

Python语言在数学地质课程教学中的应用

周子勇, 李 力

中国石油大学(北京)地球科学学院, 北京
Email: 595086313@qq.com

收稿日期: 2020年11月16日; 录用日期: 2020年11月27日; 发布日期: 2020年12月4日

摘 要

数学地质尽管是地学类专业的一门选修课程, 但也是一门重要的课程。选修该课程可以增强学生用数学思维解决地质问题的能力。Python语言是目前用于科学计算与可视化最广泛的编程语言。因此把Python语言结合数学地质教学, 可以丰富数学地质的教学内容, 也有助于学生对数学地质算法的理解, 同时可以提高学生的学习兴趣。本文简要介绍了Python语言的特点、国内外高校地学类专业开设Python语言用于地质数据处理与分析处理的相关课程, 并结合作者在数学地质教学中引入Python语言的情况, 探索了Python语言在数学地质教学中的应用。

关键词

Python, 数学地质, 教学

The Application of Python in the Teaching of Mathematical Geology

Ziyong Zhou, Li Li

College of Geosciences, China University of Petroleum (Beijing), Beijing
Email: 595086313@qq.com

Received: Nov. 16th, 2020; accepted: Nov. 27th, 2020; published: Dec. 4th, 2020

Abstract

Mathematical geology is an elective but very important course for most of Geoscience majors in China. The goal of the course is to improve the ability of students to use mathematical thinking in solving geological problems. Python is one of the most widely used programming languages for scientific computing and visualization. Therefore, the combination of Python and mathematical geology teaching can enrich the content of the subject, help students better understand the ma-

mathematical geology algorithm, and improve their interest in learning the subject. The key features of Python are briefly described in the paper, and some similar courses in foreign universities that use Python for geoscience data processing are presented. Finally, the practical application of Python in mathematical geology teaching is introduced.

Keywords

Python, Mathematical Geology, Teaching

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

数学地质(数学地球科学(mathematical geoscience))作为面向地学类专业学生开设的一门选修课,对于提高大学生数学素养、认识数学的重要性以及在地学数据的定量化研究方面都有重要意义[1][2]。但是由于数学地质课程涉及到较多的数学内容,不少同学都有一种畏难情绪,不敢选、不愿意选修数学地质。然而这个短板在本科毕业论文或研究生工作、学习期间就有了明显的表现。如 Excel 绘图不规范、不美观;对于 Excel 绘制的一些的图件(如回归分析中的拟合线)未能正确理解,对于 Excel 软件本身的错误(如一元回归分析中,出现拟合优度系数 $R^2 < 0$ 的明显错误)没有意识到等。这些问题都说明部分学生对于概率统计分析中的一些基本概念不很清楚,特别是对于概念后面的真正含义没有完全弄懂。而数学地质课程,恰好能弥补这一短板,让学生真正理解数学公式的含义及其背后的物理意义,从而更好地培养其数学素养。

为了克服学生对数学地质的畏难情绪,增强对基本概念的理解,在数学地质教学过程中,需要采用一些新的教学环节。其中的两个重要环节是实践教学和计算过程和结果的可视化。

实践教学环节主要是运用常用的多元统计软件对实际数据进行处理和分析。运用不同的算法对实际数据进行处理,不但可以让学生进一步理解各种算法的原理,同时也掌握常用的多元统计软件的使用方法,增加学生对数学地质的兴趣。对于数学地质课程中涉及的多元统计方法,Excel + SPSS 或 Excel + Minitab 基本可以满足实验课的教学需求。Excel 软件提供了非常丰富的函数库,对于一些简单的运算,如均值、方差、距离计算、相关系数与相关系数矩阵、协方差与协方差矩阵、简单的方差分析、一元与多元回归分析、随机数的产生等都可以实现,结合函数的嵌套功能,可以完成一些比较复杂的公式运算。而且 Excel 还提供了丰富的绘图功能,加上 Excel 应用广泛,基于 Excel 进行数据处理、分析与可视化是大学生必须掌握的一项基本技能。SPSS 和 Minitab 是常用的统计分析软件。SPSS 和 Minitab 界面与 Excel 相似,操作也比较简单,同时功能比较强大,除一般的数据预处理功能外,两类软件都提供了常用的多元统计方法,包括回归与逐步回归分析、层次聚类与 K-means 聚类、判别与逐步判别分析、主成分分析和因子分析等。但是这两种软件总的来说,图形功能较弱,除通用的多元统计方法外,一些新的机器学习方法没有提供。而且这两个软件都是商业软件,涉及版权问题。

