

オンデマンドとオンラインを併用した 多人数データサイエンス演習教育の実践

原田 耕治^{1),2)}, 後藤 仁志^{1),2)}

1) 豊橋技術科学大学 IT 活用教育センター

2) 豊橋技術科学大学工学部情報・知能工学系

harada@cite.tut.ac.jp

Practice of large-scale exercise education in data science using both on-demand and online

Kouji Harada^{1),2)}, Hitoshi Goto^{1),2)}

1) Center for IT-Based Education, Toyohashi University of Technology.

2) Department of Computer Science and Engineering, Toyohashi University of Technology.

概要

文部科学省が創設した数理・データサイエンス・AI 教育プログラム認定制度には、リテラシーレベルと応用基礎レベルの 2 つの区分があり、前者は基礎的な能力を、後者は実践的な能力を育成することを目指している。本学の「GIKADAI 数理データサイエンス AI 教育プログラム」は 2021 年 8 月に両レベルで認定を受けた。この制度は全大学・高専の卒業生が数理・データサイエンス・AI の能力を習得することを目指しており、認定を受けるには全学/学部レベルの大規模な教育プログラムの実施が必須となる。しかしながら、リテラシーレベルと異なり、応用基礎レベルは実践的能力の育成が求められているため演習型授業が適当であるが、多人数教育には課題がある。我々は、Jupyter Notebook 形式の演習教材と解説ビデオを組み合わせ、オンデマンドによる自学自習とオンライン質疑を通じて全学規模の演習授業を実現している。この論文では、その実践内容を報告する。

1 はじめに

2020 年度に創設された文部科学省 数理・データサイエンス・AI 教育プログラム認定制度 (MDASH)には、育成する能力のレベルによって「リテラシーレベル」と「応用基礎レベル」の二つの区分がある[1]。リテラシーレベルでは、数理・データサイエンス・AI の基礎的な能力を育成することを目指している一方、応用基礎レベルでは数理・データサイエンス・AI を活用して課題を解決するための実践的な能力を育成することを目指している[2]。2023 年 9 月の時点で、リテラシーレベルには 382 プログラム (内 25 がプラス)、一方、応用基礎レベルには 147 プログラム (内 14 がプラス) が認定された[3]。本学の「GIKADAI 数理データサイエンス AI 教育プログラム」もリテラシーレベルと応用基礎レベルで構成されており、それぞれ 2021 年 8 月に MDASH 認定を受けている[4][5]。

特にリテラシーレベルはプラスとして選定された。本制度は、文理を問わず、全ての大学・高専の卒業生 (50 万人/年) がリテラシーレベルを習得し、そのうちの 25 万人は応用基礎レベルを習得することを目指している。このことから MDASH の認定を目指す大学・高専は全学規模または学部規模の多人数に対する数理・データサイエンス・AI 教育を実践する仕組みが必要とされている。リテラシーレベルの教育プログラムは、数理・データサイエンス・AI に関わる基礎的知識の理解を目的としているのでオンデマンド型授業による実現が可能である。一方、応用基礎レベルは実践的な能力を育成することを目指しているため、プログラミング教育が可能な演習型授業が理想的である。ところが演習型授業は、計算機環境や人的リソース、教室のスペースの観点から多人数教育を実施するには適していない。そこで我々は、数理・データサイエンス・AI を実践的に学べる Jupyter Notebook 形式の演習教材とその解説ビデオを開

発し、オンデマンドによる自学自習とオンラインによる質疑を組み合わせることで全学規模の演習授業を実施している。本論ではその実践内容を報告する。

2 教材の特徴

本学の応用基礎レベルは、「データサイエンス演習基礎」と「データサイエンス演習応用」の2科目で構成されている。これらの演習では、本学と株式会社キカガク[6]の共同で開発した電子教材とその解説ビデオを用いている。

表1 基礎編 (TK Basic シリーズ)

1. イントロダクション (人工知能とは)
2. 機械学習の数学 1 (微分)
3. 機械学習の数学 2 (線形代数)
4. 機械学習の数学 3 (統計)
5. 機械学習の数学 4 (単回帰分析)
6. 機械学習の数学 5 (重回帰分析)
7. Python の基礎 1 (データ構造と制御構文)
8. Python の基礎 2 (関数)
9. Python の基礎 3 (クラス)
10. 数値計算
11. データ処理と可視化
12. 機械学習の実装 1 (教師あり学習: 回帰)
13. 機械学習の実装 2 (教師あり学習: 分類)
14. 機械学習の実装 3 (ハイパーパラメータの調整)
15. 機械学習の実装 4 (教師無し学習)

