

データ活用プログラミング教育の取り組み

佐々木 正人, 石黒 克也, 佐々 浩司

高知大学学術情報基盤図書館

sasaki@kochi-u.ac.jp, ishiguro@kochi-u.ac.jp, sassa@kochi-u.ac.jp

Efforts to Data Literacy Education in synchronous-type online classes

Masato Sasaki, Katsuya Ishiguro, Koji Sassa

Library and Information Technology, Kochi University

概要

2021年度から、共通教育科目として「データ活用のためのプログラミング入門」を開講している。この科目は、すべての学部生（理系も文系も）を対象とし、Anaconda + Pythonによる演習を中心とした同期型オンライン形式の授業となっている。ハイフレックス形式での実習やオンラインサポート等、対面形式で授業ができない中でのプログラミング実習によるデータリテラシー教育や実習サポートについての取り組みについて報告する。

1. はじめに

本学では、これまでで新入生全員を対象とした「情報セキュリティ講習会」や「情報処理」授業を実施してきた^[1]。2020年からは、コロナ禍における新入生対象情報教育にも取り組んできた^[2]。

2021年度からは、「データ活用のためのプログラミング入門」（共通教育科目）を、データサイエンスリテラシーレベル（基礎）として、1学期および2学期（同じ内容）に開講している。すべての学部生（理系も文系も含む）を対象とし、基本的なパソコン操作ができれば、プログラミングの経験については受講条件としていない（図1参照）。

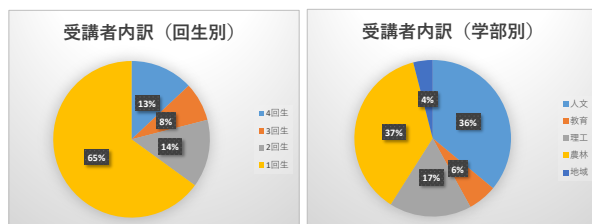


図1 本授業の受講生内訳（回生別，学部別）

Pythonプログラミングによるデータの整理・処理、グラフ等で可視化するなどの演習が中心であるため、直接指導できる対面での授業が望ましいが、新型コロナの影響により同期型オンライン形式での実施が余儀なくされた。

本稿では、授業の目的・目標、具体的な内容等について紹介し、プログラミングに不慣れな学生、

特に入学直後でパソコン操作が不慣れな新入生のサポートなどについて報告する。

2. 授業概要

2.1 目的・目標

本授業の目的は、(1) データを扱う基本的な知識とプログラミング等のスキルを学ぶこと、(2) 身の回りのデータを読み取り、分析に繋げる ことである。Pythonプログラミングを習得することが主目的ではなく、「データを読み、使い、分析し、論じる能力」（データリテラシー）を身につけることを目指しており、その目標は次の通りである。

[I] 「データを読む、説明する、扱う」といった数理・データサイエンス・AIの基本的な活用法を学ぶ。

- ①データの基礎的な事柄と概念を理解している
- ②目的に応じて適切なデータ表現ができる
- ③データ利活用のプロセスを理解し実践できる

[II] Pythonの基本的な文法・規則を習得し、社会での実例を題材としてプログラミングによるデータ活用法を学ぶ。

- ①Pythonの基本的な文法・規則を理解している
- ②Pythonの基本的なプログラミングができる
- ③プログラムの処理の流れを理解できる

2.2 受講生に期待すること

「デジタル社会/Society5.0」はどのような社会になるのか、また、どうなるべきかを常に考えながら学習し、日々の知的活動において、DS や AI がどのように活用できるのかを考え、各自の目的達成までのシナリオを考えることこそが重要であることを理解することを期待している。また、そのシナリオをコンピュータに実行させるのがプログラムであり、目的達成までには試行錯誤が必要（やってみなければ分からない）であること、つまり「知識獲得型学習から自己体系型学習に切り替える（パターンマッチングから論理的思考へ）」を、実習を通じて習得することを期待している。

2.3 実習環境

本授業では、Anaconda + Jupyter Notebook を使用している。インストールが面倒である、CUI (エディタ+Python コマンド) に比べ PC の負荷が大きい等のデメリットもあるが、コメント、プログラム、実行結果 (作図を含む) が 1 つのウインドウ内で完結できるメリットを優先させた。JupyterHub や Google Colaboratory も検討したが、自宅・下宿のネット環境が必ずしも十分でなく、またサーバの準備が困難であったため、この環境を採用した。

なお、本授業での演習では、以下の Python ライブラリを使用している。

- pandas
- numpy
- scipy
- matplotlib
- seaborn
- folium
- scikit-learn

2.4 主な内容

15 回の授業内容は、以下の通りである。Python 入門 1~3 では、以後の処理で必要となる最低限の機能を解説している。

- 第 01 回 オリエンテーション
 - 授業概要・計画の説明
 - 社会で起きている変化とデータ・AI 利活用
 - 社会で活用されているデータ
 - データ・AI 利活用のための技術、動向
- 第 02 回 演習環境の準備、動作確認
 - 統合開発環境 (IDE) Jupyter Notebook

