

基于 Python 的地理空间数据可视化探究

胡楠, 孟宪伟, 王彦超

(辽宁科技学院 曙光大数据学院, 辽宁 本溪 117004)

摘要: OpenStreetMap (OSM) 作为众源地理信息数据典型的代表, 其中涵盖了大量的开源地理信息数据, 为实现地理空间数据信息的挖掘、分析和可视化提供了可自由编辑的世界级地理数据。文章以 OSM 官网的地理信息数据为依据, 利用 Python 的地理信息数据处理技巧、地理空间数据可视化技术和图层叠加方法, 实现各类地理空间数据可视化和信息注记, 为国内的资源勘探、交通运输和疫情防控等领域提供了可靠的数据处理、可视化和分析方法。

关键词: 图层叠加; OSM; 可视化; 地理空间数据; 信息注记

中图分类号: TP391.1

文献标识码: A

1 OpenStreetMap 地理空间数据

1.1 数据下载

开放地图包含了丰富的地理数据, 其目标是向所有人提供地理数据使用及分享^[1], 用户只需在 OpenStreetMap 官网的首页“Geofabrik 下载”链接下载所需的世界各个国家地区的地理数据压缩文件。

1.2 数据描述

每个数据包的主文件为 ESRI(美国环境系统研究所公司)开发的一种空间数据开放格式数据文件, 地理数据一般比较通用的格式是 shp (Shapefile) 文件; 可以存储井、河流、湖泊等空间对象的几何位置, 主要包括描述空间对象对应的几何体的点、折线与多边形的空间数据。

2 Python 数据可视化

2.1 绘制多边形

1) 地理坐标系 (Geographic Coordinate System)

地理坐标系使用三维球面来定义地球表面的任何一个地标的位置, 并在地球表面的所有地标点应用经纬度坐标系来标注的方法, 在球面系统中, 水平线为纬线、垂直线为经线。

2) 多边形类型 (Polygon 和 MultiPolygon)

任何一个区域、国家和省份都对应一个多边形, 每个多边形是由一系列边线上的点组成的, 而点的坐标分别为该点在球面系统的纬线和经线值, 每一行为多边形边线上一个点的经纬度坐标。多边形又

分为 Polygon 和 MultiPolygon, 二者的区别如下:

(1) Polygon: 为单面封闭多边形, 且其中第一行数据与最后一行数据相同, 即第一个点与最后一个点重叠, 这样才能保证多边形的封闭。Shapefile 数据文件, 为我们绘制地理空间图形提供的数据源的数据格式如图 1(a) 所示, 该数据是某个行政区域的多边形的 shapely.geometry.polygon.Polygon 类的对象。

```
<class 'shapely.geometry.polygon.Polygon'>
POLYGON (((116.4263229378118 34.63999176825482,
116.426055982031 34.63815689086925,
116.4257888793946 34.63632282148443,
116.4263610839845 34.63288497924805,
116.4272079467776 34.63105010986322,
116.4288787841797 34.62945175170898,
116.4319534301758 34.62694168090815,
116.4350128173828 34.62535095214844,
116.4377899169922 34.62421798706066,
116.4257049560548 34.64985275268555,
...
116.4262771686446 34.64664459228527,
116.4260177612305 34.64411926269537,
116.4263229378118 34.63999176825482)))
```

(a) Polygon 数据格式

```
<class 'shapely.geometry.multipolygon.MultiPolygon'>
MULTIPOLYGON(((120.69484819205 27.8010213272119,
120.693914783313 27.8003668682123,
120.69332469733 27.7992725269342,
120.695438278033 27.798156727984,
120.697584045245 27.8003024951959,
120.69484819205 27.8010213272119),
((120.344448395 27.991393955,
120.34604707882828 27.98745719307442,
120.346404642361 27.9872565921189,
120.410475409 27.95913671,
...
120.385600285 27.980657895,
120.374984166 27.985694829,
120.368014522 27.984234814,
120.349883502 27.988421062,
120.344448395 27.991393955)))
```

(b) MultiPolygon 数据格式

图 1 封闭多边形

(2) MultiPolygon: MultiPolygon 为多面封闭多边形, MultiPolygon 类型数据特征是一个 MultiPolygon 可能存在多个 Polygon 单面多边形数据, 其数据格式如图 1(b) 所示, 第一个 Polygon 代表这个整体的 MultiPolygon 对象的整体外部轮廓, 而后续每个

收稿日期: 2021-11-28

项目来源: 2021 年校级一流课程“Python”; 2021 年校级教改项目“培养学生智能数据分析能力的线上线下混合式教学模式的研究与实践—基于 Python 课程群建设”; 2021 年产教融合项目“《可视化技术》课程体系改革”(202101016007)。

