



近年のデータサイエンス・AIの進展

— 文系・理系を問わず活用できる時代に —

富山大学 中川 大

1 はじめに

● 近年のデータサイエンスの進展

情報システムのイメージは、以前は「理数系の人の世界・難しい分野」

⇒ 現在は、「だれでも使える身近なツール」。

「現代の読み書きそろばん」

高度なツールが登場していることによって、入門は次第に簡単化。

⇒ 入門は楽しく、興味が出れば無限の可能性。

データサイエンス学習は、
データサイエンスの面白さと有用性を知ることが目標に。

● 本日の内容 下記の4つのことを実演

- ① クリックだけで、統計データを地図に表現できる。
- ② プログラミングは文系の人にとっても難しくないし、面白い。
- ③ AI技術の成果も、ほとんどカットアンドペーストのみで利用可能。
- ④ ディープラーニング・ニューラルネットワークも自分で計算できる。

2 近年のデータサイエンス 以前との違い

●データの作成・収集

(以前) 最初から自分でデータベースを構築する必要。
(現在) 多くのデータが、デジタルデータとしてwebで提供されている。有料・無料のものがあり、公的データを中心として多くのデータが無料で提供されはじめている。

●データの取り扱い・分析

(以前) 統計手法を勉強して、自ら計算する必要。
(現在) エクセルをはじめとするデータベースソフトが、広く普及し、計算のためのプログラムを自分で作成する必要はほとんどなくなっている。

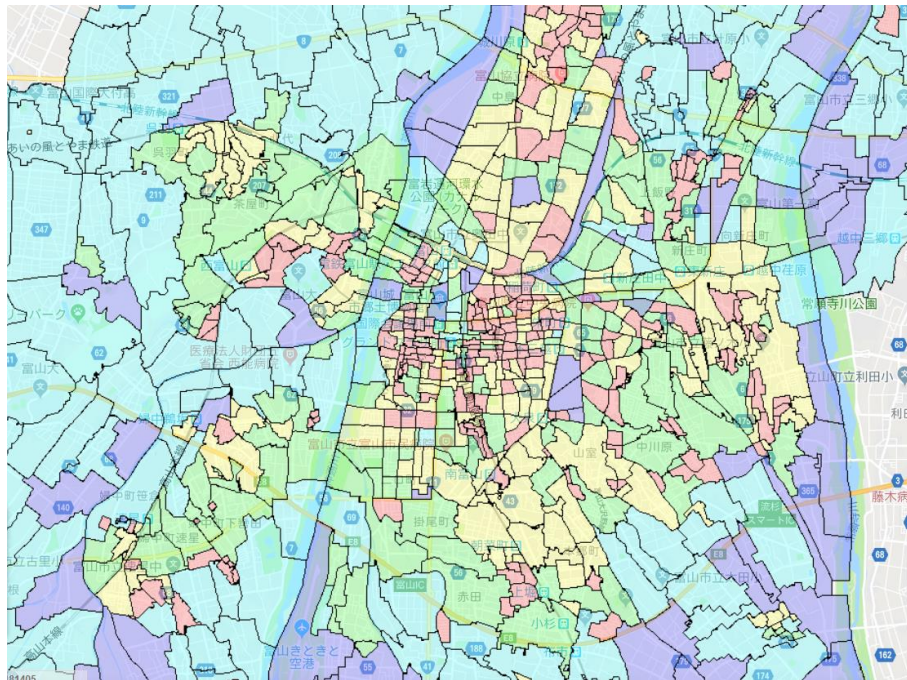
●プログラムの作成

(以前) 最初から自分でプログラムを書く必要。
(現在) ほとんどのことは、サンプルプログラム、テンプレート、ライブラリーなどとして、webなどで提供されている。有料・無料のものがあり、多くの事項は無料で提供されるもので十分処理できる。

実演1 クリックだけで、統計データを地図上に表現できる

政府統計の総合窓口 e-Stat 「統計でみる日本」
(無料・登録無しでもすぐに使える)

小地域別 (町丁目別) 人口分布
(富山市)



金沢駅から車で20分の範囲



実演2 プログラミングは難しくない

プログラミングは小学校の必修科目に ⇒ 誰でもできるということ。
プログラムは、「言語」。語学は文系の人の方が得意なのは。

文法は中学英語よりもはるかにやさしく、登場する単語の数もはるかに少ない。

また、英語とは異なって例外は少なく単純。

しかも、近年は最初から自分で書く「英作文」ではなく、「例文の書き換え」でほとんどのことが可能。

実演内容

Python (いま人気のプログラミング言語)を使って、ネットで面白そうなプログラムを見つけて動かしてみる。

まずは、ネットで「Python ゲーム 簡単」で検索。

実演3 AI技術の成果を使ってみる

AIを使ったツールを提供しているサイト。

「IBM Watson」、「Microsoft Azure」、「Google Cloud」、「Nanonets」など、すでにたくさんある。

本日は、

「IBM Watson」の「音声合成 Text to Speech」をPythonで利用。
テキストファイルを自動で読み上げて音声ファイルを作成するプログラム。

手順

- (1) IBMのwebページで利用者登録をして、自分の暗号(APIキー)を取得する。(取得済み。無料。クレジットカード登録等もなし。)
- (2) IBMのwebページに書かれているプログラムをカットアンドペーストして、「APIキーと音声化したい文章を記入して」実行する。

実演4 ディープラーニング・ニューラルネットワークを使ってみる

ディープラーニングのためのツールも多く公開されている。
例として、**Neural Network Console (Sony)**を紹介。
サンプルデータを用いれば、すぐにでも体験可能。
解説ビデオ (YouTube) も充実。

Neural Network Consoleの画面

The screenshot displays the Neural Network Console interface for an MNIST tutorial. The main window shows a neural network architecture with the following layers:

Layer	KernelShape	W
Input	-	1,28,28
Convolution	5,5	64,32,32
BatchNormalization	-	64,32,32
ReLU	-	64,32,32
Convolution_2	1,1	64,32,32
BatchNormalization_2	-	64,32,32
ReLU_2	-	64,32,32
Convolution_3	3,3	64,32,32
BatchNormalization_3	-	64,32,32
ReLU_3	-	64,32,32
Convolution_4	1,1	64,32,32

The interface also includes a left sidebar with layer and loss functions, a top navigation bar with tabs for 'Main', '学習' (Learning), and '評価' (Evaluation), and a right-hand control panel with buttons for '実行' (Execute), '事前実行(計算時間推定)' (Pre-run (Calculation time estimation)), '学習' (Learning), and '評価' (Evaluation). The control panel also shows hardware specifications for the simulation, such as CPU (9.8 H Free) and GPU options.

3 本日のまとめ

データサイエンスは近年急速に発達し、文系・理系を問わず誰でも活用できるツールになってきている。

多くの学生が楽しみながら取り組んでくれることを期待。

本日デモした内容を実際に試して頂くには、

1. 統計データを地図に表現 ⇒ **政府統計の総合窓口 e-Stat**
2. プログラミング (Pythonを動かす方法)
⇒ **Anaconda** というソフトをインストールすると本日デモした Jupyter Notebook が付いている。
3. AI技術の活用 ⇒ **IBM watson**
4. AIの計算 ⇒ **Neural Network Console (Sony)**

(いずれも無料)