计算过程的可视化,则是把不同算法的计算过程和结果进行可视化表达,从而可以更直观的展示算法的特征和计算过程,加强对算法的理解。这方面,Excel、SPSS、Minitab 难以满足要求,需要新的工具来补充。而 Python 语言则是比较适合的选择。

把 Python 语言引入数学地质教学中, 有很多的优势。首先 Python 是开源、跨平台语言, 而且函数库非常丰富; 其次, 随着 Python 语言的广泛应用, Python 用于地学数据数据分析已经越来越广泛[3]; 再者, 相对于 C 语言, Python 容易上手。作为理工科学生, 一般都学过 C 语言。C 语言是所有编程语言的基础, 对于编程方法、思路的训练是其它语言很难替代的。但是 C 语言比较难学, 特别是学完以后, 很难直接用于解决实际问题, 因此不少同学会认为 C 语言难学, 而且学了也没有用。Python 则恰恰相反。与 C 语言相比, Python 语言非常容易上手, 因此容易使学生有成就感, 从而增加其学习兴趣。基于这方面的考虑, 在数学地质教学中, 尝试把 Python 引入数学地质教学。既让学生初步了解 Python 语言的特点及编程方法, 同时提高学生学数学地质的兴趣。同时, 很多地学软件, 如 ArcGIS、ENVI 等都支持 Python 的脚本语言。因此, 掌握好 Python 可以增强学生运用计算机解决实际问题的能力。

相对于 Matlab 在数字信号处理、高等数学、线性代数教学过程中的广泛应用, Python 应用于辅助教学还比较少。

2. Python 语言及其特点

Python 语言的出现要追溯到上世纪 80 年代, 1989 年开始启用。目前已经发展到 3.7 版。

表 1 为 TIOBE 统计的近 20 年来几种语言的排名情况。由表 1 可知, Python 开始启用 5 年后, 其用户用量已经排名到了 20 位, 从本世纪初开始, 其排名一直上升, 进入编程语言的前 10 位。到 2019 年, 更是超过 C++, 上升到第 3 位, 其优势自不待言。而且在 2007 年、2010 年以及 2018 年 3 次成为年度语言(Hall of Fame)。

Table 1. Annual rankings of Python language in recent 20 years (Data source: www.tiobe.com)

表 1. Python 语言近 20 年来年度排名情况(数据来源: www.tiobe.com)

编程语言	年度平均排名						
	2019 年	2014 年	2009 年	2004 年	1999 年	1994 年	1989 年
Java	1	2	1	1	3	-	-
C	2	1	2	2	1	1	1
C++	4	4	3	3	2	2	2
Python	3	7	6	6	22	20	-

之所以越来越多的用户选择 Python 语言作为编程语言, 主要有以下几方面的原因。

- 1) Python 语言是开源语言, 可以免费获取。同时也是跨平台语言。
- 2) 相对于在数值分析中广泛应用的 Matlab 语言, Python 语言更紧凑, 可读性更强。据统计, 完成同一个任务, Python 语言比 Matlab 语言代码要少 10%~20%。
- 3) Python 语言相对于许多其它编程语言, 更容易学习和掌握。
- 4) Python 语言有丰富的数值分析库、统计库和数据可视化库。
- 5) 作为开源语言, Python 网络资源非常丰富, 在线帮助文件也非常多。

