

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

北村 美穂子*1

金光 安芸子*2

吉田 智子*1

*1: 京都ノートルダム女子大学 社会情報課程

*2: Wolfram Research, Inc.

本日の内容



- ■背景
- ■「AIとデータサイエンス入門」の授業内容
- ワードクラウドを用いた授業の紹介
- ■授業実践の結果(学生アンケート)
- ■まとめ

背景



- 京都ノートルダム女子大学
 - 1961年開学、国際言語文化学部、現代人間学部を有する文系女子大学
 - 2023年4月社会情報課程 新課程開設
 - 文部科学省「数理・データサイエンス・AI教育認定制度 |
 - 2021年度「リテラシーレベル」, 2023年度「応用基礎レベル」認定



- ■「AIとデータサイエンス入門」(2022年度後期)
 - 文系学生による学びの難しさ
 - 高校時代の数学・情報の基礎を学んでいない(忘れている)学生が多い
 - 難しい、苦手との思い込み
 - 仕組みや理論からの説明では、思考停止

「AIとデータサイエンス入門」授業の進め方



- 数学・情報の学び直し
- 将来, 学生の役に立つ知識
- 学生に苦手意識を感じさせない(これなら私でもできる!)
- ■面白いと思わせる工夫



毎回の授業は「講義+実習」の二部構成(教員2名によるチームティーチング)

[講義] 実世界で使われるAIやデータサイエンスに主眼をおき,そこで利用されている 技術や仕組みを学ぶ

[実習] <u>学んだ知識を確認するためのプログラミング実</u>習を行う(Wolfram言語を利用)

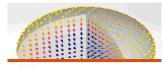
Wolfram言語



AIとデータサイエンスに欠かせない 高度な関数が豊富に揃っているため, 講義の内容に沿った実習が可能

計算知識エンジンWolfram|Alpha 数式処理システムMathematica のベースである 関数型プログラミング言語

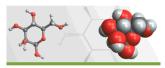
- 微積分,線形代数,統計などの数学 関数
- 画像分析や音声分析、自然言語処理 や機械学習に関する関数

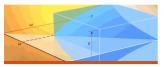


可視化





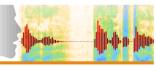


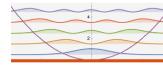


微積分と代数



最適化

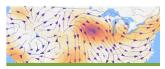




AI用ツール









授業計画

回	内容] r
第1回	ガイダンス • 授業の目的・進め方についての説明	
	• WolframAlpha を使ってみよう	-
第 2 回	AI による技術および社会の変化	1
	◆ AI による技術の変化(インターネット, IoT, 5G), AI に	
	よる社会の変化(ソサエティ 5.0, シェアリングエコノ	
	ミー他) ● Wolfram 言語の基本操作の説明と実習	
第3回	データ活用のための基礎知識	
,,,,,,,	データの種類、コンピュータ内でのデータの扱い(数字、文字)	
	● プログラミング実習(Wolfram 言語におけるデータ処理)	盾
第4回	データ活用のための技術(データ解析, 非構造化データ) ・ データ解析とは、非構造化データ(画像、音、音声) の扱い	
	プログラミング実習(画像, 音, 音声処理)	┞
第 5 回	第2回から第5回のまとめ	1
	• 小テストおよび講評	
第6回	プログラミング言語概論	
	• プログラミング(アルゴリズムとデータ構造), プロ	
	グラミング言語の歴史,種類と特徴 ■ プログラミング実習(条件分岐, 反復等の基本制御)	
第7回	プログラミング実習	1
	プログラミング実習(アートプログラミング)	
第 8 回	AI の事例 自然言語処理(1)	
	• テキスト解析(形態素解析, 構文解析)	
	プログラミング実習(文字列処理, テキストの文法構造) お	

第 9 回	AI の事例 自然言語処理(2) ■ テキスト解析(意味解析, 辞書とコーパス) ■ プログラミング実習(Wolfram 言語の知識ベース)
第 10 回	第6回から第9回のまとめ小テストおよび講評アートプログラミング課題の作品発表・ディスカッション
第 11 回	 データサイエンスの基礎 AI と統計学の関係、データの種類と収集方法、ヒストグラムと基本統計量、統計解析 プログラミング実習(可視化、基本統計量の計算、相思想係)
第 12 回	 データビジュアライゼーション 様々な可視化(グラフ, ワードクラウド, 地図上の可視化) プログラミング実習(データビジュアライゼーション)
第 13 回	機械字省とは • AI と人間の学習の違い,教師あり学習・教師なし学習とは,機械学習(分類,予測) • プログラミング実習(Wolfram 言語の機械学習,分類,予測)
第 14 回	AI が社会に与える影響 AI 活用における課題と未来(AI と仕事, AI と上手く付き合うには) データビジュアライゼーション課題のレポート発表・ディスカッション
第 15 回	全体のまとめ まとめのテストおよび講評本授業の総括

