

# 基于 Python 的数据可视化

杨凯利<sup>1</sup>, 山美娟<sup>2</sup>

(1. 北方光电股份有限公司, 陕西 西安 710043; 2. 西安欧亚学院, 陕西 西安 710065)

**摘要:** 在大数据快速发展的今天, Python 丰富的工具包在科学计算、文件处理、数据可视化等领域越来越凸显其价值。能够发现数据、清洗数据, 并使用正确的工具实现数据可视化至关重要。本文叙述了 Python 软件第三方库的安装和绘图工具的应用, 并利用 Numpy 和 Matplotlib 库进行数据可视化, 分析大学生通识能力情况, 通过实践展示了 Python 库的强大功能和使用步骤, 为学习数据可视化的读者提供参考。

**关键词:** Python; 数据可视化; 数据分析; 第三方库

**中图分类号:** TP391.1

**文献标识码:** A

**文章编号:** 2096-4706 (2019) 05-0030-03

## Data Visualization Based on Python

YANG Kaili<sup>1</sup>, SHAN Meijuan<sup>2</sup>

(1. North Electro Optic Co., Ltd., Xi'an 710043, China; 2. Xi'an Eurasia University, Xi'an 710065, China)

**Abstract:** With the rapid development of big data, Python's rich toolkits are becoming more and more valuable in the fields of scientific computing, file processing, data visualization and so on. It is essential to be able to discover, clean and visualize data using the right tools. This paper describes the installation of the third library of Python software and the application of drawing tools, and uses Numpy and Matplotlib libraries for data visualization, analyses the general knowledge ability of college students, and demonstrates the powerful functions and using steps of Python library through practice, which provides a reference for the readers who study data visualization.

**Keywords:** Python; data visualization; data analysis; third-party library

## 0 引言

Python 是目前市面上用于大数据分析的优先选择, Python 数据分析功能强大, 对数据抽取、收集整理、分析挖掘及展示, 都可以实现, 避免了开发程序的切换。Python 的数据挖掘能力和产品构建能力兼而有之, 是跨平台且开源的技术, 成本较小。

数据时代, 通过数据分析挖掘数据的价值, Python 就是很好的选择, 它包含了 Numpy、Pandas、Matplotlib、SciPy、iPython 等主要数据分析库。数据可视化是数据探索的主要途径。数据可视化的目标是通过所选方法的视觉展示、清晰有效地与用户交流信息。有效的可视化有助于分析和推理数据和证据。这使得复杂数据更容易接触、理解和使用。

## 1 数据可视化工具

俗话说, 一图胜千言。数据可视化不仅是一种技能, 更是一门艺术。Python 中的可视化工具包括通用的软件工具和特定软件组件。通用的软件工具是集成开发环境 (IDE), 这是一种同一软件包内囊括所有生产工具的应用程序。Enthought 中的 Canopy 和 Continuum Analytics 中的 Anaconda 都是比较方便的工具。特定的软件绘图组件

是 Python 绘图库, 比如 Bokeh、iPython、Matplotlib、NetworkX、Scipy、Numpy、Scikit-learn 以及 Seaborn。

Plotly 是一种在线分析和数据可视化工具。为了更好地协作, Plotly 提供了在线画图、分析学和统计工具。这种工具通过将 Python 与使用 JavaScript 的用户界面和 D3.js、HTML 和 CSS 创建的可视化库一起使用来构建。Plotly 包括多语言兼容的科学绘图库, 比如 Arduino、Julia、MATLAB、Python 和 R。

最常用的可视化类型包含以下分类: 比较和排名、相关性、分布、位置定位或地理数据、局部到整体的关系、随时间的变化趋势等。

## 2 Python 第三方库的使用

Python 语言有标准库和第三方库两类, 标准库随 Python 安装包一起发布, 用户可以随时使用, 第三方库需要安装后才能使用。强大的标准库奠定了 Python 发展的基石, 丰富的第三方库是 Python 不断发展的保证, 随着 Python 的发展, 一些稳定的第三库被加入到了标准库里面。