基于以上原因, 国内外很多地学科科研机构及研究人员都首选 Python 作为软件开发平台或开发语言。如 GEOS-Chem 技术支持团队基于 Python 开发了 GCPy 工具, 并替代以前基于 IDL (Interactive Data Language) 开发的应用 GAMAP [4]。而且有不少基于 Python 语言开发的第三方库也可以免费获取。比较常用的地学 Python 库有: Geopandas、EarthPy 库、eo-learn、PVGeo [5]。为了普及与交流 Python 的地学应用, 还有专门的国际学术会议。如专门面向地学领域 Python 应用的国际学术会议 GeoPython (<http://www.geopython.net>), 从 2016 年开始举办, 参会人员逐年增加, 2018 年有来自 30 个国家的参会代表。

3. 国内外高校地学相关专业 Python 语言应用情况

地学数据具有海量、非结构化以及空间位置的相关性、多尺度性等显著的特点。与传统多元统计方法中研究的数据特征有很大的区别。Python 语言出现的时间并不长, 广泛应用也是近 10 年才出现。由于 Python 语言在大数据处理、人工智能、网络数据获取等方面的广泛应用, 为了适应 Python 快速发展的要求, 很多高校或研究机构都开设了相关课程或速成班。国内有关 Python 教学的文章主要有两类, 一类是 Python 本身的教学方法研究, 另一类是 Python 辅助教学方法研究。图 1 是知网公开发表的论文中以 Python+教学作为关键词检索得到的文章数。由图 1 可知, 近 3 年相关文章快速增加。但是这些文章有关 Python 语言辅助地学类专业教学的论文鲜有报道。高校的课程设置方面, 主要还是以 Python 语言编程方法为主。把 Python 语言用于辅助地学教学目前还报道得较少, 用于辅助地质教学更是鲜有报道。

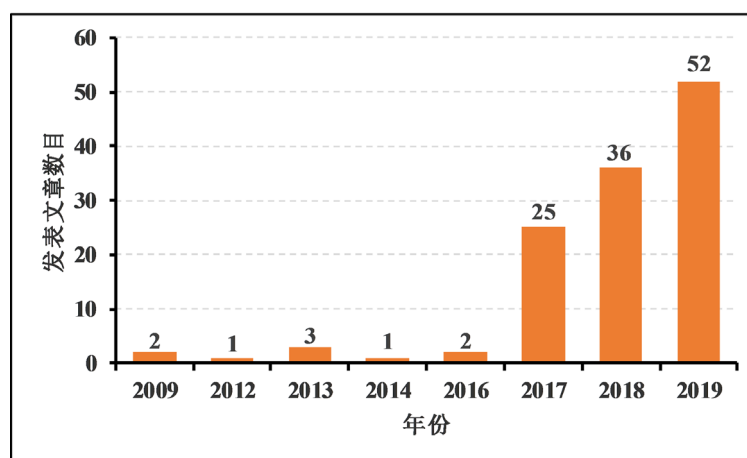


Figure 1. Number of articles on Python teaching collected by CNKI in recent 10 years

图 1. 近 10 年知网发表的有关 Python 教学方面的文章数目

国外 Python 语言的应用比国内早。把 Python 用于地学数据分析也同样比较早。同时, 为了满足地学数据大数据分析与人工智能处理的需要, 国外许多高校的地学类专业纷纷开设 Python 地学数据分析方面的课程。表 2 列出了部分国外高校地学相关专业开设 Python 语言用于地学数据分析方面的课程。这些课程不是纯粹的 Python 语言编程课程, 而是 Python 语言如何用于地学数据处理。部分课程名称尽管没有包含“Python”, 但应用平台仍然是以 Python 为主。

Table 2. Python courses open to students of geoscience major from some universities

表 2. 国外地学类相关专业开设与 Python 语言相关课程情况

序号	学校	课程名称
1	University de Montpellier	Introduction to basic data analysis and statistics in Earth sciences (using python 3)
2	University of Maryland	Python for Remote Sensing Applications in Earth Science
3	University of Edinburgh	Python for Earth Science Modelling
4	University of California San Diego	Python Programming for Earth Science Students
5	University of California San Diego	Introduction to Computational Earth Science
6	The University of Melbourne	Python for Earth Sciences
7	University of Colorado Boulder	Python Programming Tutorials for Earth Data Science
8	University of Alberta	Python for geoscientists

