

データビジュアライゼーションを用いた 親しみやすいデータサイエンス教材の提案 — Wolfram Cloud の利用 —

吉田智子

京都ノートルダム女子大学

金光安芸子

ウルフラムリサーチ

北村美穂子

京都ノートルダム女子大学

「数理・データサイエンス・AI 教育プログラム」 の教材に関して

大学において文系理系学部を問わずデータサイエンスの必修化が広がる中、2021 年度から文部科学省は「数理・データサイエンス・AI 教育」を体系的に行っている大学等に対して認定制度を開始した。しかし、初年度の 2021 年にこの制度のリテラシーレベルの認定を受けたのは 78 大学のみにとどまった^{☆1}。現在も全学的なデータサイエンス教育を模索している大学が多い。

筆者らが所属するのは、初年度に上記の認定を受けた文系女子大学である。データサイエンス教育を含む情報教育において、特に文系学部では教材の選び方が重要である。親しみやすいことに加えて、学生の興味関心との接点があれば、必要性が感じられ学ぶ意欲が高まる。さらに、学生自身のデザイン力や創造性が発揮できる教材が受け入れられやすい¹⁾。

^{☆1} 文部科学省「数理・データサイエンス・AI 教育プログラム(リテラシーレベル)」の認定等について(2021 年 8 月 4 日), https://www.mext.go.jp/content/20210804-mxt_senmon01-000016191_2.pdf



図-1 学生によるワードクラウド作品『吾輩は猫である』『みだれ髪』, ほか

本稿ではこれらの条件を満たすものとして、データの可視化(以下、「データビジュアライゼーション」)をテーマとした教材(図-1)を紹介する。ワードクラウド作品の制作には、Wolfram Research, Inc. (2022 年 4 月に京都ノートルダム女子大学と産学連携協定を締結)が提供するプログラミング言語である Wolfram 言語が持つ WordCloud 関数を利用している。ビッグデータの活用が促される昨今、どのような学問、職種でもデータを扱うため、その視覚的な表現の必要性は実感しやすい。自分のプレゼンテーションなどに利用する楽しみも生まれるからである。

「AI とデータサイエンス入門」の授業に導入

ワードクラウドは、アンケートや文章などのテキストデータを可視化する手法で、出現回数の多い単語をより大きく表示する。そのため、どのようなワードがよく多く出てきているのかを、瞬時に認識することができる。“今日ツイッターでバズったワードは?”や、“アンケートの回答で多く出てきたワードの可視化”など、ニュースや新聞、TV の情報番組などでもよく利用される手法である。

このワードクラウド作成実習を、本学の“データ活用プログラム”^{☆2}の必修科目の 1 つである「AI とデータサイエンス入門」(2 単位, 90 分授業 15 回)の中で実施した。この科目は 2021 年以降に入学し

^{☆2} 文部科学省の「数理・データサイエンス・AI 教育プログラム」で認定された筆者の大学でのプログラム名。



た全学部が2年生以降に選択できるもので、筆者の北村と金光の2名がいわゆるティーム・ティーチング形式で一緒に2つのクラス（各クラスの定員は40名）を担当する形とした。以下は、2022年度の秋学期の授業の報告である。

まず、15回の授業は原則、「講義+実習」という形式で行われ、データビジュアライゼーションは第12回目に行われた。前半の約45分の講義の概要は次のとおりであった。

1. データビジュアライゼーション
 - クロス集計と棒グラフの例
 - テキストマイニングの例（ワードクラウド、共起ネットワーク）
 - 地図上の可視化の例（ポイント分布マップ、ヒートマップ、塗りつぶしマップ）
2. データに騙されないために（チャートジャンク）
これらの内容の講義の後にPCに向かい、データビジュアライゼーションの作成実習を行った。

