

データサイエンス・AI 教育の実践報告（1）

2021/03/04

応用物理学科 玉井良則

1. はじめに

世の中 AI やビッグデータが花盛り。スマホを持っているだけで、毎日データサイエンス・AI の恩恵に預かっています。NHK E テレの高校講座「ベーシック数学」でも、3 学期に入って、箱ひげ図や相関係数など、データサイエンスの内容をやっていて驚きました。

さて、文部科学省では「数理・データサイエンス・AI 教育の全国展開」を推進しており、その関係で、今年度後期に、共通教育科目「データサイエンス・AI 入門」を担当させていただきました。また、それに合わせて、応用物理学科 1 年前期の「情報処理基礎」のシラバスも改定し、データサイエンスの内容を含めてみました。その実践報告をさせていただきます。

2. ターゲットは「リテラシーレベル + α 」

「楽しくて役に立つ授業」を目標とし、前期の「情報処理基礎」と後期の「データサイエンス・AI 入門」を続けて受講すると、文部科学省の言うところの「リテラシーレベル」+ α が身につくような設計としてみました。要素を表にまとめると以下の様になります。

表 1 授業のトピックと具体的な内容

科目	トピック	内容
情報処理基礎	情報リテラシーの基礎	PC の基本操作, クラウド活用, 図書・文献情報 情報セキュリティ, 情報倫理
	オフィスソフトの活用	レポート作成の基本, 著作権 ワープロ, 表計算, プレゼンテーション技術
	データサイエンスの初歩 (データの扱いに慣れる)	オープンデータの活用 e-Stat 家計消費データを用いた実習
データサイエンス・AI 入門	現代社会とデータサイエンス	データサイエンスとは ビッグデータ活用とデータ倫理, GAFA の台頭)
	データ分析の基礎	ヒストグラム, 箱ひげ図, 散布図, 平均値と分散 回帰直線, データ分析の注意点
	データサイエンスの手法	クロス集計, 回帰分析, ベイズ推論, クラスタ分析 決定木, ニューラルネットワーク, 機械学習
	Python を使った AI 実習	Python の基礎, COVID-19 データの可視化 古典文学 (くずし字) の文字認識

データサイエンス・AI 教育はまさにアクティブラーニングであり、遠隔授業との相性も良いです。授業は Classroom と Meet を使ったリアルタイム遠隔方式で行いました。Python の実行環境は、Google の仮想サーバをブラウザから利用できる Colaboratory を活用しました。

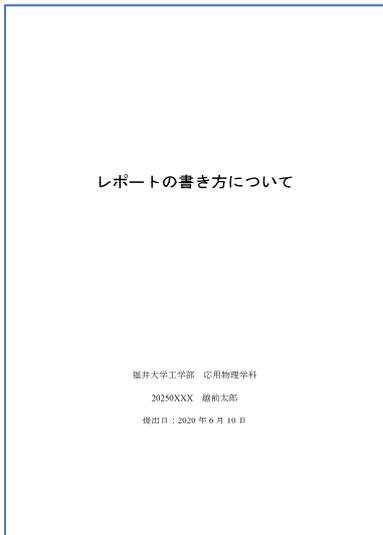
3. 情報処理基礎

3-1 情報リテラシー

データサイエンス・AIを活用するためには、前提として基本的な情報リテラシーが身につけている必要があります。いくらデータ分析を行っても、的確にレポートにまとめたり、プレゼンで表現したりできなければ、宝の持ち腐れとなってしまいます。また、情報セキュリティや著作権など、データを活用するにあたって留意すべきことが多々あります。これらを「情報処理基礎」でしっかりと身に付けておくことが重要です。

単調で堅苦しい授業にならないように、学生には「量子力学を応用したナノテクノロジーの展開」、「地球温暖化で福井の冬は寒くなる？」など、応用物理に関連した10テーマの中から好きなテーマを選び、情報検索、技術レポートの作成、プレゼン資料の作成などの一連の流れを体験してもらうようにしました。ちなみに、AI関連のテーマが1番人気でした。

今年度はリアルタイム遠隔授業で行いしましたが、クリッカーを使った「著作権クイズ」を導入したり、チャットなどで質問対応したりして、以前よりもインタラクティブな授業になったような気がします。「レポートの作成の基本」などは、他の授業のレポート作成にも役立ててもらえたのではないかと思います。



