

文章编号: 1674-599X(2020)04-0107-06

# 基于 ArcGIS 和 FME 的兴趣点数据处理技术研究

刘帅<sup>1</sup>, 王佳<sup>2</sup>

(1. 南京市城市与交通规划设计研究院股份有限公司, 江苏 南京 210018;

2. 长沙理工大学 交通运输工程学院, 湖南 长沙 410114)

**摘 要:** 针对各类兴趣点抓取时区域过大而造成兴趣点数据丢失的问题, 本研究通过 FME 对研究区域进行格网划分, 采用 Geosharp 软件按照格网经纬度从高德地图中分区获取格网小区兴趣点, 并基于 ArcGIS 构建了分区导出模型和量化处理。研究表明: 格网坐标提取模型和分区导出模型能够较好地自动处理导出各类兴趣点数据, 提高数据抓取和处理效率, 比原有裁剪导出更节约时间, 能够为交通规划提供前期的基础数据支撑。

**关键词:** 大数据; 兴趣点; FME; 模型构建器

**中图分类号:** U491.1<sup>+</sup>2      **文献标志码:** A

**DOI:**10.16544/j.cnki.cn43-1494/u.2020.04.016

## Research on POI data processing technology based on ArcGIS and FME

LIU Shuai<sup>1</sup>, WANG Jia<sup>2</sup>

(1. Nanjing Institute of City & Transport Planning Co., Ltd., Nanjing 210018, China; 2. School of Traffic and Transportation Engineering, Changsha University of Science & Technology, Changsha 410114, China)

**Abstract:** In order to solve the data loss caused by the large amount of interest point data in the study area. In this study, the research area was divided into grids using the FME. The point of interest of grid cells was obtained from the grid map which was got from the Amap according to the grid longitude and latitude using the Geosharp software. Based on the ArcGIS, the partition export model was constructed, and the quantitative processing was carried out. The results show that the point of interest data can be automatically processed and exported using the grid coordinate extraction model and the zoning model. The efficiency of data picking and processing is improved, the handling time is saved effectively compared with the original cutting export. This study can provide early basic data to support traffic planning.

**Key words:** the big data; point of interest; FME; model builder

随着大数据的发展, 电子地图提供的各类兴趣点(point of interest, 简称为 POI)为交通规划设计提供了一种新的数据支撑。国外对于 POI 数据的获取和应用研究较早。Mummidi<sup>[1]</sup>等人通过分析用户在地图上注释的标签, 发现兴趣点, 增加了 POI 数据的获取途径。Xie<sup>[2]</sup>等人提出网络核密度分析, 将线要素引入核密度分析, 提高了核密度的运算效率。Lian<sup>[3]</sup>等人将 POI 签到信息划分权重, 通过用户的

签到频率增大 POI 的影响区域, 帮助用户推荐感兴趣的 POI 信息。李伟<sup>[4]</sup>等人通过 GIS 平台对 POI 数据进行处理, 对厦门岛内常规公交线路的可达性进行了评价。于浩川<sup>[5]</sup>等人以公交线路数据和 POI 数据为基础, 通过开发软件对城市公交线路进行规划。这些研究成果表明: 国外研究主要集中于 POI 数据分析, 而国内主要集中于 POI 应用, 但 POI 数据在交通规划应用中还有诸多缺陷。因此, 作者通

收稿日期: 2019-11-08

作者简介: 刘帅(1994-), 男, 南京市城市与交通规划设计研究院股份有限公司工程师。

过分析 POI 数据的获取和处理,拟构建研究区域格网坐标,提取模型和分区导出模型,提高 POI 数据获取的准确性和处理效率,以期为交通规划提供前期的基础数据。

## 1 POI 数据简介

### 1.1 POI 数据概念

POI 与人们的日常生活息息相关<sup>[6]</sup>,其含有地理信息的空间实体<sup>[7]</sup>,是可以通过计算机编程语言从电子地图检索和获取的位置大数据来获取。通常包含 ID、名称、经纬度、类别等相关信息,其中经纬度是 POI 数据的核心属性。

### 1.2 POI 数据的优势

POI 数据具有样本量大、获取成本低、更新速度快、位置精度高及涵盖信息详细等优势<sup>[8]</sup>。POI 的空间分布可以反映城市的空间结构。POI 数据越密集,城市功能越集中,则该区域可获得的机会越多。POI 的空间分布对交通规划设计和交通基础设施的布局有重要指导作用<sup>[9]</sup>。

### 1.3 POI 数据的主要类型

本研究采用高德地图 POI 三级分类标准,一级分类 23 种,二级分类 264 种,三级分类 869 种。通过分析各分类 POI 对交通规划影响程度的大小,基于居民使用频次和设施点数量,对电子地图 POI 各类设施进行筛选和分类,筛选出与交通规划相关程度较大的设施类型。从一级分类中,选出 12 个 POI 类型,见表 1。