第12回 データビジュアライゼーションの授業



- 可視化の種類(棒グラフ, 地図上の可視化など)
- 可視化の重要性
- チャートジャンクなど
- プログラミング実習:ワードクラウドで可視化



出現回数が多いほど,大きい文字で描画される ワードクラウドで可視化する

実習後,レポート課題 文学作品のワードクラウドを2個(以上)作成し, 可視化によってわかったこと,工夫した点を記述

実習の準備



■ 青空文庫で提供されている以下の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)
「ラプンツエル」(グリム)
「若草物語」(オルコット)

■ 教員が 青空文庫(https://www.aozora.gr.jp/) から入手した全文テキストを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」などの助詞や助動詞)を除去し, テキスト データファイルとして学生に提供

プログラムの手順



(1) SetDirectory関数でファイルのパスを設定し、Import関数でテキストファイルを読み込む.

SetDirectory [NotebookDirectory []];

data = Import ["respon1.output.txt"];

(2) WordCloud関数でワードクラウドを作成する.

WordCloud [data]

WordCloud [data]

WordCloud [data]

WordCloud [data]

学生には、単にワードクラウドを作ってもらうことだけを期待していたが...

自分の好みの形でワードクラウドを作りたい!

学生からの質問 好きな画像からワードクラウドを作る方法



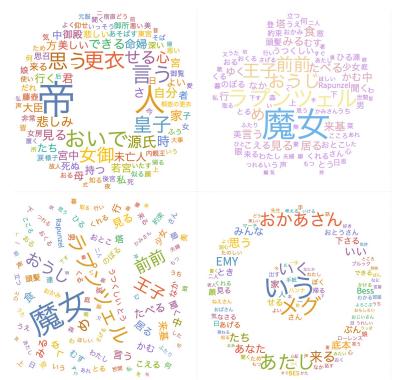
以下の関数の使い方を, Wolframのノートブックで共有

wagahai = Import["aozorabunkodata/wagahai_morph.txt"];
WordCloud[wagahai,

学生の作品







学生の作品





学生の声



- ■主人公の名前や登場するキャラクターのワードが大きく表示され、それらのワードが多く登場していることがわかった。
- すべての小説において、一人称、二人称、人物名がよく出てくるということがわかった。また、動詞も多く、特に多いのは"いう"という動詞であることがわかった。このことから、人物が何を言ったのかという文章が多いのではないかと思われる。
- 『ラプンツェル』のワードクラウドからは、主人公の"ラプンツェル"よりも"魔女"の方が多く出現していることに驚いた.この結果は、ワードクラウドを作成してみないと分からないことなので、これこそがワードクラウドの良さだと思った.

まとめ



- ワードクラウド等, 題材選びを工夫
 - 文系の学生にも受け入れられやすい身近な題材
 - 学生が自分の創造力を発揮し、楽しく意欲的に取り組めた
- Wolfram言語, Wolfram Cloudでデータサイエンスのエッセンスを体感
 - 比較的短いコードでプログラムが書けるWolfram 言語を毎回使用
 - プログラム言語に不慣れな学生も、実習や課題にスムーズに取り組めた
 - 学生がやりたいと思ったタイミングを逃さず積極的に取り組めた
- 今後の発展
 - 教材テキストをWebで公開中 https://wolfr.am/ai-datascience-textbook
 - 「応用基礎レベル」の教材開発





■ 動画公開中「楽しく学べるデータサイエンス教材の提案」https://www.youtube.com/watch?v=j73tGl9Rj-A&t=87s

参考となるWeb情報



• 発表論文

- https://conference.ciec.or.jp/pdf/2023pcc/pcc003.pdf
 PCカンファレンス2023で発表
 「初学者のための AI・データサイエンス入門授業の試み」(北村 美穂子・金光 安芸子・吉田 智子)
- https://conference.ciec.or.jp/2023pcc/program/subcommittee/presentation/pcc011.html
 「体感しながら学ぶAIとデータサイエンス入門授業の試み」(金光 安芸子・北村 美穂子・吉田 智子)

- 冊子掲載

情報処理学会 学会誌 2023年10月号(9月15日発行)ペた語義 「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」 (吉田 智子・北村 美穂子・金光 安芸子) https://www.ipsj.or.jp/magazine/peta-gogy.html にてWebページとしても公開(9月15日に公開予定)

- 教材テキストをWebで公開中
 - 京都ノートルダム女子大学 Wolfram実習教材 https://wolfr.am/ai-datascience-textbook
 - 第4回 データ活用のための技術(データ解析,非構造化データ) https://wolfr.am/KNDU_DSTextbook_4
 - ND講演会発表NoteBook (ユーザ登録すれば利用可)

https://www.wolframcloud.com/obj/akikok/Published/KNDU Wolfram Seminar2023 DataScienceText Presentation akikok.nb

■ PCカンファレンス2023発表NoteBook

https://www.wolframcloud.com/obj/akikok/Published/pcc011 2023 Kanamitsu presentation.nb