

AI…Pythonだけじゃないもう一つの定番データ解析「R」入門

ご購入はこちら

足立 悠

データ分析向き「R」を使う理由

● データ分析2大言語「R」と「Python」

データ分析に使えるソフトウェアはいろいろあります(表1)。中でも「R」と「Python」と呼ばれる言語環境がよく使われています。

RとPythonは、どちらにも、次のような特徴があります。

- (1) データ分析に必要な「統計解析」、「機械学習」、「ディープ・ラーニング」のアルゴリズムが備わっており、ライブラリ・APIとして利用できます。
- (2) Windows, Mac, Linuxで動作し、オープンソースであり無料で使えます。すぐにインストールして利用することができます。

● 対応している機能はどちらも同じようなもの

どちらも主だったデータ分析機能を備えています。

▶ 統計解析機能

基本統計量の計算や各種検定などを行います。

▶ 機械学習機能

回帰や分類などのアルゴリズムを用いた予測分析、グループ化やパターン抽出などの知識発見、いわゆる

データ・マイニングを行います。

予測分析は例えば、機械・設備の故障予測、障害の検出などに利用できます。

知識発見は例えば、ウェブ・コンテンツのオススメ抽出などに利用できます。

▶ ディープ・ラーニング機能

機械学習の一種であるため、同じような機能ですが、機械学習に比べて次元の多いデータの扱いに強いというメリットがあります。センサ・データや購買履歴などの構造化データは機械学習のアルゴリズムを用いる、画像・音声・言語データなどの非構造化データはディープ・ラーニングのアルゴリズムを用いるといった、データの形式に応じて使い分ける必要があります。

● データ解析の世界では双壁

大量のデータの中から、意味のある情報を見つけ出す技術 データ・マイニングに関する情報サイト「KDnuggets」が2016年に実施したアンケート「過去1年間でデータ分析に使用したソフトは何か?」の結果において、Rは1位を、Pythonは2位を獲得しました⁽¹⁾。

データ解析の世界では、双壁といってもよいと思います。

表1 データ分析に使う代表的なソフトウェアあれこれ

ソフト名	データ加工	機械学習アルゴリズム				可視化	操作	形態	ライセンス
		回帰	分類	グループ化	パターン発見				
R	△ ○	○	○	○	○	○	CUI	無料	GPL
NYSOL	○ ←(得意) ×	×	△	×	○	△	CUI	無料	AGPL
Python	△ ○	○	○	○	○	○	CUI	無料	PSFL
SQL	△ ×	×	×	×	×	×	CUI	無料	-
Excel	△ ○	○	×	×	×	○	GUI	商用	-
RapidMiner	△ ○	○	○	○	○	○	GUI	無料/商用	AGPL/-
Hadoop	△ ○	○	○	○	○	△	CUI	無料	Apache 2.0
Spark	△ ○	○	○	○	○	△	CUI	無料	Apache 2.0
Tableau	△ ×	×	×	×	×	○	GUI	商用	-
KNIME	△ ○	○	○	○	○	△	GUI	無料	GPL
scikit-learn	×	○	○	○	○	△	CUI	無料	BSD