### 2.1 Python 第三方库的安装

Python 第三方库有三种安装方式, pip 工具安装、自定义安装和文件安装。最常用且最高效的 Python 第三方库安装方式是 pip 工具安装, 本文只介绍 pip 工具安装方法。

pip 是 Python 内置命令, 需要通过命令行执行。通过

pip-h 命令将列出 pip 常用的子命令。安装一个库的命令格式为: pip install<拟安装库名>, 比如安装 Matplotlib 库, 就可以用命令 pip install Matplotlib, 打开 Windows 命令程序 cmd.exe, 进入 Python 安装目录所在的文件夹中, 并进入 Scripts 文件 (pip 在 Scripts 目录下), 输入命令, pip 工具默认从网络上下载 Matplotlib 库安装文件, 自动安装到系统中。第三方库安装好后, 可以通过 pip list 命令列出当前系统中已经安装的第三方库。

## 2.2 矩阵运算库 Numpy 和绘图库 Matplotlib 概述

Numpy 是用于处理含有同种元素的多维数组运算的第三方库。Numpy 库还包括三角运算函数、傅里叶变换、随机和概率分布、基本数值统计、位运算、矩阵运算等非常丰富的功能。

Python 有很多绘图包, 而 Matplotlib 是它们中极其热门的一个软件包。Matplotlib 是提供数据绘图功能的第三方库, 其 pyplot 子库主要用于实现各种数据展示图形的绘制。Matplotlib 库由一系列有组织有隶属关系的对象构成, 这对于基础绘图操作来说显得过于复杂。而 pyplot 子模块将绘图所需要的对象构建过程封装在函数中, 对用户提供了更加友好的接口。

## 3 利用 Python 第三方库进行数据可视化

雷达图是通过多个离散属性比较对象的最直观工具, 掌握绘制雷达图的方法将会为生活和工作带来乐趣。本例通过问卷调查, 获取了某大学本科一年级不同分院学生在五种核心通识能力方面的数据, 经过数据整理, 使用 Python 来绘制多级雷达图, 即在一组同心圆上填充不规则五边形, 其每个顶点到圆心的距离代表分院学生的某种能力。

### 3.1 大学生通识能力雷达图绘制代码

```
import numpy as np
import matplotlib.pyplot as plt
import matplotlib
matplotlib.rcParams['font.family']='SimHei'
matplotlib.rcParams['font.sans-serif']='SimHei'
radar_labels=np.array(['表达与沟通','团队合作',
'分析与综合信息','创新思维','批判性思维']) # 雷达标签
nAttr=5
data=np.array([[5, 6, 8, 6, 7],
               [9, 6, 5, 7, 6],
               [4, 8, 9, 8, 7],
               [7, 8, 6, 7, 8],
               [8, 7, 6, 7, 6]]) # 数据值
data_labels=('金融学院','文化传媒学院','信息学院',
'管理学院','人文教育学院')
angles=np.linspace(0, 2*np.pi,nAttr,endpoint=False)
data=np.concatenate((data, [data[0]]))
angles=np.concatenate((angles, [angles[0]]))
fig=plt.figure(facecolor="white")
```

```
plt.subplot(111, polar=True)
plt.plot(angles,data,'o-',linewidth=1.5, alpha=0.2)
plt.fill(angles,data, alpha=0.25)
plt.thetagrids(angles*180/np.pi, radar_labels,1.2)
plt.figtext(0.52, 0.95,'大学生通识能力分析', ha='center', size=20)
legend=plt.legend(data_labels, loc=(0.94, 0.80), labelspace=0.1)
plt.setp(legend.get_texts(), fontsize='large')
plt.grid(True)
plt.savefig('tongshi.png')
plt.show()
```

大学生核心通识能力包含 5 个属性, 设置属性标签 labels, 并预设各分院汇总数据。np.concatenate() 函数用于将数据和角度的数组首尾闭合起来, 便于调用 plot()。图形保存格式为 Python 可支持的 PNG 类型, 最后通过 show() 函数显示图形, 默认图形文件保存路径与代码保存路径相同。