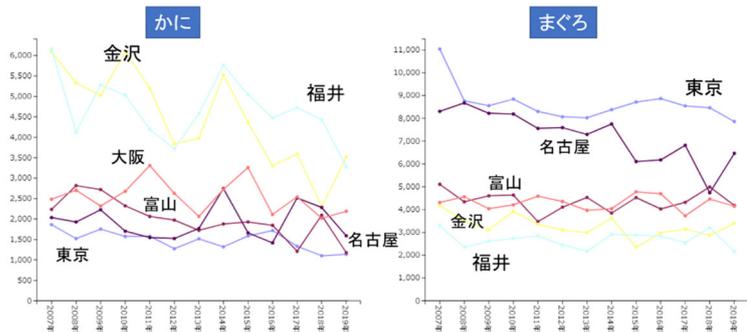
3-2 データサイエンスの初歩～データに慣れる～

データサイエンスについては、「データの扱いに慣れること」を目標とし、無料のオープンデータ e-Stat (政府統計の総合窓口) <https://www.e-stat.go.jp/> を活用しました。下図はその一例ですが、世帯 (2人以上) あたりの、かにとマグロの年間消費支出の年次推移をグラフに表したものです。都市別に比較してみると、越前ガニの福井、築地の東京、かに道楽の大阪など、いろいろと連想が湧いてきて楽しめます。



自由にデータを選んで解析してもらえると、興味深いデータを掘り起こしてくれる学生が多数いました。例えば、アラブ首長国連邦の人口ピラミッドは、社会・経済情勢を反映した非常に興味深いものでした。ぜひ、e-Stat で調べてみて下さい。

「かに」と「マグロ」の消費支出 (年次データ)

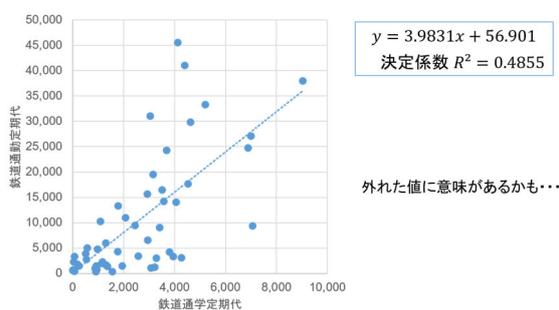


3-3 データ間の相関を考察

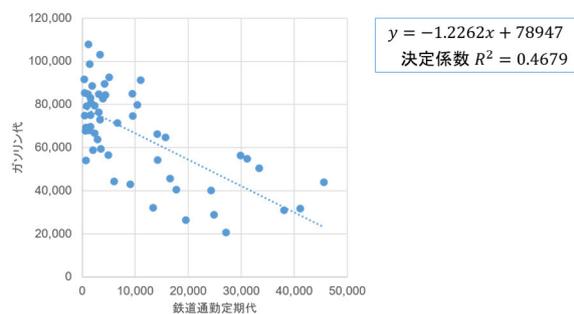
実習では、e-Statの家計調査データから2種類を選び、その相関について考察するという課題に取り組みました。下図はその一例で、世帯当たりの定期代やガソリン代の県庁所在地（+政令指定都市）別データを散布図に表したものです。通学定期代と通勤定期代には正の相関、通勤定期代とガソリン代には負の相関があることがわかります。これだけでも、回帰、相関係数、決定係数などデータサイエンスの基礎が詰まっています。また、左側のグラフをよく見ると外れ値があります。上にずれている（通勤定期代が多い）のはさいたま市や川崎市、下にずれている（通学定期代が多い）は神戸市です。こういった外れ値には特別な意味がある場合もあり、社会・経済情勢などとも絡めて情報検索をしてみると、いろいろな考察ができます。

データの組み合わせは無数にあるため、まったく同じレポートが出されることはありません。非常に面白い分析をしてくれた学生も多く、レポートを読んでいるだけでかなり楽しめました。例えば、ワイシャツと乳液の消費支出の相関を都道府県別に解析し、家庭内における夫婦間の力関係について論理展開したレポートは、着眼点もよく、読み応えがありました。情報リテラシーの基礎を学んだ直後ですので、グラフの作成やレポートの書き方も含めて、非常によくまとまったレポートが多くありました。