作者简介: 胡楠(1974-), 女, 辽宁铁岭人, 辽宁科技学院曙光大数据学院讲师, 硕士。

Polygon 对象是前一个 Polygon 对象的内部空洞或内部线条。

3) Python 绘制多边形

本文分为两种情况探讨绘制多边形方法。

(1) 如果绘制单面封闭多边形, 只需将 Polygon 对象转对应的每一行数据拆分成每个点的经纬度值, 并分别构造列表为 lats 和 lons, 使用 matplotlib.pyplot.plot(lats, lons) 函数来绘制一个封闭多边形, 如图 2(a) 所示。

(2) 如果是绘制多面封闭多边形, 先将 Multi-Polygon 对象转换为 Polygon 对象, 具体可以使用递归的方法, 依次将每个 Polygon 对象依次转换为两个列表 lats 和 lons, 再使用列表的 list.extend() 函数把后 Polygon 对象对应列表追加到前一个列表末尾, 构造出包含外部多边形和内部空洞多边形或内部线条总的经纬度列表 lats 和 lons, 然后再绘制多边形, 如图 2(b) 所示。

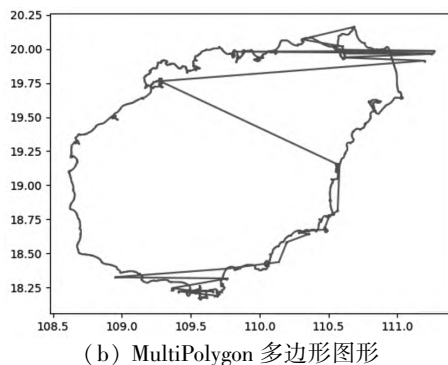
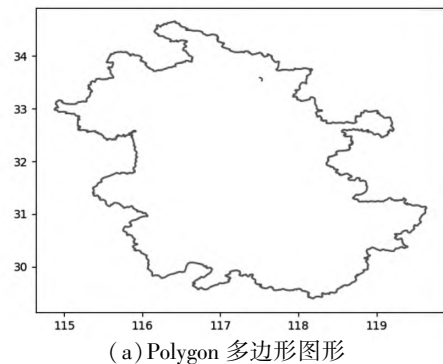


图2 绘制多边形

2.2 地理数据信息注记

1) 注记信息的目的和意义

随着地理信息系统 (Geographic Information System, GIS) 的广泛应用, 有效准确的地图信息标注是 GIS 的基本功能也是重要的研究问题, 更是备受社会各界关注的问题, 因为无论应用在哪个具体的行业, GIS 数据质量的好坏直接影响到所要建立的地理空间数据库所带来的经济效益和社会效益, 并直接影响到基于该空间地理数据信息的应用、分析、决策的准确性和可信度。

2) 注记信息的属性

注记一个区域文本信息的意义要远大于绘制多边形的边线, 由于地理空间数据可视化不仅包括每个封闭区域边线的绘制, 还需要每个区域文本信息的标注, 而确定信息注记位置也是实现地理空间数据可视化的关键点和重要组成部分, 注记标签中的文字是要素的某个字段属性, 且标记文本具有自己地理位置 (文本的注记经纬度) 和属性 (文本要素的显示字样), 注记文本应有自己的显示模式和风格, 但是地理空间数据对应注记文本的注记位置并没有标准。

3) 注记信息的位置

中心点是 2012 年公布的地理信息系统名词, 本文使用地理区域中心点来标注该区域文本信息^[2]。Python 的 geopandas.geodataframe.GeoDataFrame.representative_point(GeoDataFrame) 函数返回每个区域多边形的中心点; 每个多边形的中心点为 Python 的 geopandas.geoseries.GeoSeries 对象, 其数据格式为 Point 类对象, 即该点的经纬度值。

4) 带注记信息的地理区域绘图

实现封闭区域多边形绘制, 其核心理念是将绘图与数据分离, 且按图层叠加的方法进行作图。最底层是地理区域的地理空间数据; 然后依次叠加使用底层数据的边界绘图 (包括区域填充和线条颜色等样式设置); 再次叠加封闭区域内散点标注; 最后叠加对应区域文本信息的注记, 如图 3 所示。

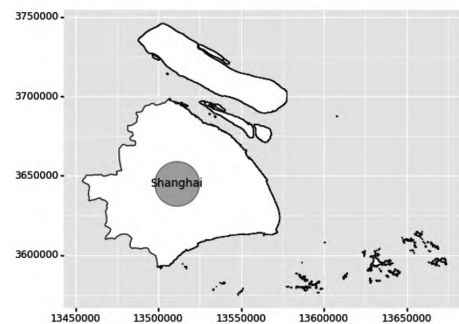


图3 带地理数据信息注记的区域图

3 基于 shapely + geopandas + plotnine 地理空间数据可视化设计

3.1 shapely

shapely 是用来表示地理空间数据的 Python 第三方库, 它的确定性空间分析是解决农业、交通、流行病、行政区域划分和许多其他领域相关问题的计算方法的重要组成部分。shapely 对应的数据类型的三要素为: 点 (Point)、线 (LineString) 和面 (Polygons)。每个平面中都包含这点、线和面组合的有限或无限的点关联。这三个要素的内部、边界和外部集是相互排斥的, 但是三者的并集与整个平面重合。