Wolfram 言語を使った実習内容

第12回目の授業後半の実習では、まずワードクラウドを作成する手順を学んだ。利用したプログラミング言語は、先にも述べたように Wolfram 言語であった。これは科学技術計算ソフトウェア Mathematica の中核をなすプログラミング言語であり、機械学習を含むテキスト・画像・音声・動画などのデータ分析に必要な関数を豊富に備えた関数型の言語である。

□ ワードクラウド作成手順

具体的にワードクラウド（図-2）の作成手順を紹介



図-2 同じデータの異なる形でのワードクラウド（授業のアンケート結果より）

介する。授業では Wolfram Cloud 環境^{☆3}を用いた。その場合の手順は、

- (1) 形態素解析済みのテキストファイルのアップロード
- (2) ファイルの読み込み
- (3) ワードクラウドの作成

となる。以下に、具体的に紹介する。

● 形態素解析済みのテキストファイルのアップロード

Wolfram Cloud 環境でワードクラウドを作成するために、あらかじめ形態素解析を行ったテキストファイルを Wolfram Cloud にアップロードする。形態素解析とは、文章を分かち書きし（単語に分ける）、標準形に直す（思っ→思う、など）処理のことである。今回は、教員が日本語形態素解析エンジン MeCab を利用して形態素解析をした後、ストップワード（「を」「は」「です」などの助詞や不要語）を除去したテキストファイルを学生に提供した。

● ファイルの読み込み

ファイル（たとえば、output.txt）の読み込みのためのプログラム記述は、次の2行である。1行目の記述を実行することで、現在のディレクトリにパスを指定する。2行目では、Import 関数で「output.txt」を読み込み、変数 data にセットする。

```
SetDirectory[NotebookDirectory[]];
```

```
data = Import["output.txt"];
```

data 内容の確認のため、WordCounts 関数でファイル中にあるワードをカウントしてみる。ちなみにこの data は学生の授業コメントである（図-3）。

● ワードクラウドの作成

WordCloud 関数で、読み込んだテキストのワードクラウドを作成する（図-4）。

四角形のほかに、ハートやドーナツ型などの

^{☆3} ブラウザ (<https://www.wolframcloud.com/>) で Wolfram 言語を利用できるクラウド環境。アカウント登録をしておくことで、PC、スマホ、タブレットから共通して利用できる。無料の Basic Plan と有料プランがあり、今回の授業の学生は有料プラン (Mathematica Online Student 4 カ月間) を利用した。なお、Basic Plan ではファイルのアップロードができないため、Basic Plan でワードクラウドを作る場合は、形態素解析済みのテキストファイルを PC 上にテキストエディタで開き、コピー&ペーストで Wolfram Cloud のノートブックに貼り付ける方法を使う。

好きな形を指定して作ることもできる(図-2)。

□ 個性的なワードクラウド作品の誕生

授業では、上述の操作を一緒に実行した後、レポート課題の説明を行った。その課題とは、提供された以下の10個の文学作品全文のテキストファイル(青空文庫^{☆4}から入手)から2作品を選んでワードクラウドを作り、そこから分かることや工夫した点を記述することだった。

<10個の文学作品>

『吾輩は猫である』(夏目漱石), 『人間失格』(太宰

治), 『みだれ髪』(与謝野晶子), 『源氏物語 桐壺』(紫式部, 與謝野晶子訳), 『源氏物語 夕顔』(紫式部, 與謝野晶子訳), 『母を尋ねて三千里』(アミーチス), 『みにくいアヒルの子』(アンデルセン), 『白雪姫』(グリム), 『ラプンツェル』(グリム), 『若草物語』(オルコット)

学生から提出された作品では、作品のイメージに合わせた、とてもバラエティに富んだワードクラウドが集まった(図-5)。図-5のワードクラウドの文学作品名は次の通りである。

1行目: 『ラプンツェル』『ラプンツェル』『みにくいアヒルの子』

2行目: 『若草物語』『母を尋ねて三千里』『みにくいアヒルの子』

☆4 青空文庫, <https://www.aozora.gr.jp/>



図-3 dataの表示結果



図-4 標準的な形で作成



図-5 さまざまな文学作品のワードクラウド(学生の作品より)



