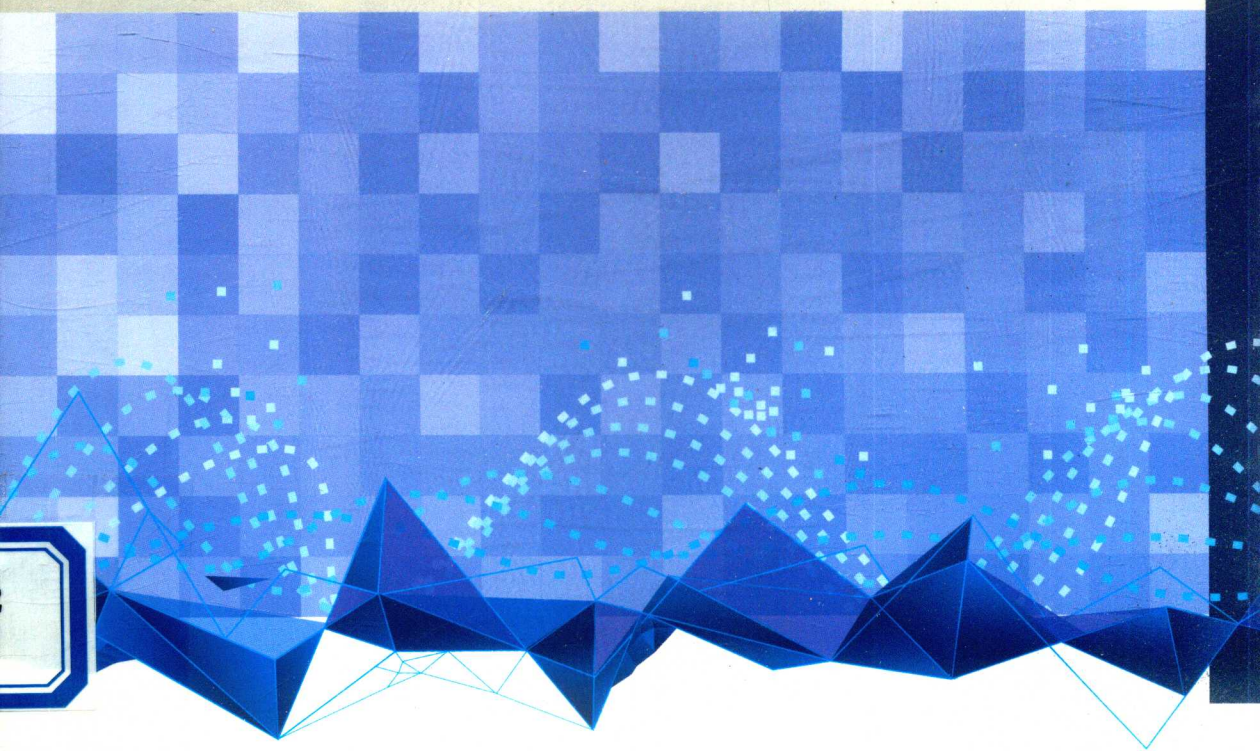


Python与开源GIS

——数据处理、空间分析与地图制图

卜坤 著



 科学出版社

Python 与开源 GIS

——数据处理、空间分析与地图制图

卜 坤 著



科学出版社

北京

内 容 简 介

本书从应用开发角度,根据作者多年的工作经验,介绍 Python 语言在开源 GIS 中的应用。希望能够借此机会,使得开源 GIS 得到应用,并进一步推广开源 GIS 的理念与技术。本书主要以空间数据的处理、分析以及地图制图为主线。在选择内容时,本书以目前最为经典、常用的类库为主,目的是给初学者系统地讲解基本的概念。书中用到一些数据,并有代码,这些资源都可以从网站上下载,并且网站上的内容也会有相应的更新。书中代码经过了测试,可以在 Linux 操作系统中运行,大部分也可以在 Windows 操作系统中运行。

