

1. 序論: プログラムで何ができる

プログラミング・データサイエンス I

2023/4/13

1 序論

1.1 プログラムとは? 手続き的プログラムの場合

プログラムとは? 手続き的プログラムの場合

- コンピュータの処理を並べたもの
 - 書いた順番に実行
 - 条件分岐
ある条件の時は A を、それ以外は B を実行
 - 繰り返し
各〇〇に処理 A を行う

プログラムというのをイメージしますか。呪文でもありませんし、秘密の暗号でもありません。

プログラムと呼ばれるものには、いくつかの種類があります。通常は、「手続き的プログラム」を差しています。手続き的プログラムとは、コンピュータに行わせる処理を順番に並べたものです。例えば、皆さんが学校の先生であるとしましょう。教室で出席をとる手順を考えましょう。名簿を持って、名前を呼び、返事を確認してチェックマークを付けます。返事がない場合には、欠席者としてマークします。これを全員分繰り返す。最後にチェックマークの数を数える。これが手続きです。

「名前を呼ぶ→チェックマークを付ける」のように、予め定めた通りに実行する部分があります。「返事がない場合→欠席マークを付ける」のように、条件に応じて違う動きをする部分があります。さらに、「全員分繰り返す」のように、繰り返しがあります。

これをプログラミング用の言語を使って書き、実行するのがプログラミングです。とくに難しくありません。

1.2 プログラム (コンピュータ) が得意なのは?

プログラム (コンピュータ) が得意なのは?

- ある処理を (非常に多数回) 繰り返す
- 大量のデータから、条件を満たすものを探す
- 毎日、決まった時刻にある処理を実行する
- ネットワークを使って情報を集める

コンピュータが得意なことは何でしょう。

今回は、試験結果の集計のことを考えましょう。採点結果が表になっていて、ここから平均点を計算しましょう。クラス的人数が少ないうちは、手で計算しても、電卓を使ってもよいでしょう。しかし、すごく多い、例えば大学入試センター試験は50万人が受験しますが、その人数になるとコンピュータが必要です。コンピュータは、同じ処理を大量に繰り返すのが得意です。

大学入試センター試験の次に各大学の個別試験があります。各大学は、大学入試センターから、受験者の成績をもらいます。50万人の中から、ある大学の志願者のデータだけを取り出す必要があります。これも、コンピュータが得意とする処理です。

コンビニエンスストアでは、各商品の売り上げをPOS (Point of Sales) というシステムで集めています。決まった時刻に集計し、必要な商品を店舗に届けます。このように定時に何かの処理を実行することもコンピュータならば簡単です。さらに、この情報は、ネットワークで収集・集計されていることはいまでもありません。

1.3 データサイエンス

データサイエンス

- データ駆動型社会
 - － データが利益を生む
 - － データに基づく経営判断や政策立案
- データサイエンス
 - － データを分析する
 - － データから新たな知識を生み出す

現代はデータ駆動型社会と言われています。例えば、オンラインショッピングサイトは単に商品をオンラインで売っているだけではありません。商品購入のデータから、サイト来訪者に商品を推薦することがありますね。多数の顧客の購入情報は、売れ筋商品の情報にもなります。

SNS (Social Networking Services) も、ユーザに情報交換の場所を提供しているだけではありません。ユーザが交わすメッセージを分析し、何が話題になっているかを抽出し、場合によっては広告等に利用しています。

GAF A と呼ばれる大規模情報企業は、オンラインショッピングサイトや SNS、その他のサービスを提供していますが、それと同時に大量のデータを収集・分析し、そこから利益を得ています。大量のデータは利益を生むのです。

また、企業経営や、行政における政策立案では、根拠、特にデータに基づく判断が重要となってきています。企業経営では、自社の製品やサービスの売上や素材の仕入れ状況、競合他社の状況、消費や景気の動向、自社の人員や施設設備の状況など、具体的な根拠に基づく判断が必要です。さらに、新たな商品やサービスのインパクト、自社の資源の再編成による改善など、データとモデルに基づくシミュレーションを行い、判断をしなければなりません。行政においても、対象となる地域の状況を適切に把握し、何が必要なのかを判断するとともに、施策の影響をシミュレーションしなければなりません。つまり、データが重要になっています。