3行目：『白雪姫』『白雪姫』『白雪姫』

4行目：『源氏物語 桐壺』『みだれ髪』『吾輩は猫である』

5行目：『人間失格』『人間失格』『源氏物語 夕顔』
複雑な色合いやオリジナルの形のワードクラウドが多く見られるが、ここまでのアレンジはレポート課題を学生に課した時点で教員側は想定していなかった。学生らは自らオンラインドキュメント^{☆5}で調べ、「ハートやドーナツ型ができるのであれば、文学作品のイメージに合う画像で形を作りたいから方法を教えてください」という質問が寄せられた結果、オリジナルの形のワードクラウドを作るための手順書を追加で用意したのである。

ここで簡単にその方法を紹介しておく(図-6)。

- 形にしたいイラストを用意して、白黒に2値化した形を作成
- データファイルをその形でワードクラウド化(WordCloud 関数を利用)

『ラプンツェル』のワードクラウドをカメレオンの形にしたのは、ラプンツェルの大親友がカメレオンだからということを学生の作品から知った。ここで紹介できなかったワードクラウド作品も含めて、以下のような工夫が観察できた。

- 『白雪姫』のワードクラウドを「リング型」や「小人型」で作成

.....
☆5 <https://reference.wolfram.com/>



図-6 好きな画像を使う方法

- 『母を尋ねて三千里』を「サル型」で(主人公マルコのペットのサルをイメージ)作成
- 『ラプンツェル』を「お花型」で色は「紫色」で
- 『人間失格』は文豪作品なので、「白黒」で表現、など

この実習から得られた主な効果

今回、学生たちの力作が揃った要因の1つは、扱うデータを文学作品にしたことが、文系学生に受け入れられたからだと考えている。そして、Wolfram言語の操作やファイルの呼び出しや保存に苦戦しながらも「自分の理想のワードクラウドを作りたい!」という一心で、一生懸命(そして、恐らく楽しみながら)取り組めたのは、自分の創造力を発揮できる課題であったからであろう。さらに、Wolfram CloudはWebブラウザから手軽に利用できるもので、週末に自宅でも作業が続けられたことも要因の1つだと思われる。

また、近年の学生のPC利用においては、ファイルの置き場所を意識しない傾向が強い。しかし、Wolfram言語をWolfram Cloudで使う場合にはファイルの置き場所を意識する機会が度々ある。たとえば、講師から提供されたプログラムファイルを自分が利用する場合に「Make your own copy」のボタンを押すと「Copied Files」と呼ばれるフォルダに保存される(図-7と図-8)。この場合、自分が利用するファイルに行きつくには、「Copied Files」をクリックして階層を下りる必要があり、フォルダの階層構造を意識せざるを得ない(図-9)。

また、実習のために必要なデータファイルや画像ファイルを自分のPCからWolfram Cloudにアップロードすることもある。ファイルがどこに保存されているかを意識しないと、自分が作りたい作品が作れない状況を体験することで、階層構造となっているクラウド側のフォルダにファイルを保存し、呼び出す力がつくことも確認できた。

女子大学が協力し合っの教材開発に期待

本学の「AIとデータサイエンス入門」の授業では、今回紹介したクラウドの教材以外にも、自らが写した写真や録音した音声を用いて情報科学の基礎（デジタル化の仕組み）を学ぶ教材や、プログラミングによるアート作品制作、お菓子の種類を見分けるための Wolfram 言語による機械学習プログラミング教材など、さまざまな教材を学生に提供している。そして、それらの教材を積極的に公開しており²⁾、学外向けの講演会や市民向けの DX リカレント講座でも紹介する。