- Python と Jupyter Notebook のインストール
- テキストエディタのインストール
- データリテラシーは現代人の基礎知識
- データの種類
- CSV ファイルとは
- 第 03 回 データを読む
 - データを読む、データの種類、データの分布、代表値、データのばらつき
 - 観測データに含まれる誤差
 - アルゴリズムとプログラミング
 - プログラミング言語 Python と各種ライブラリ
- 第 04 回 Python 入門 1
 - データの種類 (整数, 実数, 文字列,)
 - 簡単な計算 (式, 演算子, 四則演算,)
 - 変数とは, 代入文と等号, 文字列, 関数
- 第 05 回 Python 入門 2
 - 文字列操作, 関数とメソッド
 - 文字列の演算子と型変換
 - 条件式と分岐 1
 - コメントと Markdown セル
- 第 06 回 Python 入門 3
 - 条件式と分岐 2
 - ブール型と論理演算子
 - 処理の繰り返し
 - オブジェクト, リスト・辞書・タプル
- 第 07 回 外部データの読み込み
 - Python ライブラリの利用 (pandas 等)
 - 外部ファイルデータの読み込み
 - 列・行データの表示, 追加, 削除
 - データの集計, 並べ替え, グラフ表示
- 第 08 回 データを説明する
 - データ表現
 - Python でグラフ作成 (matplotlib)
(棒グラフ, 積み上げ棒グラフ, 円グラフ等)
- 第 09 回 データを扱う
 - データの特徴や傾向をとらえる
データの集計 (和, 平均, 記述統計)
 - データの並び替え
 - エクセルや R の活用
- 第 10 回 データの関係性をとらえる
 - 基本的なグラフ (大小関係, 変化, 要素の割合)
 - ばらつきをとらえるグラフ (比較, 関係性)
 - 見やすいグラフの作成
- 第 11 回 データの関係性から予測する
 - 回帰分析, 散布図と相関係数, 回帰直線で予測
- 第 12 回 データを地図上にプロットする
 - コンビニの緯度, 経度データを集める
 - データを地図上に表示する
- 第 13 回 機械学習 (AI) に挑戦 1
 - 機械学習とは

- ・ scikit-learn ライブラリでの例
- ・ 「教師あり学習」で手書き数字の認識
- ・ 第14回 機械学習 (AI) に挑戦2
 - ・ 手書き数字を認識させてみる
 - ・ AI の活用事例
- ・ 第15回 まとめ
 - ・ データサイエンスと Society5.0
 - ・ AI, IoT, 5G/6G, ビッグデータ

2.5 授業内で実習に使用したデータ

授業 (演習) で実際に使用したデータは、インターネットで公開されているものを使用した。

- ・ 新型コロナウイルスに関するデータ^[3]
- ・ 過去の気象データ^[4]
- ・ 震度データベース^[5]
- ・ コンビニチェーン売上等データ^[6] 等

公開データを CSV 形式でダウンロードし、エクセルで一部整形後、pandas で DataFrame 形式で読み込み処理することとした。ただし、PDF 形式のデータはエクセルで変換・整形して CSV 形式に変換して利用した。

2.6 実習の進め方

プログラミングに不慣れな学生も多いので、授業内で実行するプログラム例はすべて LMS で事前に提示し、授業内ではプログラム例をコピーまたはキーボードから直接入力することで、まずは実際に動作させることを重視した。当初はすべてキーボードから入力させていたが、スペルミスによるエラーや入力速度が遅いなどにより、途中で作業を止めてしまうことがあった。このため、解説の際は、受講生のスキルに応じてまずはコピーで実行し、授業後にあらためてキー入力して動作を確認・復習するように指導した。

2.7 質問、感想・コメント

毎回授業後 LMS で、質問はもちろん、感想やコメントを必ず記入してもらっている。質問とその回答・補足説明等は、次回授業までに LMS で全員に公開し、次回授業の冒頭 (15 分前後) に解説している。また、感想やコメントのうち、全員で共有が望ましい事項も同時に紹介し、必要に応じて授業改善にも活用している。さらに、個別事情によるトラブルは別途メールやオンラインサポートで支援している。もちろん、簡単な質問は授業内でも受け付けて対応した。

2.8 課題・演習例

本授業での主な課題は、「データを読み、使い、分析し、論じる」を実践するため、できるだけ身近な話題を取り上げた。

●Python 入門の主な課題

- ・ 123 から 567 までの整数を足すといくら？
- ・ 体重 (単位は Kg) と身長 (単位は m) を入力し、BMI を計算して表示するプログラムを作成。さらに、適正体重も表示。