Continued

9	The University of Iowa	Scientific Computing with Python
10	Royal School of Mine, Imperial College	Introduction to programming for geoscientists
11	Rice University	Machine Learning in Geoscience
12	The University of Wyoming	Introduction to GeoModeling
13	Brown University	Mathematical and Computational Earth Sciences
14	University of Texas at Austin	Python for Geoscience Research
15	Columbia University	Research Computing in Earth Sciences
16	The University of Oklahoma	Machine Learning for Petroleum Engineers, Engineers and Geoscientists
17	Texas A&M University	Python for Geosciences
18	Utrecht University	Python in Earth Science

4. Python 辅助数学地质教学内容设计

Python 用于数学地质教学中, 如何设计教学内容, 是需要考虑的重点。首先, 目的要明确。在数学地质教学中引入 Python 语言, 其目的有两个, 一是通过引入 Python 语言, 让学生对 Python 语言有基本的了解, 引起学生的兴趣。其次, 对教学过程中涉及的算法及结果运用 Python 进行可视化, 让学生直观的了解算法的实质, 同时也激发学生学习 Python 的兴趣。其次分清主次。由于 Python 主要是用于辅助教学, Python 语言应该是从属的, 因此 Python 的编程方法不能作为主要内容。基于以上考虑, Python 用于数学地质教学主要从以下几方面入手。

4.1. Python 安装与开发环境的选择

Python 语言作为一门开源语言, 有各种各样的安装包, 让初学者难以选择, 同时, 学习运用 Python 语言, 离不开代码编辑器或集成开发环境(IDE, Integrated Development Environment)。因此, 首先需要给学生大致介绍 Python 语言的安装方法及 IDE。根据作者 Python 编程体会, 初学者采用 Anaconda 安装包是一种高效、快速的方法。关于 Python 的编程环境, 目前到底有多少 Python IDE, 并没有一个明确的统计, 而且新的 IDE 还在不断涌现。但最常用的 IDE 还是有共识的。据统计, 目前最常用的 Python 开发环境主要有: Jupyter、PyCharm、Spyder、visual studio code、IDLE、Sublime Text 等[6]。而 Anaconda 安装包集成了 Jupyter 和 Spyder 两个开发环境。基于以上考虑, 在 Python 教学中采用了 Anaconda 以及提供的 Spyder 开发环境。

4.2. Python 常用库介绍

丰富的开源函数库是 Python 的一大亮点, 也是 Python 语言近年来快速成为重要的开发语言的一个主要原因。Python 库包括两类: 标准库和第三方库, 两种库的调用方式是一样的。我们讲的 Python 语言丰富的函数库, 实际暗指的就是第三方库, 即非 Python 语言自带的库, 需要下载安装方能使用。由于 Anaconda 安装包集成了多达 1500 个库, 这些库包括了数学地质教学内容所涉及的所有函数, 因此安装好 Anaconda 后, 一般情况, 不需要再安装其它库, 这也是推荐学生采用 Anaconda 包的一个重要原因。尽管库非常多, 但最常用, 最重要, 学生必须掌握的库主要有以下几个: Numpy、Scipy、Pandas、Matplotlib、Scikit-learn。其中 Numpy、Scipy、Pandas 为科学计算与数据分析软件包, Matplotlib 为图形可视化包。Numpy、Scipy、Pandas、Matplotlib 4 个函数库构成了 Python 科学计算与可视化的支柱。而 Scikit-learn 函数库则是建立在 Numpy、Scipy、Matplotlib 上的机器学习工具, 提供了非常丰富的机器学习算法[7]。

包括了常用的监督学习与非监督学习常用算法: 线性模型(普通最小二乘法、岭回归、Lasso 回归模型)、线性和二次判别分析、支持向量机分类、贝叶斯分类、最近邻分类、决策树分类、神经网络分类、随机森林分类(Random forest)、梯度树提升算法(Gradient Tree Boosting)、自适应增强算法(Adaboost), 层次聚类、K-means 聚类、基于密度的聚类(DBSCAN)、谱聚类(Spectral Clustering)等。显然, 这些机器学习算法比数学地质教学内容要丰富得多。教学过程中应该结合具体的方法, 通过示例, 简要介绍这些库的使用方法。