### 3.2 代码运行及实验分析

在 Python3.7 IDLE 编辑环境中输入代码并运行, 可以得到如图 1 所示的内容, 显示各分院学生通识能力五个维度的效果。从图中可以看出, 学校金融学院学生的团队合作能力较高, 信息学院的分析与综合信息能力较高。学校在通识课程设置和能力培养方面可以有所侧重, 通过一些工作坊和俱乐部加强学生薄弱能力的培养, 旨在培养具有积极心态、独立思考和开阔视野的全人。

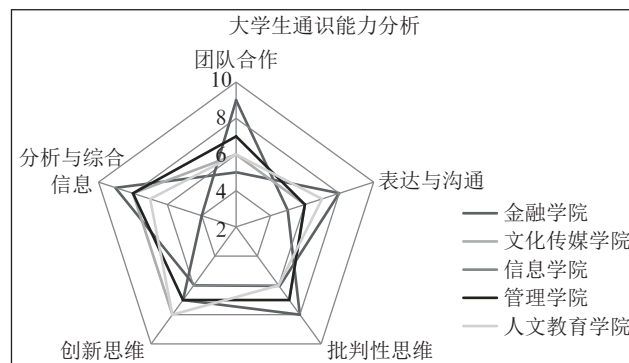


图 1 通识能力雷达图

## 4 结 论

近二十年的开源运动产生了深植于各信息技术领域的大量可重用资源, 直接且有力地支撑了信息技术超越其他技术领域的发展速度, 形成了“计算生态”。Python 语言有大量的第三方库, 形成了庞大的“计算生态”。从游戏制作、数据处理、数据可视化分析等方面, 这些“计算生态”Python 使用者都提供了更加便捷的操作, 以及更加灵活的编程方式, 值得我们不断尝试与实践。

### 参考文献:

- [1] [美] Wes McKinney. 利用 Python 进行数据分析 [M]. 徐敬一, 译. 北京: 机械工业出版社, 2018.
- [2] [印度] Kirithi Raman. Python 数据可视化 (下转 34 页)

任何警察不必真正独自进入不可控情况区域。

### 3 智慧警务数据分析

#### 3.1 警务数据 AI 分析

我们生活在充斥着数据的世界里,利用大量数据来预防犯罪是现代警务工作的一个重要需求。随着许多部门开始部署增强现实、车身摄像头、车牌阅读器和智能传感器等技术解决方案,每天必然会产生比整个模拟时代历史时期更多的数据。未来执法战略的成功,则在于是否能够快速、有效地利用这些庞大的数据来支持调查及其他执法行动。

现实的问题是这些数据量之大,任何人即使穷其一生都用于搜索也是显得过于庞大。为了能够真正地洞察数据中的线索,人工智能(AI)和机器学习必将是未来调查的关键。例如:2013年9月,搜狐有新闻中报道,“在周克华案件中,长沙全市一千名视频侦查民警短短两个月内观看了将近30万G的监控视频,相当于83万部电影”,当时新闻中用了“短短两个月”形容民警效率之高,然而,在目前的技术条件下,借助于人工智能和视频分析技术,使用100台电脑,花费4个小时基本就能完成这一任务。

通过分析模式,传感器馈送和记录数据库,AI可以帮助执法部门确定关键位置,找到嫌疑人之间的关键联系,并探索隐藏在数据海洋中的其他见解。

#### 3.2 预测性警务数据分析

预测性警务的目的是通过缩小潜在的犯罪活动,支持犯罪调查,提高传统警方调查工作的效率。预测性警务数据分析涉及大型数据集的处理,利用预测模型,根据不同类型的输入数据集提供输出。预测性警务数据分析包括利用地理空间技术和基于证据的警务模式,以此来帮助执法机构进行犯罪分析。

借助警务大数据及高级分析,预测可能犯罪的时间和地点、未来的罪犯、受害者的人口统计资料和潜在犯罪分子过去的犯罪行为,以帮助预测犯罪,不但使执法机构的决策能力得到明显改善,还为执法机构提供了先发制人的机会,可以重点关注犯罪易发地区,重点关注潜在犯罪者或被攻击目标。可能情况下,在犯罪发生前制止它。这对社会、潜在犯罪者和被攻击目标都是有意义的。