全教材は大きく「基礎編」(TKBasic)と「応用編」(TKAdvance)の二つのシリーズからなり(表1,2)、データサイエンス演習基礎では基礎編、演習応用では応用編を使用している。基礎編は15教材からなり、主に「機械学習の数学」、「Pythonの基礎」、「機械学習の実装」の内容を扱っている。一方、応用編は10教材からなり、「ニューラルネットワークの数学」、「画像処理とディープラーニング」、「系列モデリングとディープラーニング」、「自然言語処理とディープラーニング」の内容を扱う。

表2 応用編 (TK Advance シリーズ)

1. ニューラルネットワークの数学 1 (順伝播)
2. ニューラルネットワークの数学 2 (逆伝播)
3. ニューラルネットワークの実装 1 (分類)
4. ニューラルネットワークの実装 2 (回帰)
5. 画像処理とディープラーニング
6. 画像分類の実装
7. 系列モデリングとディープラーニング
8. 時系列解析
9. 自然言語処理とディープラーニング
10. 機械翻訳の実装

全ての教材は、Jupyter Notebook形式で作られている。Jupyter Notebookとは、PythonなどをWebブラウザ上で記述・実行できる統合開発環境です。各教材の前半部は、教材で学ぶ内容の解説となっており、入力済みのPythonコードを実行しながら体験的に学べるようになっている。一方後半部では、教材内容の理解を定着させるための練習問題が用意されている。演習講義では、この練習問題を受講生に課題として課している。また、各教材にはそれぞれ解説ビデオが補助教材として用意されている。上述のように、本教材は座学的な要素と演習的な要素を含んでいることから、自学自習に適した演習教材となっていると言える。また、本教材は自学自習型の演習教材として以下の優れた点がある。

- Google Colab[7]を利用することでGPUを無料で使えるので、手持ちの計算資源が十分でない場合でもディープラーニングのような重い計算を実行できる
- 「コードセル」を利用することで、学習途中で気になったこと、不明なことを実験・確認しながら学習を進められるので、学習効果が高い。
- 学習の途中で気づいたことなどを「テキストセル」にメモしながら学習できる。

- ・ 自前の PC に Jupyter Notebook をインストールすることで、オフラインでも学習が可能である。

3 演習授業の実践

データサイエンス演習基礎（3 年次・後期）と演習応用（4 年次・前期）はともに半期 15 回の演習授業である。演習基礎は、基礎編が 15 教材あるので、1 教材/1 回のペースで進み、一方、演習応用は応用編の 10 教材に新たに 2 教材を追加し、12 教材を 15 回で学ぶ。どちらの授業も授業時間は 18 時から 19 時半に設けられており、オンラインで質問を受け付ける時間としている。また各回は、講義日を中心に 2 週間で完結するように設計しており、具体的には以下の流れに従う。

1. 講義の 1 週間前に Moodle で教材の配信を開始し（図 1）、受講生は教材をダウンロードする
2. 受講生は各自で講義日までに教材を自学する。
3. 質問のある学生は、講義日に Google Meet で質問をする。
4. 次週の講義日の深夜 0 時までに課題を提出する。



図 1. データサイエンス演習応用の Moodle コース

本学は工学部だけの単科大学であり、5 つの専門課程（「機械工学」、「電気・電子情報工学」、「情報・知能工学」、「応用化学・生命工学」、「建築・都

市システム学」）からなる。データサイエンス演習基礎と演習応用はどちらも全課程対応科目であるため、受講者数が多く、演習基礎は約 250 名、演習応用は約 100 名ほどになる。また両演習科目は情報知能工学の受講者が最も多く、演習基礎では 40%、演習応用では 42%である。また受講者数は課程による大きな偏りがある（図 2）。各科目は教員 1 名と TA1 名で担当している。全受講生を 1 教室に集めて実施する通常の演習授業とした場合、この人的リソースでは対応することはできないが、オンデマンドでの自学自習とオンラインでの質問対応という授業形態をとることで可能となる。

