてください。
データサイエンスの授業において、不要だと思った項目を上げてください。

表2 各年度の参加者数

年度	2019年	2020年	2021年
受講者	2175	2152	2219
授業評価アンケート回答者	1712	1391	1328
効果測定研究参加者	1757	1493	1537
全設問に回答した研究参加者	1382 (64%)	1028 (48%)	1033 (47%)

5. 効果測定

2019年度のデータサイエンスの授業開始と同時に、継続して効果測定を行っています。2019年度の授業開始当初は、全学的にデータサイエンス教育を行っている大学は少なかったため、データサイエンスの授業の効果を客観的に測り、今後の授業改善に活かしていくことが重要だと考えたためです。

本学でのデータサイエンスの効果測定は、第1週目と第9週目に授業時間の中で時間をとって行っています。第1週目では、説明書を読み、同意書を作成後、効果測定に関する設問に答えてもらっています。効果測定に関する設問は、動機に関する設問（授業が楽しみか、今後の自身の学習・研究に活かそうか）、授業で習う35個の専門用語のうち受講開始前に説明できると思う言葉の自己申告、それと学生の属性に関する設問（所属学類、受講しているクラス、高校在籍時の数学の履修状況等）です。

第9週目の時には、授業評価アンケート、動機に関する質問、専門用語が主観的に説明できると思うかどうかと客観テストを行っています。授業評価アンケートは、通常、本学の授業で行っている授業評価アンケートです。説明は、前節の通りです。動機に関する質問では、授業が楽しみだったか、今後の自身の学習・研究に活かそうかを聞いています。専門用語に関する質問では、1週目に聞いた35個の専門用語のうち曜日時限別にランダムに割り当てた7個の専門用語が、説明できると思うか、主観的に答えてもらっています。客観テストに関しては、問題のプール41個の中から、ランダムに選んだ6問を、学生の受講している曜日時限別に割り当てて、実施しています。

効果測定の全設問は、全て本学で使用している学習支援システムであるmanaba上で、回答するように設定しています。

2019年度の授業開始前に、本学システム情報系の倫理審査委員会において、倫理審査を経た上で、効果測定を実施しています。今までに、4回倫理審査で修正を行い現在に至っています。

2019年度は、対面型授業の中で、効果測定を行いました。2020年度、2021年度は、オンデマンド型の授業の中で、効果測定を行いました。

2022年度も、現在のところ対面型授業の中で、効果測定を行っています。

2019年度から2021年度までの参加者の数を表2にまとめました。受講生のうち7割以上が効果測定に参加しています。また、最終的に全設問に答えた研究参加者の割合は、受講生のおよそ半分でした。

また、客観テストに関して、年度ごとの変化を図4に示します。これは、研究参加者の6点満点中の得点分布です。点数が高い（左側に寄っている）ほど良い結果を示しています。データサイエンスの授業の内容が、年を経るにしたがって、改善されて来ていることがわかります。Wilconxonのrank sum testを用いると、2019年度（平均点：3.42）よりも2020年度（平均点：3.79）、2020年度よりも2021年度（平均点：3.96）の客観テストの点数が、有意に良くなっていました。授業評価アンケートの結果も含めた詳細な解析は、別の機会に報告する予定です。

本学のデータサイエンスの授業で行っている効果測定の結果を紹介しました。効果測定の手順を述べた後で、研究参加者数の統計と、客観テストスコアの経年変化を示し、データサイエンスの授業が改善して来ている様子を示しました。

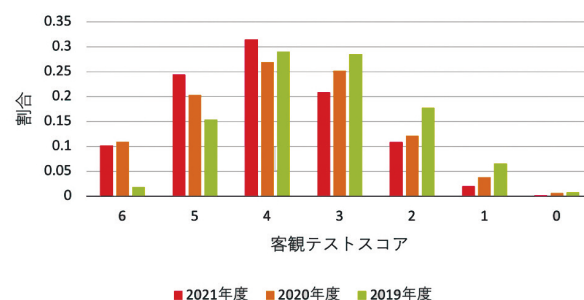


図4 客観テストスコアの経年変化

6. おわりに

本稿では本学の1年次必修科目である「データサイエンス」について、その授業実施上の工夫について述べました。今後は授業評価アンケートや教育効果測定の結果分析を進めつつ、応用基礎レベル、エキスパートレベルにおけるデータサイエンス教育との連携を深め、さまざまな到達レベルを見据えたリテラシー教育のありかたと実施方式の改善を継続的に行う予定です。