通学定期代と通勤定期代の関係 ～正の相関～



通勤定期代とガソリン代の関係 ～負の相関～



3-4 情報処理基礎とデータサイエンス・AI教育

情報処理基礎にデータサイエンスの内容を含めることのメリットは、学んだばかりの情報リテラシーの基礎を即実践に使うことにより、知識やリテラシーの定着が確実になることだと思います。一方で、情報セキュリティ、情報倫理、著作権なども含めて、今後ますます高度な知識が要求されることが予想され、情報処理基礎のももとの内容を削ってまでデータサイエンスの内容を詰め込むことには、限界があります。今回試行したように、「データに慣れる」ことに焦点を絞ると、既存の情報処理基礎の内容と比較的相性よく、内容が拡張できると思いました。

情報処理基礎をご担当のみなさんは、4月からの授業にデータサイエンスの要素を取り込んでみてはいかがでしょうか？授業を担当している側も結構楽しめますので、おすすめします。参考資料としてシラバスを添付しておきます。

（「データサイエンス・AI入門」については、また後日、報告させていただきます。）

授業科目名	(B0019) 情報処理基礎AP	単位数	2
開講開始学期	前期	曜日/限目	火/4
授業区分	週間授業	科目授業種別	共通教育/基礎教育科目/情報処理基礎科目
担当教員 (E-mail、電話番号 (内線)、研究室、オフィスアワー、実務経験の有・無)			
玉井 良則 (質問は Classroom のコメントを利用、総合研究棟I 東館 5階、木16:30~18:00、実務経験:民間企業)			

ナンバリングコード

GB-BAC-101 共通教育 / 情報処理基礎科目〔1年次レベル〕

授業概要

【受講対象: 応用物理学科】

実務において不可欠な情報リテラシーおよびデータリテラシーの基礎を習得します。

インターネットの活用法、および、オフィスソフトの活用法を習得し、データサイエンスの初歩的な実習に取り組みます。

到達目標

- ・インターネットを利用したコミュニケーション、情報収集ができる。
- ・PCを利用してレポート、グラフ、プレゼンテーション資料が作成できる。
- ・オープンデータを活用した初歩的なデータ分析ができる。

授業内容

=== 第1部 インターネットとデータ活用 ===

- 第1回 Windows PCの基本操作 (ログオン, ログオフ, ファイル操作, ネットワーク, アプリケーションソフト)
- 第2回 電子メールの利用と情報セキュリティ
- 第3回 Webによる情報検索, 学内の情報リソース (キャリアサポートシステム解説を含む)
- 第4回 図書・文献情報の収集 (図書館職員による講義と演習)

=== 第2部 オフィスソフトの活用 ===

- 第5回 ワードプロセッサ(1) 日本語入力, 文章の作成, 文字修飾, ヘッダとフッタ, ページ番号
- 第6回 ワードプロセッサ(2) 画像の挿入, 図形の作成, 数式の挿入, 校閲, 印刷
- 第7回 表計算ソフト(1) 数値・文字データの入力, 数式セル, グラフ作成 (棒グラフ, 折れ線, 散布図)
- 第8回 表計算ソフト(2) データの並び替え, 相関係数, 近似曲線
- 第9回 レポートの基本作法 (表紙, 章立て, 文体, 図表の配置とキャプション, 文献引用と著作権)
- 第10回 プレゼンテーションソフト(1) タイトルページ, スライド作成, 図の作成, アニメーション
- 第11回 プレゼンテーションソフト(2) 発表原稿 (ノート), スライドショー, 配布資料印刷

=== 第3部 データサイエンスの初歩 ===

— Excelによるデータ分析 —

- 第12回 e-Stat (政府統計の総合窓口) の利用法, データ取得, データ分析のための前処理
- 第13回 ヒストグラム, 平均値と分散, 相関係数
- 第14回 回帰分析
- 第15回 分析結果のまとめ方と表現, 最終レポート執筆

準備学習 (予習・復習) 等

- ・授業ホームページに課題があらかじめ公開されている。必ず予習をして授業に臨むこと (30分/週)。
- ・課題の提出締め切り日があらかじめ設定され、ホームページに明記してある。締め切りまでに計画的に自己学習時間を確保し、課題に取り組むこと (平均3時間30分/週)。

授業形式