データの収集と分析は、従来は「統計」という名で呼ばれてきました。しかし、今必要なのは、もっと広い切り口でのデータ分析です。例えば、大量のデータから特徴を取り出す際に、伝統的な数学的手法だけでなく、機械学習という方法を使うなどの変化が起きている。機械学習は、人工知能の一種であり、コンピュータによるデータ分析を行います。

このような新しい分野を、「データサイエンス」と呼びます。

1.4 様々なプログラミング言語

— 様々なプログラミング言語 —

- 高級言語
 - 英語のような単語で記述
 - スクリプト型言語
 - コンパイラ型言語
- 低級言語
 - 機械に近い言語：読みにくい

プログラミング言語には、「高級言語」と「低級言語」があります。べつに、価値が「高級」「低級」という意味ではありません。「低級言語」は、よりコンピュータという機械に近い言語です。従って、読み書きは難しいものです。一方、ほとんどのプログラマーは、「高級言語」を使っています。「高級言語」では、プログラムの全体構造や処理が、ヒトによってわかりやすく記述することができます。世界中で使われるために、英語に似た単語で記述するものがほとんどです。

また、プログラミング言語の別の分類方法があります。プログラムを書くとただちに実行できる「スクリプト型」と、実行前にコンピュータが高速に実行できるように翻訳する「コンパイラ型」です。

Python はヒトが読みやすい「高級言語」であり、すぐに実行できる「スクリプト型」の言語です。

1.5 Python とは

— Python とは —

- 1989 年ごろから、Guido van Rossum が開発開始した言語
- 読みやすく、書きやすいと言われている
- 無償で利用できる
- データ処理や Web アプリなど、非常に広汎に利用されている

この講義で使う Python (パイソン) について説明しましょう。

Python は、無償、つまりタダで使うことができるプログラミング言語で、読みやすく、

書きやすいと言われていました。なぜ、「言われています」という言い方をするかというと、プログラムを書いてきた人々のなかには、「そうかなぁ」という声も少なくないからです。しかし、利用者が最も多い言語の一つであることは間違いありません。

しかし、これまでプログラムを書いてこなかった人々が使い始めているので、読みやすく、書きやすいのかもしれませんが。実際に、データ処理や Web アプリ開発など、広汎な分野で使われています。書籍も多数出版されています。

1.6 この講義の目標

この講義の目標

- 簡単な Python プログラムを書ける
マニュアル、本、Web を見ながらでも可
- プログラムを書いて、何か処理を実行しようと思うことがある

この講義では、皆さんが、簡単な Python のプログラムを書けるようになるのが目標です。仕事で Python を書けるレベルを求めるわけではありませんから、教科書、マニュアル、そして Web 上の記事を見てかまいません。仕事でプログラムを書く人だって、教科書やマニュアルなどを参照します。

そして、この後の学習や就職後に、「これって、ちょっとプログラム書けば、サクッとできる」と思うことができるようになれば、大成功です。プログラムを書くのは理系の子、と考えず、自分もできると思ってください。実際に、情報系の技術者の中には、文系の学部出身者が少なくありません。

1.7 教科書と講義の進め方

教科書と講義の進め方

- 大重美幸「詳細! Python 3 入門ノート」(ソーテック、2017)
- **必ず、教科書を読んでくる**
次回は、2 章から 4 章の部分

プログラミングを学ぶということは、言語を習得することですから、自習が重要です。授業で、全てを教えることはできません。また、英語の学習と同じで、文法を学んだだけでは使えるようにはなりません。他の人の書いた良いプログラムを読むことと、自分でプログラムを書くことでしか、習得できません。

この講義では、大重さんの教科書を使います。教科書を入手し、必ず読んで来てください。もちろん、どんどん先に進んでくれても何も問題ありません。

2 次回

次回は、2章から4章の部分を扱います。必ず読んでおいてください。もちろん、実際にプログラムを動かしてみても構いません。