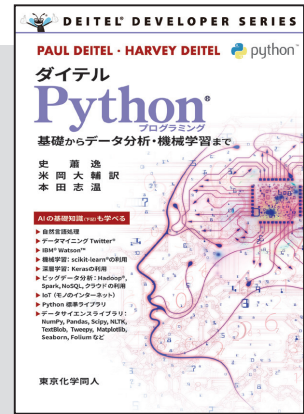


## ダイテル Python プログラミング 基礎からデータ分析・機械学習まで

P. Deitel, H. Deitel 著

史 蕭逸, 米岡大輔, 本田志温 訳

東京化学同人 (2021), 5,280 円 (税込), 576p., ISBN: 978-4-8079-2002-0



### 書籍の概要

本書は、Python プログラミングと題した欧米の教科書の翻訳版である。国内の関連教科書より厚くて分量が多く、まさに欧米の教科書という雰囲気がある。Python そのものは汎用プログラミング言語であるが、簡潔で読みやすく書きやすく、科学技術計算や機械学習のライブラリが豊富であることから、データサイエンスで人気である。Python の書籍は、最近たくさん出ているが、本書は網羅性が高く、記述も詳細であり、教科書的な使用には適する。また、自然言語処理や深層学習、ビッグデータの記述も充実しており、データサイエンス分野での使用を指向した学習者に適している。しかし、数学や統計などの数式的な記述や説明はあまり詳しくなく、ほかの書籍で補足する必要がある。

### 本書の内容

本書は、4部構成で、これに16の章が配置されている。第1～3部では、章末に「データサイエンスの導入」という節が設けられ、各章の話題においてデータサイエンスにおける特別な用法を紹介している。

第1部は「Pythonの基礎的なクイックスタート」である。

第1章の「コンピュータとPythonへの導入」では、Anacondaを用いたPythonのインストール、Jupyter Notebookの導入などが丁寧に解説されて

いる。章末には、AI、IoT、データサイエンスとのかかわりが解説されており、Pythonのデータサイエンスに関連する情報が詳しく載っている。

第2章の「Pythonプログラミング入門」では、演算子や変数の使い方など、基本事項を解説している。章末に組み込み関数 `min`, `max` を用いた記述統計が紹介されている。

第3章の「制御文」では、`if`, `for`, `while` などを用いた制御文とイテレータ、組み込み関数 `range` などを用いた基本的なリスト処理を解説している。章末に `statistics` モジュールの関数 `mean`, `median`, `mode` を用いた要約統計が紹介されている。

第4章の「関数」では、関数の作成方法、Python標準ライブラリの紹介と、モジュールの使い方、`import` の使い方などが紹介され、再帰的なプログラムの作成方法や関数型プログラミングも紹介されている。章末に `statistics` モジュールを用いたばらつきの求め方が紹介されている。

第5章の「リストとタプル」では、2種類のコレクション（複数の要素からなるデータ型）であるリスト（可変型）とタプル（不変型）の使い方を紹介している。リスト内包表記や、ラムダ式、`filter`, `map`, `reduce` など関数型プログラミング的な使い方も紹介している。章末に、`Matplotlib` や `Seaborn` を用いたシミュレーションの結果の描画法を紹介している。

第2部は「Pythonのデータ構造、文字列、ファ

イル」である。

第6章の「辞書と set」では、辞書と set の作成法、関数、内包表記など、一連の使い方を紹介している。章末に、Matplotlib の animation モジュールを用いた動的可視化の方法を紹介している。

第7章の「NumPy を用いた配列指向プログラミング」では、NumPy が提供する n 次元配列型の ndarray の使い方を紹介している。ndarray では、ブロードキャストという強力な操作法を紹介している。章末に、pandas の機能拡張配列である Series と DataFrame を紹介している。

第8章の「文字列：より詳しく」では、文字列の使い方を紹介している。最初に、文字列フォーマットの方法を示し、連結、反復、split メソッド、join メソッドなど基本的使い方を示した後に re モジュールによる正規表現の使い方を説明している。章末に、pandas の Series と DataFrame におけるデータマニピュレーション（データクリーニングとデータ変換からなるデータの前処理）を紹介し、re モジュールによる正規表現を使ったデータの再変換などを紹介している。