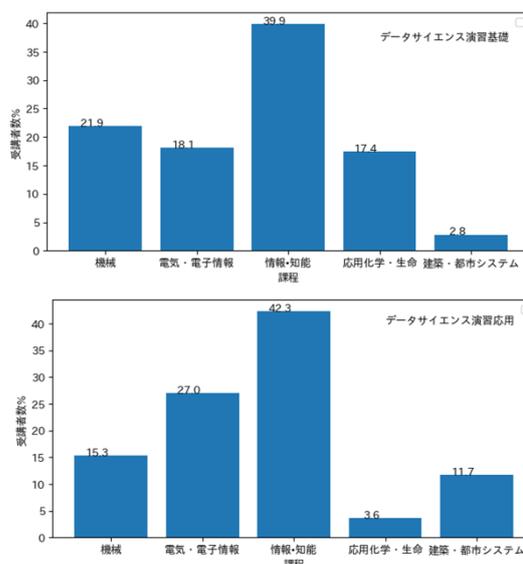


図 2. 2022 年度データサイエンス演習基礎と演習応用の受講者分布

4 アンケート結果と成績

データサイエンス演習基礎と演習応用では独自に授業アンケートを取っている。2022 年度は演習基礎では 146 名（回答率 61.6%）から回答を得た。一方、演習応用では 46 名（回答率 45.5%）から回答を得た。

4.1 理解度

理解度に関しては、「よく理解できた」「どちらかという理解できた」「どちらかという理解できていない」「理解できていない」の 4 段階で評価した。図 3 より「理解できた/どちらかという理解できた」の回答を合わせた割合は、データサイエンス演習基礎は 87.7%であり、演習応用では 97.6%となった。この結果から、どちらの科目も理解度はかなり高いと考えられる。

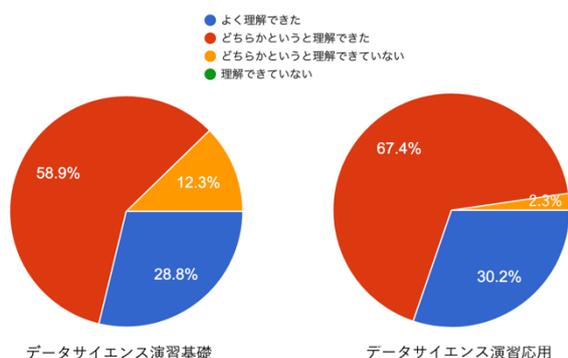


図 3. 2022 年度データサイエンス演習基礎と演習応用の理解度

両科目とも理解度はかなり高いが、これは自己評価であるので次に実際の成績の分布をしてみる(図 4)。

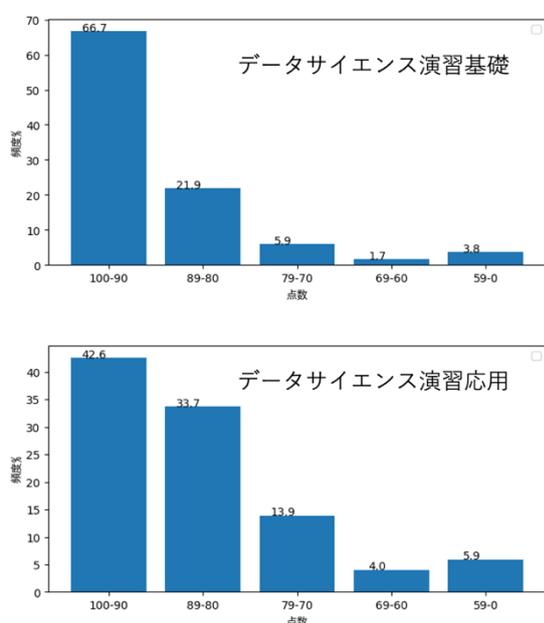


図 4. 2022 年度データサイエンス演習基礎と演習応用の成績分布

データサイエンス演習基礎は S 評価 (100-90 点) の割合が最も高く 66.7% となり、S と A 評価 (89-80 点) を合わせると 88.6% となる。一方、データサイエンス演習応用では演習基礎の成績と比較すると S 評価が減り 42.6% となるが、逆に A 評価は 33.7% と上がっているため、S と A 評価を合わせると 76.3% となる。この値は、演習基礎の 88.6% よりも低いですが全体として成績は良い。