4.3. 基于 Python 进行计算结果的可视化

结果的可视化, 指的是运用 Python 编程展示一些无法或者难以通过 Excel、SPSS、Minitab 软件展示的图件。在教学过程中可视化的图件主要有以下几类: 曲线的填充(如概率分布、假设检验拒绝域等图件中部分区域的填充)、散点图矩阵(多变量中两两变量散点图的矩阵排列)、相关系数热力图(图 2)、不同线性判别函数对比、K-Means 聚类过程(图 3)、样本的高维映射(利用核函数把二维变量映射到三维特征空间)等。

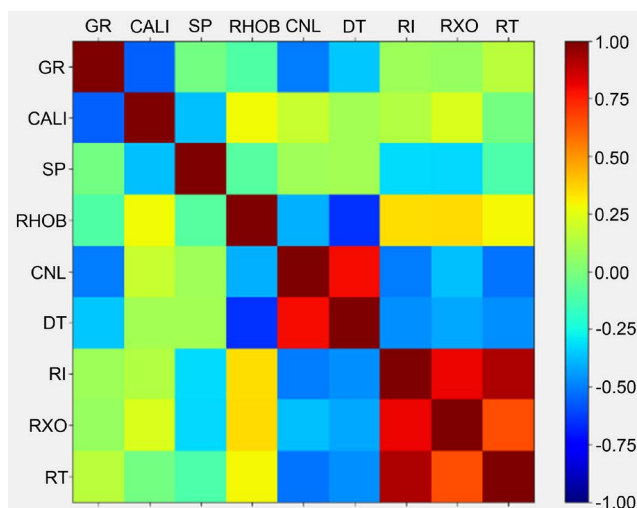
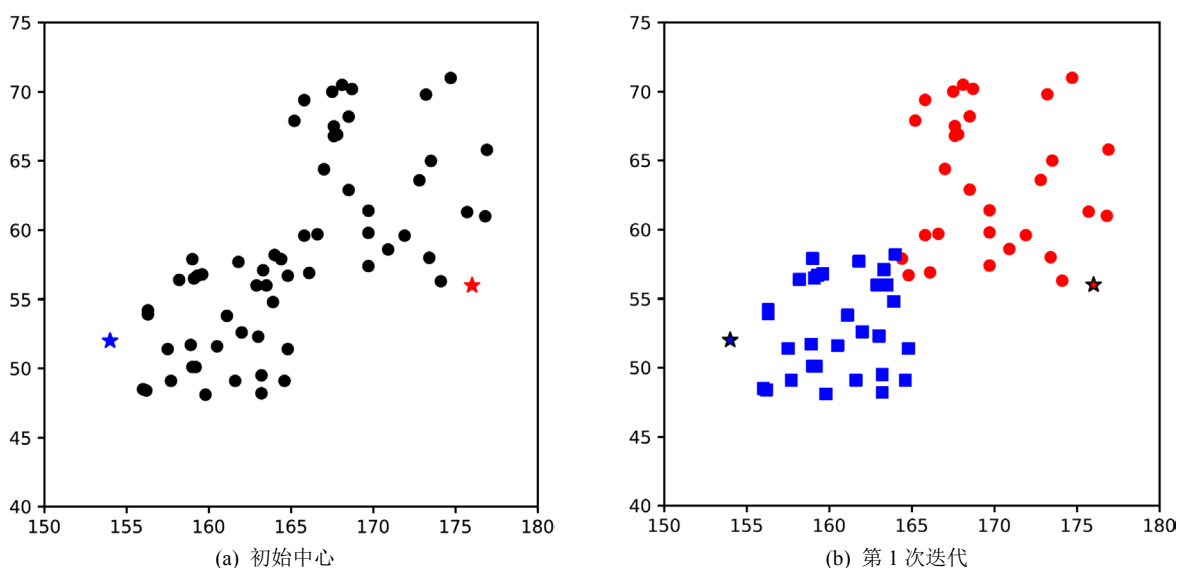