本书适合地理信息等相关专业的学生、研究人员和开发人员阅读与参考。

图书在版编目(CIP)数据

Python 与开源 GIS : 数据处理、空间分析与地图制图/卜坤著. —北京: 科学出版社, 2019.11

ISBN 978-7-03-062927-2

I. ①P… II. ①卜… III. ①地理信息系统 IV. ①P208.2

中国版本图书馆 CIP 数据核字(2019) 第 241937 号

责任编辑: 陈 静 董素芹 / 责任校对: 严 娜

责任印制: 吴兆东 / 封面设计: 迷底书装

科 学 出 版 社 出 版

北京东黄城根北街 16 号

邮政编码: 100717

<http://www.sciencep.com>

北京中石油彩色印刷有限责任公司 印刷

科学出版社发行 各地新华书店经销

*

2019 年 11 月第 一 版 开本: 720 × 1 000 1/16

2020 年 1 月第二次印刷 印张: 22 1/2 插页: 1

字数: 500 000

定价: 108.00 元

(如有印装质量问题, 我社负责调换)

序

“夫地形者，兵之助也。”这是《孙子兵法·地形篇》中的一句话。自古以来，地理信息就是一国重要的信息资源。如何处理这些信息资源，需要科学技术的支持。从早期的计里画方制图，到后来的手扶跟踪数字化处理，再到后来的地理信息系统（GIS），都是在解决地理信息的数据处理问题。大数据时代的到来，既提供了海量地理信息数据获取和使用的机会，又提出了提高技术处理能力的需求。

引进与使用开源 GIS 软件是一种可以快速实现地理信息数据处理技术突破的开放、有效方法。而且开源软件的开发没有商业公司的生存压力，在很多方面有先进的理论与实现，非常有利于学生与技术人员掌握与使用。

卜坤博士多年来致力于推广开源 GIS 应用，开展了许多实践研究和应用。在此过程中，我也与他有一些实际合作，学到许多知识。据我的一些了解，他先后参加和承担了国际科联世界数据系统（ICSU-WDS）中国中心门户和 WDS 可再生资源与环境世界数据中心网络平台建设，联合国教育、科学及文化组织（UNESCO）国际工程科技知识中心防灾减灾知识服务系统网站平台建设，中国科学院大数据驱动的资源学科领域创新示范平台网站建设，一带一路国际科学家联盟平台网站建设等。这些平台建设和应用集中体现了开源 GIS 在数据管理、处理、分析、可视化等方面的技术。

除此之外，卜坤博士还积极参加许多相关的公益性工作，包括 OSGeo 中国中心网站维护、开源 GIS 文档翻译和编写等，为国内开源 GIS 社区的发展做出自己的努力和贡献。

本书正是结合他这些年的实践开发工作而著，内容涉及开源 GIS 应用、WebGIS 开发、地理信息科学数据共享等。非常值得一读！也值得操作实践！特此给大家推荐以为序。

王卷乐

中国科学院地理科学与资源研究所 研究员
世界数据系统可再生资源与环境数据中心 主任

前 言

在知识与经济全球化的时代，地理空间信息是现代社会的战略性信息资源，地理空间信息产业已成为现代知识经济的重要组成部分。

因此，充分利用国际开源地理空间信息技术与资源，从底层入手，面向行业应用需求，则有可能实现我国地理信息系统（geographic information system, GIS）技术的跨越发展，突破核心关键技术的封锁，推进我国地理空间信息产业的新发展。

开源 GIS 的发展较早，现在技术体系也已经比较完善，在数据处理、制图、Web 应用中都有所发展，在国外的学校、科研机构以及商业中都有应用。但是在国内，由于宣传力度不够，以及国人版权意识淡薄的原因，从学校到企业，对开源 GIS 的了解都相对较少。国际开源地理空间基金会（Open Source Geospatial foundation, OSGeo）中国中心作为国内开源 GIS 的推广组织，现在也只是由几名技术爱好者在推动。

GIS 业界已逐渐认识到数据采集和生产是建立 GIS 的一项最大的投资。从国内的现状来看，GIS 的数据处理还是采用人工处理方式，但是在处理过程中使用编程方式已经越来越普遍了。