4.1 質問状況

授業時間に行われる質問に関しては、課題提出締切日を授業日に設定することで、締切直前に課題について質問できるように工夫をしている。し

かしながら、質問件数は平均 1 回未満/コマであり、少ない傾向が見られる(図 5)。その要因を知るため、アンケートで質問のしやすさについて訊いた(図 6)。データサイエンス演習基礎のアンケートでは、「質問したいことはなかった」が約 92%、一方、演習応用のアンケートでは 84% であった。このことから、質問が少なかったのは、質問したいことがなかったことが要因であると考えられる。

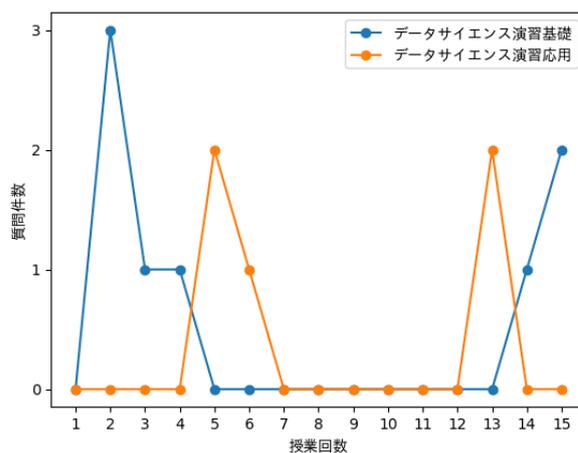


図 5. 2022 データサイエンス演習基礎と演習応用の授業時間中の質問件数

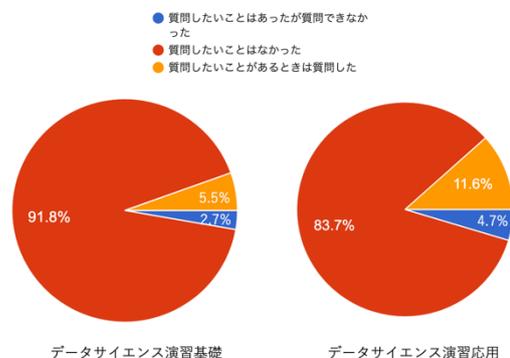


図 6. 質問のしやすさについて

4.1 満足度

オンデマンドによる事前学習とオンラインによる質問対応という形式での演習を実施しているが、結果として質問件数は少なく、直接、受講生とコミュニケーションを取る機会は限られている。このことにより、学生の授業に対する満足度に影響があるのではないかと考え、データサイエンス演習基礎のアンケートの中で「満足度」に関して質問した。結果は「大変満足/どちらかという満足」を合わせると、約 72% となり、満足度は高かった(図 7)。

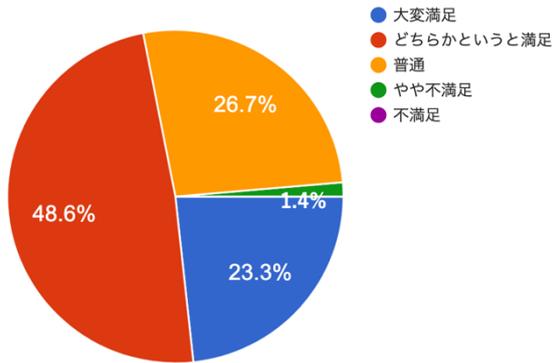


図 7. データサイエンス演習基礎の満足度

また、「オンデマンドで実施したことでよかった事はありましたか」という質問に自由記述で回答してもらったところ、総じてオンデマンドに肯定的な意見がしめた。一例としては、

- ・ 自分のペースで自由に進めることができた
- ・ 授業時間外でも空いた時間に取り組むことができる
- ・ 遅い時間に授業があったが、オンデマンドであったため、バイト後など、夜の落ち着いた時間に学習ができ、眠たいと感じる時間や余裕のない時間に受ける授業よりも内容をよく理解出来た。

また、回答全体のワードクラウド（図 8）からも「好き」「時間」「自分」「ペース」などから好きな時間や自分のペースといった肯定的な単語が多く含まれていることがわかる。