さらに、2023年4月には、文部科学省の推進する「数理およびデータサイエンスにかかわる教育強化」の協力校(20校)の1つであるお茶の水女子大学と「文

理融合データサイエンス教育に関する覚書」を締結した^{☆6)}。お茶の水女子大学は「AI・データサイエンス教材」を複数の女子大学に提供しており、すでに日本女子大学、東京女子大学、東京家政学院大学と同様の覚書を交わしている。これらの女子大学が教育研究分野で協力体制をさらに強化することにより、意欲・能力のある幅広い分野を専攻する学生がデジタル社会で活躍する機会が増えることを期待している。


参考文献

- 1) 吉田智子, 有賀妙子, 真下武久: ジェンダーインクルーシブなプログラミング教材の開発, 情報処理, Vol.62, No.12, pp.667-671 (2021), doi/10.20729/00213743
- 2) 金光安芸子, 北村美穂子, 吉田智子: AIとデータサイエンス入門—Wolfram言語を使ったプログラミング実習テキスト(Mathematica ノートブックにて提供) (2023), <https://wolfram.ai-datascience-textbook> (2023年6月23日受付)


☆6) お茶の水女子大学と京都ノートルダム女子大学が文理融合データサイエンス教育に関する覚書を締結(2023年4月6日), <https://www.ocha.ac.jp/news/d012598.html>




図-7 講師からの提供ファイルは「Make Your Own Copy」を押すことで編集可能に(Wolfram Cloud 画面)



吉田智子 (正会員) tyoshida@notredame.ac.jp
 京都ノートルダム女子大学社会情報課程長教授。
 2000年より京都ノートルダム女子大学専任教員。情報教育の教材開発に興味を持ち活動中。



金光安芸子 (正会員) akikok@wolfram.com
 Internationalization Specialist / Wolfram Certified Instructor, Wolfram Research, Inc. 2022年度より京都ノートルダム女子大学「AIとデータサイエンス入門」講師。



北村美穂子 (正会員) kitamura@notredame.ac.jp
 京都ノートルダム女子大学社会情報課程特任教授。Wolfram Alpha LLC, Consultant. 博士(工学)。自然言語処理一般、および、その教育利用に興味を持ち活動中。

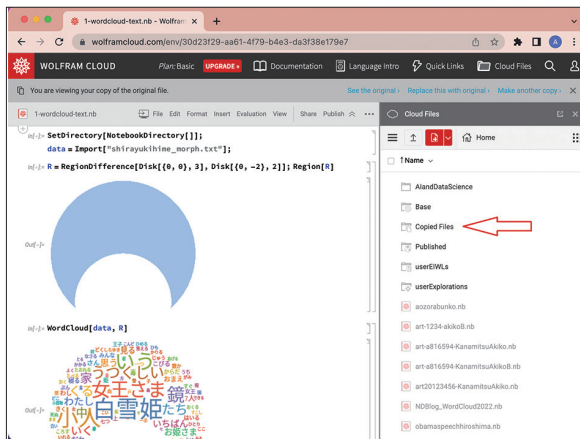


図-8 編集可能ファイルは Copied Files フォルダに保存される

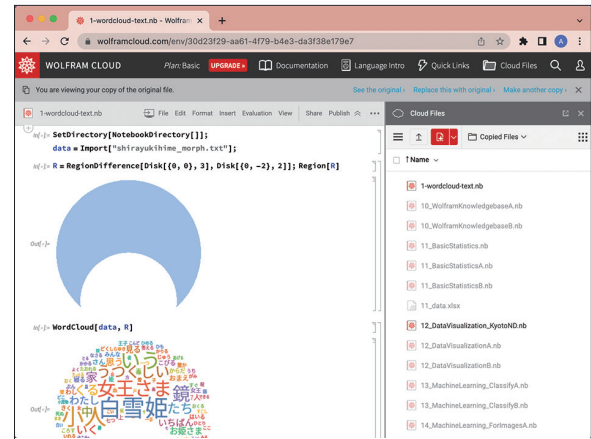


図-9 Copied Files フォルダのファイルを表示したところ

