Python程序设计

第八讲科学计算、数据分析与可视化 Python生态圈的扩展模块

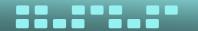


张 华 WHU

- *科学计算
 - ▶ Numpy: 表达N维数组,数据分析和科学计算的基础库
 - ▶ Scipy: 数学、科学和工程计算功能库
 - ▶SymPy: 用一套强大的符号计算体系完成诸如多项式求值、求极限、解方程、求积分、微分方程、级数展开、矩阵运算等等计算问题
- *数据分析
 - ▶ Pandas: Python数据分析高层次应用库



- ☀数据可视化(二维/三维)
 - ▶ Matplotlib: 高质量的二维数据可视化功能库
 - mplfinance: 金融数据可视化分析库
 - ➤ Seaborn: 统计类数据可视化功能库
 - ▶ TVTK: 科学计算三维可视化功能库
 - ➤ Mayavi: 三维科学数据可视化功能库
 - ➤ Chaco、 Vpython
- * 图形用户界面
 - ▶PyQt5: Qt开发框架的Python接口
 - ▶wxPython: 跨平台GUI开发框架
 - ▶ PyGObject: 使用GTK+开发GUI的功能库
 - ➤ TraitsUI: 建立在Traits库基础上的用户界面库



■ Python生态圈的扩展模块:

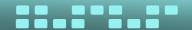
- * 文本处理
 - ▶ PyPDF2: 用来处理pdf文件的工具集
 - ▶NLTK: 自然语言文本处理第三方库
 - ▶ Python-docx: 创建或更新Microsoft Word文件的第三方库

* 机器学习

- ▶ Scikit-learn: 机器学习方法工具集
- ▶MXNet: 基于神经网络的深度学习计算框架
- ▶ TensorFlow: 谷歌公司推动的开源机器学习框架
- ▶ Keras: 基于Python的深度学习库
- ▶ PyTorch: 使用GPU和CPU优化的深度学习张量库



- *网络爬虫
 - ▶ Requests: 最友好的网络爬虫功能库
 - ▶ Scrapy: 优秀的网络爬虫框架
 - ▶ pyspider: 强大的Web页面爬取系统
- * Web信息提取
 - ➤ Beautiful Soup: HTML和XML的解析库
 - ▶Re: 正则表达式解析和处理功能库
 - ▶ Python-Goose: 提取文章类型Web页面的功能库



- * 网站开发
 - ▶ Django: 最流行的Web应用框架
 - ▶ Pyramid: 规模适中的Web应用框架
 - ▶ Flask: Web应用开发微框架
- * 网络应用开发
 - ➤ WeRoBot: 微信公众号开发框架
 - ▶aip: 百度AI开放平台接口
 - ▶ MyQR: 二维码生成第三方库



■ Python生态圈的扩展模块:

*游戏开发

- ▶ PyGame: 简单的游戏开发功能库
- ▶ Panda3D: 开源、跨平台的3D渲染和游戏开发库
- ▶cocos2d: 构建2D游戏和图形界面交互式应用的框架

* 虚拟现实

- ▶ VR Zero: 在树莓派上开发VR应用的Python库
- ▶pyovr: Oculus Rift的Python开发接口
- ▶ Vizard: 基于Python的通用VR开发引擎

* 图形艺术

- ▶Quads: 迭代的艺术
- ➤ ascii_art: ASCII艺术库



- * 计算机视觉
 - ▶PIL: Python的图像处理库
 - ➤ OpenCV: 开源的计算机视觉库,提供了Python接口
- * 语音识别与合成
 - ➤ apiai
 - > google-cloud-speech
 - > pocketsphinx
 - > SpeechRecognition
 - > watson-developer-cloud
 - > wit
- *



扩展库安装包

■ NumPy、Pandas和Matplotlib的网站:

- NumPy: https://www.numpy.org/
- Pandas: https://pandas.pydata.org/
- Matplotlib: https://matplotlib.org/
- 大量科学扩展库安装包下载:
 - * http://www.lfd.uci.edu/~gohlke/pythonlibs/
- 国内下载源
 - https://pypi.tuna.tsinghua.edu.cn/simple
 - https://mirrors.aliyun.com/pypi/simple
 - https://pypi.douban.com/simple



科学计算扩展库

numpy

**科学计算包,支持N维数组运算、处理大型矩阵、成熟的广播函数库、矢量运算、线性代数、傅里叶变换、随机数生成,并可与C++/Fortran语言无缝结合。

scipy

** scipy依赖于numpy,提供了更多的数学工具,包括矩阵运算、线性方程组求解、积分、优化、插值、信号处理、图像处理、统计等等。



数据分析和可视化扩展库

pandas

* pandas (Python Data Analysis Library) 是基于numpy 的数据分析模块,提供了大量标准数据模型和高效操作大型数据集所需要的工具,可以说pandas是使得Python能够成为高效且强大的数据分析环境的重要因素之一。

matplotlib

* matplotlib模块依赖于numpy模块和tkinter模块,可以绘制多种形式的图形,包括线图、直方图、饼状图、散点图、误差线图等等,图形质量可满足出版要求,是数据可视化的重要工具。