还有一些内容要说明。

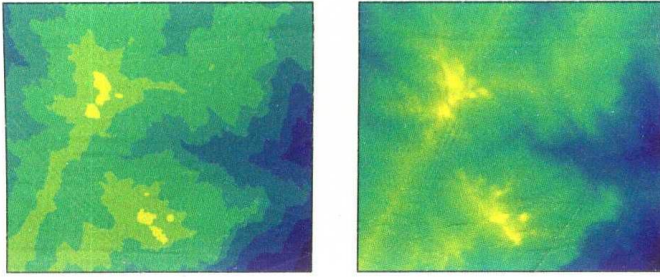
（1）本书使用 Latex 与 Python 3 两种语言写就，其中 Python 脚本是从 Latex 代码中提取出来的，这样保证书中内容与可运行程序的一致性。

（2）本书的源代码中没有任何图片，书中的图片全部是从代码中生成的。

（3）本书的所有代码使用单元测试进行维护，以保证代码的正常运行。测试环境为 Debian 9、Ubuntu 18.04 与 Debian 10（按发布时间排序）。本书相关资料详见 <https://www.osgeo.cn/pygis>。

感谢中国工程科技知识中心建设项目（CKCEST-2016-3-7、CKCEST-2017-3-1、CKCEST-2018-2-8、CKCEST-2019-3-8）、中国科学院“十三五”信息化专项科学大数据工程项目（XXH13505-07）等支持。

彩 图



(a) 子图p1

(b) 子图p2

图 8.20 填充高程地图

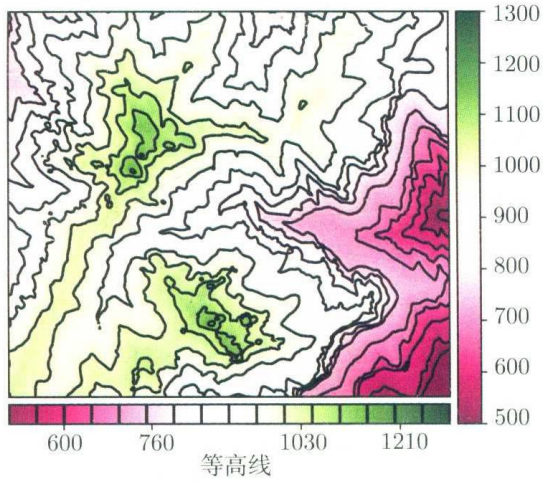


图 8.21 使用颜色条图例

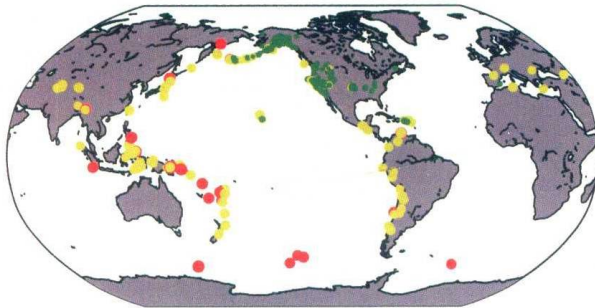


图 8.24 地震震级分类空间位置图

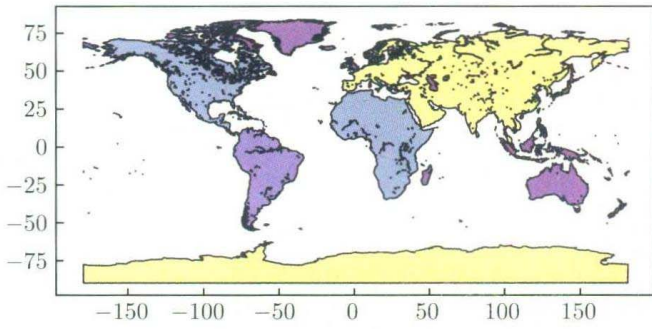


图 9.6 多度分色地图

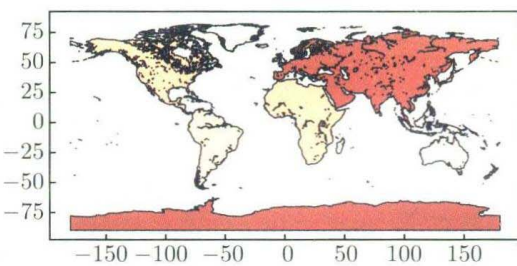


图 9.7 多度分色地图 (浅色)