Figure 2. Heatmap of correlation matrix

图 2. 相关系数矩阵热力图



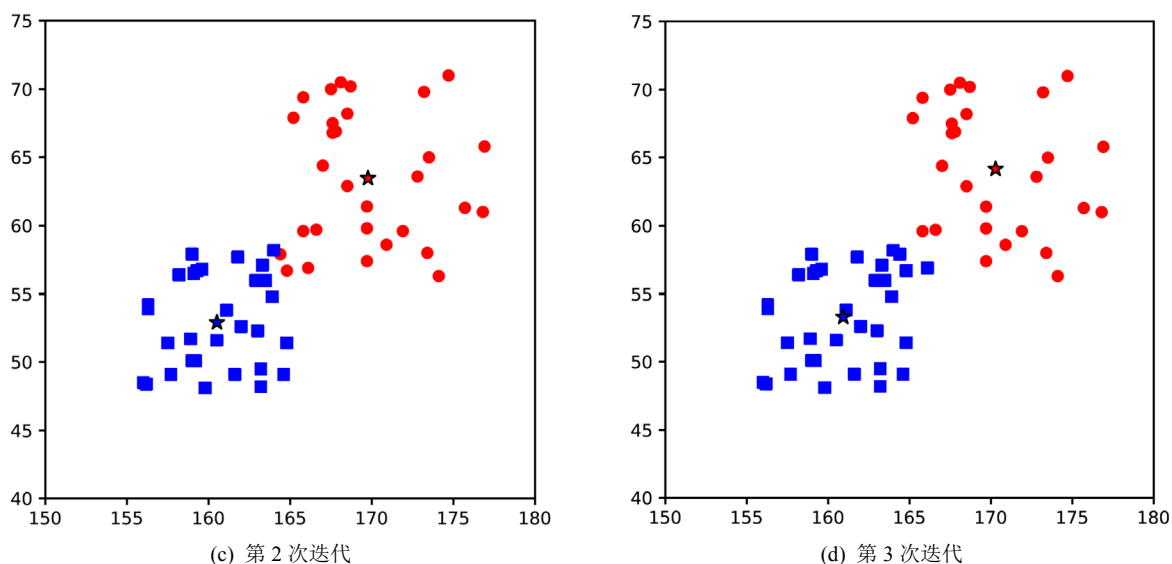


Figure 3. The change of clustering center in K-means clustering
图 3. K-means 聚类过程中聚类中心的变化

5. 结语

本文结合作者引入 Python 辅助数学地质教学的体会, 作了简单的总结。由于是首次把 Python 语言用于数学地质教学, 因此对于教学内容的设计以及教学过程的把控还存在不少问题, 需要在以后的教学过程中不断完善。

参考文献

- [1] Fowler, A. (2011) *Mathematical Geoscience*. Vol. 36. Springer, London.
- [2] Kempler, S. and Mathews, T. (2017) Earth Science Data Analytics: Definitions, Techniques and Skills. *Data Science Journal*, **16**, 1-8. <https://doi.org/10.5334/dsj-2017-006>
- [3] Lin, J.W.B. (2012) Why Python Is the Next Wave in Earth Sciences Computing. *Bulletin of the American Meteorological Society*, **93**, 1823-1824. <https://doi.org/10.1175/BAMS-D-12-00148.1>
- [4] Python Tools for Use with GEOS-Chem. <http://wiki.seas.harvard.edu/geos-chem/index.php/>
- [5] Sullivan, C. and Trainor-Guitton, W. (2019) PVGeo: An Open-Source Python Package for Geoscientific Visualization in VTK and ParaView. *Journal of Open Source Software*, **4**, 1451. <https://doi.org/10.21105/joss.01451>
- [6] Gregory, P. Here Are the Most Popular Python Ides/Editor. <https://www.kdnuggets.com/2018/12/most-popular-python-ide-editor.html>
- [7] Julian, A. (2017) *Scikit-Learn Cookbook: Over 80 Recipes for Machine Learning in Python with Scikit-Learn*. Packt Publishing.