図 8. 「オンデマンドで実施したことでよかった事はありましたか」という質問に対する回答のワードクラウド

5 考察

我々は先のアンケート結果から、現状では運用面で大きな問題はないと考えているが、質問件数が少ない点が気になる。ここでは質問件数が少ない理由について考察する。

質問件数が少ない一つ目の理由としては、当該演習科目の主な受講者が、データサイエンスを専門的に学ぶ「情報・知能工学」と「電気・電子情報工学」の学生であり、彼らにとって教材内容が比較的「易しい」ためであると考えられる。今後、データサイエンスを専門としない、「機械工学」、「建築・都市システム学」、「応用化学・生命工学」の受講生が増えた場合、質問件数が大幅に増えると予想され、オンライン以外の質問対応が必要になると考えられる。その場合には、時間外の質問受付用に slack などのコミュニケーションツールを活用する方法がある[8]。しかしながら大規模授業では、Jin らが指摘するように教員が対応することは現実的に困難であろう[9]。教員に負担をかけない形で質問対応を実現するには、質問対応チャットボットが有効かもしれない[10,11]。

質問件数が少ない二つ目の理由としては、本教材はコードが入力済みのため、理解が浅い学習者でもコードを実行し、結果を確認することで、学習を先に進めることができってしまう点が挙げられる。これは、いろいろな理解レベルの学習者に対応できるという点で本教材の利点であるが、一方、深い理解を得るには注意深く学習を進めていく必要があるということでもある。本教材は社会人のリスクリング講座[12]でも活用しているが、その場では本教材を深く学習していれば当然疑問に思うことが質問としていくつも出てくる。本来であれば、同様の質問が授業の中で出てきてもおかしくないが、そのような質問が出てくることはほとんどない。このことから授業では受講生の理解が浅い可能性があり、深い理解に到達しているのか検証する必要がある。

6 今後の課題

実践的なデータ分析技術を習得してもらうため、プログラムコードの作成を課題として課しているが、採点にかなりの人的リソースを必要とする。我々は、IT 活用教育センターで 4、5 名の RA を雇用することで対応している。採点の負担を減らすために、課題を穴埋めにするなどして LMS の

自動採点を活用できるようにする方法が考えられるが、課題が簡単になり、教育効果が下がる懸念がある。教育効果を維持しつつ、採点の負荷をいかに減らすかが今後の課題である。

参考文献

- [1] 文部科学省高等教育局専門教育課, 「数理・データサイエンス・AI教育認定制度」, https://www.mext.go.jp/a_menu/koutou/suuri_datascience_ai/00001.htm (2023年6月1日確認)
- [2] 数理・データサイエンス・AI教育プログラム認定制度実施要綱, https://www.mext.go.jp/content/20210315-mxt_senmon01-000012801_3.pdf (2023年9月25日確認)
- [3] 「数理・データサイエンス・AI教育プログラム認定制度」の概要/申請に当たってのポイント, https://www.mext.go.jp/content/20230920-mxt_senmon01-000012801_1.pdf (2023年9月25日確認)
- [4] 豊橋技術科学大学IT活用教育センター, 「GIKADAI数理データサイエンスAI教育プログラム」, <https://cite.tut.ac.jp/program-series/mdash> (2023年9月25日確認)
- [5] 後藤仁志, ものづくり技術者のための『GIKADAI数理データサイエンスAI教育プログラム』, 大学教育と情報, (1), 47-52, 2023-06.
- [6] 株式会社キカガク, <https://www.kikagaku.ai/> (2023年9月25日確認)
- [7] Google Colaboratory, <https://colab.research.google.com/> (2023年9月25日確認)
- [8] 名古屋大学応用基礎レベル申請書, <https://www.mds.nagoya-u.ac.jp/mda-progra> (2023年9月25日確認)
- [9] Sung-Hee Jin & Soobong Shin, The Effect of Teacher Feedback to Students' Question-asking in

Large-sized Engineering Classes: A Perspective of Instructional Effectiveness and Efficiency, The Asia-Pacific Education Researcher 21:3, 497-506, 2012.

- [10] Eric Hsiao-Kuang Wu, et.al., Advantages and Constraints of a Hybrid Model K-12 E-Learning Assistant Chatbot, IEEE Access, 8, 77788-77801, 2020.
- [11] Kaur Anjuli et.al. Qualitative exploration of digital chatbot use in medical education: A pilot study, Digital Health, 7, 1-11, 2021.
- [12] 先端データサイエンス実践コース, <https://www.sharen.tut.ac.jp/program/2022/2022-s11.html> (2023年9月25日確認)