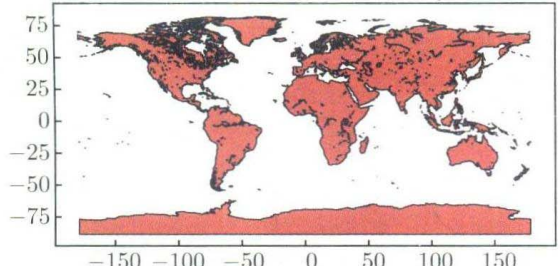


图 9.8 多度分位数分色地图

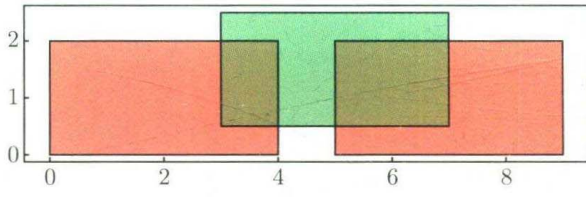


图 9.10 GeoPandas 中叠加操作的数据

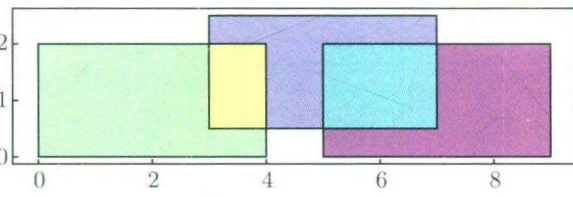


图 9.11 GeoPandas 中 union 操作的结果

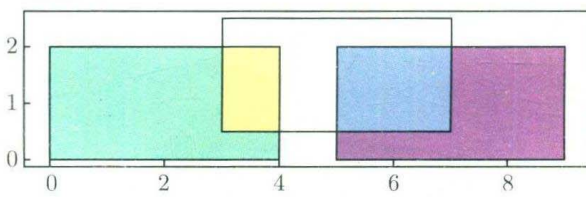


图 9.15 GeoPandas 中 identity 操作的结果

目 录

序	
前言	
第 1 章 引言	1
1.1 GIS 与开源 GIS 的基本概念	1
1.1.1 GIS 的概念	1
1.1.2 位置、距离、度量与比例尺	4
1.1.3 地图投影	7
1.1.4 空间分析的基本概念	11
1.1.5 开源 GIS 的概念	13
1.1.6 开源空间信息软件体系与技术概述	16
1.2 GIS 中的数据结构与数据类型	17
1.2.1 空间数据模型与数据格式	17
1.2.2 GeoTIFF 文件格式与颜色空间	19
1.2.3 常用矢量数据格式与文件格式介绍	20
1.3 软件安装与环境配置	22
1.3.1 本书介绍的开源 GIS	23
1.3.2 Debian Linux 的安装与配置	23
1.3.3 虚拟机的使用	25
1.3.4 编辑器与 IDE	26
1.4 Python 语言基本用法	27
1.4.1 Python 基础用法	28
1.4.2 Python 的基本语句代码结构与数据类型	30
1.4.3 流程控制	33
1.4.4 Python 中的列表、元组与字典数据结构	36
1.5 本书的约定与注意事项	40
第 2 章 使用 GDAL 操作栅格数据	42
2.1 GDAL 简介	42
2.1.1 GDAL 库简介	43
2.1.2 GDAL 数据模型	44
2.2 使用 GDAL 获取栅格数据集信息	48