第9章の「ファイルと例外」では、ファイルの出入力、書き込みなど、ファイルシステムの扱い方と JSON データフォーマットについて、紹介している。また、例外の発生と try, except, else, finally を用いた例外の処理方法について紹介している。章末に、csv ファイルの読み書きと、pandas モジュールを用いた csv ファイルの DataFrame による処理について紹介している。

第3部は「Python の発展的な話題」である。

第10章の「オブジェクト指向プログラミング」は、Python のオブジェクト指向の紹介である。カスタムクラスの定義、オブジェクトの扱い方、特殊メソッド \_\_init\_\_ による新規オブジェクトの作成

について紹介している。さらに、オブジェクトの属性へのアクセス制御とオブジェクトのプロパティの使用、基底クラスとサブクラスの作成によるクラスの継承階層、基底クラスのメソッドのオーバーライドによる機能追加など Python のポリモーフィズム的な機能などを紹介している。

第4部は「AI、ビッグデータ、クラウドの基礎的なケーススタディ」である。

第11章の「自然言語処理 (NLP)」では、TextBlob による感情分析や言語間翻訳、グラフやワードクラウドによる可視化、Textstatistic を用いた可読性評価、spaCy による固有表現抽出、類似表現抽出などの事例が紹介されている。自然言語処理関連のライブラリやツール、機械学習や深層学習におけるアプリケーション、自然言語のデータセットの紹介も豊富である。

第12章の「Twitter のデータマイニング」は、Twitter 分析の事例であり、Tweepy ライブラリを用いた Twitter 情報の取得、Twitter Trends API を用いたトレンドの発見、ツイートの感情分析などが紹介されている。

第13章の「IBM Watson とコグニティブ・コンピューティング」は、IBM Watson の紹介であり、Watson サービスの導入および使用方法とそのサービス内容が紹介され、事例として旅行者向け携帯翻訳アプリが紹介されている。

第14章の「機械学習：分類、回帰、クラスタリング」は、Python の機械学習用ライブラリ scikit-learn の紹介であり、k-近傍法による Digits データセットの分類、時系列データを用いた単回帰、California Housing データセットを用いた重回帰、教師なし学習としての時限削減と、k-means 法が紹介されている。

第15章の「深層学習」では、Keras を用いた深層学習システムの構築、MNIST データセットによ

る畳み込みニューラルネットワーク (CNN), IMDb データセットによる再帰型ニューラルネットワーク (RNN), TensorBoard ダッシュボードによるニューラルネットワークの学習の可視化, 学習済みニューラルネットワークモデルの利用などについて紹介している。

第 16 章の「ビッグデータ: Hadoop, Spark, NoSQL, IoT」は, ビッグデータの各種プラットフォーム事例の紹介である。最初に SQL データベースの概要が紹介され, 次に SQL スキームによらない NoSQL が紹介されている。さらには, SQL と NoSQL の両方の機能を併せ持った NewSQL が紹介されている。事例として, JSON ドキュメント型データベースである MongoDB を用いたツイートの解析が紹介されている。その後, Apache Hadoop および Apache Spark の説明と, Spark ストリーミングによる事例が紹介されている。最後に, IoT (モノのインターネット) で使用される publish/subscribe モデルが紹介されている。

## 本書籍をだれに薦めるか

データサイエンスの道具として, Python を利用したいすべての学生や研究者, 技術者にお薦めしたい。Python のプログラミング言語としての機能を網羅するほか, 自然言語処理, Twitter, 機械学習, 深層学習, ビッグデータなど, チュートリアルマニュアルやリファレンスではなかなか得難い事例での Python の活用が書かれており, 目から鱗が落ちる思いである。

(2021 年 8 月 15 日受付)

石井一夫 (正会員)  
kishii@rs.sus.ac.jp

公立諏訪東京理科大学工学部情報応用工学科教授, 久留米大学医学部内科学講座心臓・血管内科部門客員准教授。専門分野: ビッグデータ分析, 計算機統計学, データマイニング, 数理モデリング, 機械学習, 人工知能。医療ビッグデータ, 気象ビッグデータ研究に従事。2015 年度本会優秀教育賞受賞。日本技術士会フェロー, APEC エンジニア, IPEA 国際エンジニア。