3.2 geopandas

geopandas 是用来处理地理空间数据的 Python 第三方库, geopandas 使得 Python 处理地理空间数据更为简洁和准确, geopandas 沿用了 pandas 的数据类型, Series 和 DataFrame, geopandas 中也有两种数据类型, GeoSeries 和 GeoDataFrame, 它们继承了 pandas 数据结构的大部分方法, 且结合了 pandas 和 shapely 的功能, 还提供了操作地理空间数据的高级接口, 使得在 Python 中进行 GIS 操作变成可能^[3]。

3.3 plotnine

plotnine 是用来实现地理空间数据可视化的 Python 第三方库, plotnine 的图形语法把绘图过程分为三部分: Plot(图) = data(数据集, 主要是 GeoDataFrame 数据) + Aesthetics(映射, 变量映射给 x , y 轴坐标, 或映射给大小、形状和颜色等图形属性) + Geometry(几何对象, 柱形图、直方图、散点图和线图等等), 且通过“+”对相关图形语法进行叠加, 且各图层之间的逐级靠“+”号叠加越后面图层的层级越高, 从而使得繁琐的绘图过程逻辑清晰。

3.4 地理空间数据可视化

Python 语言实现地理空间数据可视化的主要依据 shapely 的点、线和面的互斥和合并平面的原理, 结合 plotnine 的“+”号逐级叠加的方法, 实现地理空间数据可视化步骤如下:

(1) Python 使用 GeoPandas 读取地理空间数据文件(.shp), 并返回 GeoDataFrame 格式数据。

(2) 使用 plotnine 的 ggplot(df) 函数作为开始(ggplot 类生成一个图形数据集)。

(3) geom_map(aes(fill = 'ID_1')) 函数进行映射和绘制地图(aes() 函数实现“ID_1”字段到颜色或样式的映射)。

(4) geom_point(aes(x = 'lat', y = 'lon', shape = 'k', colour = 'k', size = 2, fill = 'r', alpha = 0.4)) 函数绘制点图(aes() 实现散点 x 和 y 轴位置到数据源字段的映射, 其他参数实现映射给大小、形状、颜色和透明度等图形属性)。

(5) geom_text(aes(x = 'lat', y = 'lon', label = 'NAME_1'), size = 5) 函数绘制文本。

(6) scale_fill_gradient2(low = 'red', mid = 'white', high = 'blue', midpoint = df.ID_1.mean()) 函数实现三色梯度, 顺序由低到高, 参数 low、mid 和 high 用于控制此梯度颜色, 参数 midpoint 设定中点值。

基于 OSM 的精准数据源, Python 的 geopandas、GeoDataFrame 的灵活数据格式, plotnine 的优质的图层叠加技术, 从而实现多层 baseplot 数据对象的可视化, 本文充分结合三者的优势, 对不同类型封闭区域数据进行格式的归一化和数据内容的整理, 从而实现任何一个地域的河流、公路交通、铁路交通、疫情防控等数据分布图的绘制和信息注记。

4 结语

地理信息系统空间数据的迅猛增长, 加之地理大数据自身的多元性和丰富性的特点, 地理信息大数据的应用已经逐步渗透到人们生活的各个方面。本文使用了丰富完备的地理空间数据, 依据完备的地理空间数据的处理与分析, 使用 Python 功能强大的图形拼接组合方法和地理空间数据可视化设计策略, 以单个地理空间区域为起点, 通过把每个区域的点线面的整合和合并, 设计了合理的地理空间数据的可视化方案。本文从地理大数据的特性出发, 强调地理空间大数据的发现规律和趋势预测的能力, 展示出地理信息可视化效果, 体现了地理信息标注的重要作用, 为各行各业的用户提供辅助决策支持。

参考文献:

- [1] 丁思磊. 基于 OpenStreetMap 数据的交通要素处理[J]. 测绘与空间地理信息, 2020(43): 55-56.
- [2] 刘秋雨. 基于具身认知理论的沉浸式地理空间数据可视化设计[J]. 工业设计, 2021(9): 26-27.
- [3] 朱懿. 大数据在 GIS 中的应用进展与展望[J]. 科技创新与应用, 2021, 11(32): 129-132.

Exploration of Geospatial Data Visualization Based on Python

HU Nan, MENG Xian-wei, WANG Yan-chao

(School of Sugon Big Data, Liaoning Institute of Science and Technology, Benxi Liaoning 117004, China)

Abstract: Open Street Map (OSM), a typical representative of multi-source geographic information data, covers a large amount of open source geographic information data, thus providing world-class freely editable geographic data for the mining, analysis and visualization of geospatial data. In this paper, various kinds of geospatial data visualization and information notes were realized based on geographic information data of OSM website, using the Python geographic information data processing techniques, the geographical spatial data visualization technology as well as layer stack method. It therefore provides a reliable method of data processing, visualization and analysis for domestic resources exploration, transportation, and epidemic prevention and control.

Key words: Layer stack; OSM; Visualization; Geospatial data; Information note