以预测为主导的警务数据分析涉及如下内容:首先,收集数据并选择使用的分析方法。其次,分析用于提供给特定部门特定犯罪所涉及细节的数据。这里的分析包括热点分析这样的挖掘分析,主要用于预估更多可能的犯罪行为,提示过去发生的事件,提醒即将发生的潜在犯罪行为。最后,最好分析防止未来犯罪的安全措施。例如:可利用社会分析进行风险地形分析等。

## 4 数字警务的关键支柱

#### 4.1 协作与投资

世界各国政府都在投入大量资金用于整合数字技术,以加强其数字警务方面的准备。例如:英国政府通过“警察转型基金”为英国警察部队的转型提供1亿欧元,该基金主要用于解决网络犯罪的取证问题。我国公安系统也在积极建立统一的云平台。

在加大投资和与执法部门加强内部合作的同时,与外部机构和服务提供商合作,是改善警务的最重要步骤之一。目前在全球范围内,执法机构都正在与新技术提供商合作,试图开发一个较好的解决方案。

#### 4.2 教育与培训

在智慧警务时代,必须确保每个直接参与公众服务的警务人员都能够通过数字解决方案应对犯罪活动。为此,须向每一个警察提供一些关于此类技能组合的必要培训。这可以通过公安院校提供类似网络犯罪培训课程的方式来实现。

## 5 结论

技术的发展与应用已经证明,利用大数据,执法部门能够极大地提高执法效能,进而获得巨大收益。大数据技术可以自动完成当前艰巨而耗时的任务(例如:对通信数据的人工分析);有效使用预测分析则可以制定积极的防范和打击犯罪策略,避免发生案件——回应犯罪事件这一简单模式的出现,将资源放在最需要的地方;算法风险评估工具可用于预测与个人相关的风险,如预测重新犯罪的概率或识别处于增加状态的弱势群体失踪或受到伤害的风险;可通过开源数据来获得数据并深入了解社会大部分地区的不同犯罪问题。

然而要真正实现数字警务变革,有效利用数据的技术能力是关键。预计在未来发展中,大数据会整合到制定警务策略所需的基础架构中,而执法人员必须具备相应的技术才能有效利用这些大数据。

#### 参考文献:

[1] Police Foundation. Law Enforcement Executive's Guide to Open Data [EB/OL].<https://www.policefoundation.org/publication/law-enforcement-executives-guide-to-open-data/>, 2017-11-01.

[2] 张浩. 民警两月看30万G监控视频一眼认出周克华 [EB/OL]. <http://news.sohu.com/20130925/n387200192.shtml>, 2013-09-25.

[3] 李俊莉. 大数据多维感知与智慧警务探析 [J]. 网络安全技术与应用, 2018(12): 76-77.

[4] 王海燕, 胡婷, 王俊. 智慧警务系统在地市公安局的应用研究 [J]. 计算机科学, 2018, 45(S1): 518-522.

作者简介: 庄海燕(1976-), 女, 汉族, 河南南阳人, 副教授, 硕士, 研究方向: 信息安全、电子数据取证与分析。

(上接31页) [M]. 程毫, 译. 北京: 机械工业出版社, 2017.

[3] 高扬. 数据科学家养成手册 [M]. 北京: 电子工业出版社, 2017.

[4] 嵩天, 礼欣, 黄天羽. Python语言程序设计基础 [M]. 北京: 高等教育出版社, 2017.

[5] 叶惠仙. Python在学院招生数据分析中的应用 [J]. 计算机时代, 2018(11): 102-106.

作者简介: 杨凯利(1978-), 男, 汉族, 陕西户县人, 本科, 工程师, 研究方向: 自动控制、电路与系统; 山美娟(1978-), 女, 汉族, 陕西户县人, 硕士, 讲师, 研究方向: 数据分析、图像处理。