2.2.1	开始使用 GDAL	49
2.2.2	读取遥感影像的信息	51
2.2.3	使用 GDAL 获取栅格数据波段信息	54
2.3	访问栅格数据中的像元	55
2.3.1	GDAL 中的栅格数据类型	55
2.3.2	访问数据集的数据	56
2.3.3	读取波段中的数据	58
2.4	创建与保存栅格数据集	59
2.4.1	使用 CreateCopy 方法创建影像	60
2.4.2	使用 Create 方法创建影像	61
2.4.3	创建多波段影像	62
2.4.4	GDAL 写操作时的空间投影处理	63
2.4.5	建立影像金字塔	64
2.5	GDAL 的其他问题	64
2.5.1	GDAL 和 Pillow 的互操作	64
2.5.2	GDAL 工具集介绍	67
2.5.3	访问索引图像中的数据	73
2.5.4	地图代数计算	77
2.5.5	GDAL 中使用仿射地理变换进行空间定位	78
第 3 章	使用 OGR 库操作矢量数据	81
3.1	OGR 简介	81
3.1.1	OGR 的命令行工具	82
3.1.2	在 Python 中的 OGR 基本类	83
3.2	使用 OGR 获取 Shapefile 信息	83
3.2.1	导入 OGR 库	83
3.2.2	读取矢量数据	84
3.2.3	获取图层信息	85
3.2.4	获取要素信息	87
3.2.5	矢量数据的空间参考	90
3.3	使用 OGR 创建矢量数据	91
3.3.1	使用 OGR 创建 Shapefile	91
3.3.2	使用 OGR 创建要素几何形状	92
3.3.3	使用 WKT 创建数据集的几何形状	96
3.3.4	使用 OGR 复制方法创建新的 Shapefile	99
3.3.5	OGR 属性字段的定义与使用	100

3.3.6	OGR 中投影的处理方法	102
3.4	根据条件选择数据	104
3.4.1	根据属性条件选择与生成要素	104
3.4.2	空间择舍器	106
3.4.3	在 OGR 中使用 SQL 语句进行查询	108
3.5	使用 Fiona 进行矢量数据读写与处理	109
3.5.1	Fiona 简介	110
3.5.2	读取矢量数据	111
3.5.3	Fiona 的数据模型	114
第 4 章	空间参考与坐标转换	118
4.1	空间参考与坐标转换原理	118
4.1.1	大地水准面、地球椭球体与基准面	118
4.1.2	PROJ.4、osgeo.osr 模块及投影表示方法简介	120
4.1.3	在 PROJ.4 中了解椭球体与基准面	121
4.2	PROJ.4 命令行工具的使用	124
4.2.1	proj 命令的用法	124
4.2.2	地图投影设置	127
4.2.3	cs2cs 程序的用法	130
4.2.4	geod 程序的用法	131
4.3	在 Python 中使用 PROJ.4 的功能	133
4.3.1	Proj 类	134
4.3.2	投影变换	136
4.3.3	Geod 类的使用	137
4.4	osgeo.osr 的使用方法	139
4.4.1	osr 简介	139
4.4.2	定义空间参考系统	139
4.4.3	空间参考对象的使用	142
4.4.4	从文件中获取投影信息	143
4.4.5	不同坐标系统之间转换坐标	145
第 5 章	矢量数据的空间分析：使用 Shapely	147
5.1	Shapely 介绍	147
5.1.1	JTS、GEOS 与 Shapely	148
5.1.2	Shapely 中的空间数据模型	148
5.1.3	Shapely 的基本使用方法	150
5.2	Shapely 中的几何对象	151

