

編集後記：ここ数年で、機械学習、ビッグデータ、IoTといったワードが気象業界でも頻出用語となり、今や、気象庁の業務においても毎日のように耳にする。機械学習、ビッグデータに関して言えば、ここ数年で統計解析や機械学習の敷居が大幅に下がったことが要因の一つと感じられる。その背景として、データ分析ツールとして利用できるR、Pythonといったスクリプト言語の存在があるように思う。

筆者(開発素人)も利用経験のあるPythonを例にして言えば、機械学習用のライブラリ、統計解析用のライブラリ、それらの利用方法を記述した指南書が充実しており、ほんの数行で、教師あり機械学習における教師データの学習から、精度評価、その可視化まで実行できたり、ほんの数行でクラスタ解析からカーネル密度推定、その可視化までできたりしてしまうなど、コーディング上の敷居が低いのである。特に最近では、深層学習向けのライブラリが充実し、やはり容易に導入可能と聞く。これだけ便利なので、環境構築が大変かと思えば、Anacondaと呼ばれる無償のデータサイエンス向けプラットフォームのセットアップファイル(Linux版、Windows版あり)をダウンロードして実行すれば基本OKという簡便さである。ちなみに、Pythonユーザのことを俗にヘビ使いと言うらしい。

実際に運用するシステムは別の言語で作成するにせよ、こういった敷居の低い解析ツールの存在は、手を動かして直感的に理論を理解するうえで非常に便利であり、機械学習やデータ分析を学ぶ人口の裾野を間違いなく広げていると思う(少なくとも筆者を含めその周辺には多い)。このまま裾野が広がっていけば、今後、気象学界限でも、機械学習による研究アプローチが当たり前となるのではないかとさえ感じる。そんな時代に取り残されないよう、今の内から、ヘビ使いに弟子入りして、この分野を勉強しておこうと思う。

Pythonに関して蛇足だが(これが書きたかった)、前述のような機械学習・統計解析用のライブラリ群だけでなく、分野によっては、解析用のライブラリが公開されていることがある。例えば気象レーダーで言うと、Py-ART: The Python ARM Radar Toolkitと呼ばれるツールキットがあり、レーダーデータの可視化や、既往研究で示されている手法(例えば降水粒子判別や降水強度推定ほか)も容易に適用でき、調査や研究それ自体の敷居を下げている。このように理解を促すツールを提供するあたり、知恵の実を勤めるヘビのイメージが重なる。

(梅原章仁)