●データ活用に関する主な課題

- ・ 地震の震度の度数分布 (ヒストグラム)。また、震源の深さとマグニチュードの散布図作成。
- ・ 平均気温、最高気温、最低気温の推移折れ線グラフ (1 つのグラフの中に) を作成
- ・ 2 行 1 列で、2020 年、2011 年の高知市の平均気温の折れ線グラフを。さらに、年間平均気温を計算し、水平線 (赤色) を描き入れる。
- ・ 6 行 2 列で、COVID-19 の感染者数の推移 (全国、指定の都道府県) の折れ線グラフを作成。
- ・ コンビニチェーンの売上高と来店客数の散布図から関係性を考察。さらに、COVID-19 の感染状況の影響も考察。
 - ・ ある会社の広告費から売上を予測。
 - ・ 地図上に、① AED 設置場所データと自分で調査した 2 箇所をプロット、②追加した 2 箇所のアイコンを他とは変える、③高知県庁に Home アイコンをプロットし、県庁を中心にサークルマーカーをプロット。さらに、AED 設置数によりアイコンや色を変えるなど。
- ・ scikit-learn ライブラリの「教師あり学習」サンプル (手書き数字の認識) の学習データ (イメージ) を表示。それを参考に手書き数字を認識させる。

●最終課題 (A と B は選択)

- ・ A 新型コロナ感染防止対策についての考察
- ・ B 興味のある事項について、データにもとづいた考察
- ・ C 高知県の観光スポット地図の作成
- ・ D scikit-learn の機械学習ツールで、正しく認識される "3", "5", "7" の手書きデータの作成

3. 授業サポートと課題

本学では、2021 年 1 学期共通教育科目 (実技、

実習，実験を除く）については原則オンラインで行うこととなった．本授業も同期型オンラインで開始したが，受講生，特にパソコン操作の不慣れな学生から，対面授業を望む声がある一方，オンラインを望む声も少なくなかった．そこで，対面授業を同時にライブ中継（ハイフレックス形式）を実施した．また，研究室でのサポートはもちろん，オンラインサポートも実施するなど，受講生サポートに関して取り組んできた．

3.1 ハイフレックス授業の取り組み

事前の調査で，約 40%の学生が対面授業を希望していたため，収容可能な教室（教室定員の 50%に制限）を選び，ハイフレックス形式で授業を実施した．本学では，ハイフレックス形式での授業環境は準備されていないため，自前で環境を準備（事前準備や当日の操作を含む）する必要があったが，大きな問題もなく無事に実施できた．残念ながら，この後すぐに新型コロナ感染拡大のため，すべての授業がオンラインになり，以後ハイフレックスでの実施ができなかった．

なお，ハイフレックス授業に参加した学生からは，「解説資料が，教室内のプロジェクタで参照できるので，パソコンでプログラミングに集中できた」，「ちょっとした疑問点がある場で聞けた」など，概ね好評であった．オンライン参加の学生からも，特に問題は指摘されなかった．今後徐々に対面授業に戻ることも予想されるが，同時にハイフレックスも必要になると考えている．

3.2 オンラインサポートの取り組み

オフィスタイムに研究室で質問対応はもちろんのこと，大学に出て来ることができない学生には，オンラインで個別サポートも行なった．同期型オンライン授業でも使用している，Cisco Webex Meetings で終日サポート窓口を開き，質問がある学生はサポート時間内に自由にアクセスして質問できるようにした．Jupyter Notebook 上でのプログラムエラーなどの相談では，学生の画面を共有することで効果的にサポートすることができた．サポートするスタッフが多い場合は，ブレイクアウトルーム側にサポーターが待機し，個別支援することを検討している．

4. まとめ

AI, IoT, ビッグデータ, Society5.0, データ駆動型社会など，「データサイエンス」に興味を持っている学生が予想よりも多かった．本授業が「数理・データサイエンス（リテラシーレベル）」修了のための選択科目であることが受講を促したのかもしれない．パソコン操作の習熟度やプログラミング経験の有無により，個別にサポートする方法を試行錯誤しながら授業を実施した．オンラインでの演習の場合，トラブルやプログラミングのエラーの発生状況などは，リアルタイムで学生 PC をモニタするか，学生の申告がないと把握するのが困難である．これまでは，授業後のオンラインサポートや資料の事前配布によるトラブル・エラーの回避等を行ってきた．今後は，オンライン授業中でもブレイクアウトルームにて TA により質問に応じるなど，サポート体制を強化することを検討している．

参考文献・データ

- [1] 佐々木 正人，石黒 克也，佐々 浩司，高知大学における新入生に対する情報セキュリティ教育の改善，大学 ICT 推進協議会年次大会，2019
- [2] 佐々木 正人，石黒 克也，佐々 浩司，コロナ禍における新入生対象情報教育の取り組み，大学 ICT 推進協議会年次大会，2020
- [3] NHK 特設サイト 新型コロナウイルス，<https://www3.nhk.or.jp/news/special/coronavirus/data-widget/>
- [4] 気象庁 過去の気象データ，<https://www.data.jma.go.jp/gmd/risk/obsdl/index.php>
- [5] 気象庁 震度データベース，<https://www.data.jma.go.jp/svd/eqdb/data/shindo/>
- [6] 日本フランチャイズチェーン協会 コンビニチェーン売上等データ，<https://www.jfa-fc.or.jp/particle/19.html>