5.2.1	通用属性与方法	151
5.2.2	Shapely 中的几何要素	152
5.2.3	Shapely 中的几何集合	156
5.3	Shapely 中谓词与关系	160
5.3.1	一元谓词	160
5.3.2	二元谓词	162
5.3.3	DE 九交模型关系	164
5.4	使用 Shapely 空间分析方法构建新对象	165
5.4.1	基于集合论方法构建新的几何对象	166
5.4.2	构建新对象的方法	168
5.5	Shapely 中其他操作	173
5.5.1	合并线状要素	173
5.5.2	级联合并	175
5.5.3	制备几何操作	175
5.6	Shapely 互操作的接口与使用	176
5.6.1	WKT 格式	176
5.6.2	NumPy 与 Python 列表	177
5.6.3	<code>__geo_interface__</code> 接口与 Shapely 中的实现	178
第 6 章	使用 SpatiaLite 空间数据库	181
6.1	开源空间数据库的概念	182
6.1.1	SQLite 与 SpatiaLite 介绍	182
6.1.2	安装与基本使用	184
6.2	在命令行中使用 SpatiaLite Shell	185
6.2.1	开始运行 SpatiaLite 命令行	185
6.2.2	SpatiaLite 中的基本 SQL 数据库查询用法	188
6.2.3	导出 GIS 数据	190
6.2.4	创建 SpatiaLite 数据库	191
6.2.5	SpatiaLite 中管理空间表	193
6.3	空间索引的概念与使用	196
6.3.1	空间索引的概念	196
6.3.2	空间索引: 在 SQLite 中使用 R-Tree	197
6.3.3	空间索引: 使用 MbrCache	201
6.4	在 Python 中使用 SpatiaLite 进行数据管理	203
6.4.1	在 Python 中使用 SpatiaLite	203
6.4.2	导入 Shapefile	206

6.4.3	在表中进行空间查询查找	207
6.5	SpatiaLite 几何类型的定义与使用	208
6.5.1	熟悉 Geometry	208
6.5.2	几何要素	210
6.5.3	最小外包矩形 (MBR)	214
6.5.4	创建与更新数据表	214
6.5.5	Python 中 SQLite 的事务操作	217
6.5.6	管理坐标参考与坐标转换	218
6.6	使用虚拟表链接其他格式数据执行 SQL 语句	221
6.6.1	在 CSV 与带分隔符的 TXT 文件上进行 SQL 查询	221
6.6.2	在 Shapefile 上执行 SQL 查询	223
6.7	SpatiaLite 中空间关系比较与空间运算	225
6.7.1	评价 MBR 关系	226
6.7.2	评价几何对象之间的关系	228
6.7.3	几何对象之间空间运算	231
第 7 章	GIS 制图: 使用 Mapnik 进行地图制图	232
7.1	Mapnik 地图制图基本概念与技术框架	233
7.1.1	Mapnik 简介	233
7.1.2	地图制图的背景知识	233
7.1.3	Mapnik 制图的工作流程	235
7.1.4	使用 XML 渲染地图	239
7.2	Mapnik 制图的技术细节	240
7.2.1	地图和图层	240
7.2.2	地图及图层对象的属性和方法	242
7.2.3	地图晕渲与保存	244
7.2.4	渲染结果输出	246
7.2.5	Mapnik 制图中的地图投影	247
7.3	数据的读取与设置	248
7.3.1	数据源	248
7.3.2	常用数据格式的读取	248
7.3.3	使用 GDAL/OGR 包进行数据的读取	249
7.3.4	使用文本数据源	250
7.4	绘制不同的要素	251
7.4.1	绘制栅格影像	251
7.4.2	绘制线	252

7.4.3	绘制多边形	257
7.4.4	绘制标注	259
7.4.5	绘制点状要素	264
7.5	数据显示的规则	267
7.5.1	数据显示的规则	267
7.5.2	按比例尺显示要素	270
第 8 章	使用 Basemap 进行地图可视化	272
8.1	Basemap 简介与基本使用方法	272
8.1.1	简介	272
8.1.2	Basemap 使用简介	273
8.1.3	设置地图投影	275
8.1.4	绘制地图背景	279
8.1.5	控制制图中的数据细节	282
8.2	在 Basemap 中添加 Matplotlib 绘制功能	284
8.2.1	使用 annotate 方法绘制标注	284
8.2.2	使用 plot 函数绘图	286
8.2.3	使用 text 方法绘制文本	288
8.2.4	混合使用 plot 方法与 text 方法	289
8.3	在 Basemap 中使用 GIS 数据	291
8.3.1	使用 Shapefile	291
8.3.2	在 Basemap 中绘制 DEM 数据等高线	292
8.3.3	在 Basemap 中使用颜色渲染 DEM 数据	294
8.4	USGS 地震数据可视化实例	297
8.4.1	全球地震数据集读取	297
8.4.2	绘制地震数据集	298
第 9 章	Python 下面其他开源 GIS 库使用	302
9.1	使用 PyShp 读写 Shapefile	302
9.1.1	PyShp 的介绍与安装	302
9.1.2	读取 Shapefile	303
9.1.3	创建 Shapefile	309
9.2	使用 geojson 库处理 GeoJSON 数据	313
9.2.1	geojson 模块的安装	313
9.2.2	geojson 中的几何对象、要素与要素集合	313
9.2.3	geojson 中的方法	317
9.3	使用 Descartes 进行绘图	318

9.3.1	Descartes 的安装与使用	319
9.3.2	使用 Descartes 绘制 Shapefile 的实例	320
9.4	数据分析与可视化库 GeoPandas 的基本用法	322
9.4.1	数据结构: GeoSeries	322
9.4.2	数据结构: GeoDataFrame	323
9.4.3	地图工具	325
9.4.4	几何图形的叠加	328
9.4.5	几何操作	331
9.4.6	管理投影	334
9.5	使用 Folium 进行 WebGIS 应用	336
9.5.1	Folium 的基本用法	337
9.5.2	在 Folium 中添加用户数据	340

彩图

第 1 章 引 言

本章先来介绍本书的一些概念与理论基础，以及基本的实验环境配置。大概涵盖以下几个方面。

- (1) 地理信息系统 (geographic information system, GIS) 概念的介绍。
- (2) 开源 GIS 的概念与技术。
- (3) 本书实验环境软件的安装与配置。
- (4) Python 语言的简单介绍。
- (5) 本书的一些约定与注意事项。

1.1 GIS 与开源 GIS 的基本概念

GIS 帮助人类了解政治、经济、景观等方面的变化。在 20 世纪 80 年代之前，GIS 主要面向专业机构，如国家机构和城市机构，但在今天，它无处不在，这得益于准确且无任何费用的全球定位系统 (global positioning system, GPS)、卫星图像商品化与开放源代码软件等。

GIS 的出现与发展，得益于多种学科理论的支撑以及技术的发展，从而导致不同学科的人员对 GIS 有不同的理解，且 GIS 作为一门相对较新的学科，一些概念与理论尚未形成定论，再加上 GIS 本身的快速发展，其概念的内涵与外延也在不断进化，因此，给出大家都认可的 GIS 定义还是比较困难的。

本书是从 Python 开发的角度来分析的，认为 GIS 是对地理相关数据进行处理、分析与应用的计算机技术。这一理解有其偏颇之处，但是对于本书的宗旨而言，则是比较恰当的。为了慎重起见，下面对 GIS 的基本概念进行一些阐述。阐述的内容会涉及 GIS 的核心概念，如距离、投影、数据模型、数据结构、地图投影、空间关系、空间分析等。

1.1.1 GIS 的概念

GIS 是一门综合性学科，其结合了地理学与地图学，已经广泛地应用在不同的领域，是用于输入、存储、查询、分析和显示地理数据的计算机系统，它可以把地图这种独特的视觉化效果和地理分析功能与一般的数据库操作（如查询和统计分析等）集成在一起。

GIS 与其他信息系统的最大区别, 在于 GIS 可以对空间信息进行存储、管理与分析, 从而使其广泛地应用于公众、个人和企事业单位中解释事件、预测结果、规划战略等, 具有实用价值。

在很大程度上, GIS 是一门数据驱动的学科。尤其对于工程应用来讲, 可能建设项目的绝大部分资金和时间投入都是在数据的加工与处理上。

地理空间数据是指那些直接、间接地与地理空间位置的分布、时间的发展有关的自然、经济和人文等方面的物体、事实、事件、现象和过程的描述。从专业角度而言, 地理空间数据可以理解为标识地球表面上自然或人为要素及边界地理位置和特性的信息, 包含了地理空间实体的空间特征和属性特征。这些数据可以通过各种方法获得, 如遥感、制图和测量等手段。

关于信息的定义也非常复杂, 有一门专门的学科称为“信息学”。关键在于理解数据与信息区别: 数据是从调查中获得的数量或质量的度量值, 而信息则是指数据经过记录、分类、组织、连接或翻译后出现的意义。

1. 对 GIS 概念的不同理解

由于 GIS 学科较新, 不同的人对其有不同的理解角度。一方面, GIS 是一门学科, 是描述、存储、分析和输出空间信息理论与方法的一门新兴的交叉学科; 另一方面, GIS 是一个技术系统, 是以地理空间数据库 (geospatial database) 为基础, 采用地理模型分析方法, 适时提供多种空间和动态的地理信息, 为地理研究和地理决策提供服务的计算机技术系统。

从学术观点来看, 人们对 GIS 有如下三种观点: 地图观点、数据库观点、空间分析观点。

地图观点: 持地图观点的人主要来自于景观学派和制图学派, 他们认为 GIS 是一个地图处理和显示系统。在该系统中, 每个数据集可以被看成一张地图或一个图层 (layer), 也可以被看成一个专题 (theme) 或覆盖 (coverage)。利用 GIS 的相关功能对数据集进行操作和运算, 就可以得到新的地图。

数据库观点: 持数据库观点的人主要来自于计算机学派, 他们强调数据库理论和技术方法对 GIS 设计、操作的重要性。

空间分析观点: 持空间分析观点的人主要来自于地理学派, 他们强调空间分析和模拟的重要性。实际上, GIS 的空间分析功能是它与计算机辅助设计 (computer aided design, CAD)、管理信息系统 (management information system, MIS) 等的主要区别之一, 也是 GIS 理论和技术方法发展的动力。

随着地理信息学科与技术的发展, GIS 的内涵也在不断发展。有人称 GIS 为“地理信息科学” (geographic information science), 近年来, 还有些专家称 GIS 为“地理信息服务” (geographic information service)。

2. GIS 的功能

从 GIS 的概念上,已经能够看到 GIS 具有哪些功能。下面对相关概念做进一步解释。

首先从数据处理、信息提取的角度来看, GIS 具有以下功能。

(1) 数据采集与编辑:包括图形与属性数据的采集、编辑和分析计算。

(2) 地理数据库管理:包括数据库定义、数据库的建立与维护、数据库操作、通信等。

(3) 地图制图:根据 GIS 的数据结构及绘图仪的类型,用户可获得矢量地图或栅格地图。

(4) 空间查询与空间分析:包括拓扑空间查询、缓冲区分析、叠加分析、拓扑分析等。

(5) 地形分析:包括数字高程模型的建立、坡度分析、流域提取等。

GIS 不仅可以为用户输出全要素地图,而且可以根据用户的需要分层输出各种专题地图,如行政区划图、土壤利用图、道路交通图、等高线图等。另外还可以通过空间分析得到一些特殊的地理学分析用图,如坡度图、坡向图、剖面图等。

从解决问题的角度来看, GIS 功能遍历数据采集、分析、决策应用的全部过程,并能回答和解决以下五类问题。

(1) 位置:在某个地方有什么。

(2) 条件:符合某些条件的实体在哪里。

(3) 趋势:某个地方发生某个事件及其随时间的变化过程。

(4) 模式:某个地方存在的空间实体的分布模式。

(5) 模拟:某个地方如果具备某种条件会发生什么。

3. GIS 的发展历史

20 世纪 60 年代早期,在核武器研究的推动下,计算机硬件的发展推动了通用计算机“绘图”的应用。在加拿大安大略省的渥太华,罗杰·汤姆林森(Roger Tomlinson)博士开始研发世界上第一个真正投入应用的 GIS,这个系统被称为加拿大地理信息系统(Canada geographic information system, CGIS),主要用于存储、分析和利用加拿大统计局收集的土地方面的数据,并增设了等级分类因素来进行分析。

由于 CGIS 设计并实现了现代 GIS 技术的一些基本功能,罗杰·汤姆林森被称为“地理信息系统之父”,尤其是因为他促进了地理数据的空间分析中对于图层这一概念与技术的应用。

20 世纪八九十年代产业成长刺激了 UNIX 工作站与个人计算机的飞速增长。硬件的发展使得像美国环境系统研究所公司(Environmental Systems Research In-