表 1 溆浦县城区主要 POI 数量表

Table 1 Quantity table of main POI in Xupu County

| 序号 | POI 类型 | POI 数量 | 序号 | POI 类型    | POI 数量 |
|----|--------|--------|----|-----------|--------|
| 1  | 餐饮服务   | 365    | 7  | 商务住宅      | 83     |
| 2  | 风景名胜   | 6      | 8  | 生活服务      | 292    |
| 3  | 公司企业   | 200    | 9  | 体育休闲服务    | 85     |
| 4  | 购物服务   | 57     | 10 | 医疗保险服务    | 200    |
| 5  | 金融保险服务 | 89     | 11 | 政府机构及社会团体 | 207    |
| 6  | 科教文化服务 | 115    | 12 | 住宿服务      | 101    |

## 2 POI 数据获取

传统的 POI 获取方式是实地调研采集,但数据量小不能大范围的进行数据分析。随着地理信息大

数据的发展,人们对全球地理信息数据的需求也日趋增加,获取方式也更加便利。

### 2.1 POI 数据获取方式

目前,POI 获取主要有 3 种途径:①通过爬虫代码从电子地图获取;②通过电子地图开放平台(百度地图 API、高德地图 API 等)来抓取 POI,这些网站开放了较为完善的开发接口<sup>[10]</sup>;③通过集客大数据、Geosharp、火车头和八爪鱼等软件,直接采集 POI 数据。利用 Geosharp 软件获取 POI 数据的操作相对简单、可行性高,并且 POI 数据的准确率高,因此,通过它获取研究区域的 POI。

### 2.2 Geosharp 获取 POI 数据的思路

Geosharp 是一种网络地图数据采集软件,主要包括地理编码、坐标转换和数据采集 3 个模块。地理编码模块可以把地址信息解析为经纬度;坐标转换提供百度、WGS84 和火星坐标系之间的转换;数据采集模块可采取所有类型的 POI 数据和天气预报数据。

Geosharp 获取 POI 数据可分为 3 个步骤:①选择需要获取的 POI 类型。根据高德地图分类标准,采用 POI 的一级分类获取。②确定研究区域,设置研究区域的经纬度。如果研究范围不大,则可以直接获取。但研究范围过大,需要进行格网划分,分区获取。③确定数据保存路径。

### 2.3 Geosharp 获取 POI 步骤

Geosharp 的“高德地图 POI 采集工具”只需研究范围的经纬度,即可采集该矩形范围内所有 POI 数据,根据需要导出到 Excel 中。通过经纬度在地图上添加 X,Y 坐标,生成位置信息。通过 POI 类别,研究各类型 POI 的出行特征。通过 ID,确定 POI 的唯一性。

### 2.4 FME 区域格网划分

POI 数据在进行大范围采集时,由于研究区域过大,如果一次输入整个范围的经纬度,不仅会花费大量时间,而且容易造成软件异常和数据丢失。因此,会增加获取数据的难度,其结果的误差也较大。其解决办法是对该区域划分成格网单元<sup>[11]</sup>,按照每个网格的经纬度坐标抓取 POI,再对数据进行合并。通过多种区域格网划分方法进行比较,发现 FME<sup>[12]</sup>数据的交换与处理操作最简单,自动化程度高。因此,采用 FME 对区域进行格网划分,并自动提取格网坐标。FME 可以实现超过 200 种不同空

间数据格式(模型)之间的转换,是一款无缝连接多平台空间数据的操作工具。FME 的操作流程主要分为读模块、转化器模块和写模块3个部分。读模块可以从外部数据源读取要素数据。转换器可以在使用者的控制下,将这些数据合并或分割,从一种表达格式转换为另一种表达格式,也可以将这些要素挂接到外部数据库上。写模块将这些要素以一种支持的格式进行输出。

本研究通过 FME 软件进行建模,自动对城市的格网进行划分,提取格网坐标。具体流程:

- 1) 通过读模块将研究范围地理文件进行导入;
- 2) 通过 2DGridAccumulator 转换器生成格网,按照行列数进行拆分,创建格网类型,选择 Polygons;
- 3) 通过 GeometryCoercer 转换器将研究范围面转换为线,几何对象类型改为 fine\_line;
- 4) 通过 BoundsExtractor 转换器提取格网坐标;
- 5) 通过 AttributeManager 转换器将经纬度合并成一个单元格;
- 6) 通过 counter 转换器进行计数,设置计数从 1 开始;
- 7) 通过写模块将格网经纬度输出到 Excel 文件。

根据 FME 构造模型,将研究区域进行自动化划分格网,得到每个格网左上角和右下角的经纬度坐标。再将经纬度坐标依次输入 Geosharp 高德地图的 POI 采集工具。最后,通过 Geosharp 分区获取 POI 数据,得到整个研究区域所有的 POI 数据。

### 3 POI 数据处理

#### 3.1 POI 数据区域整合

通过 Geosharp,从高德地图中采集 POI 数据类型较多,但采用手动方法对各类 Excel 进行合并的,操作比较繁琐。因此,采用 FME 进行建模,自动将各类型 POI 进行合并。FME 模型构建流程为:

- 1) 读取文件夹,将 POI 的 Excel 文件通过读模块进行导入。
- 2) 通过 AttributeSplitter 转换器,对 POI 类别进行拆分,创建 3 个属性。Attribute Value 值分别为 type\_list{0}, type\_list{1}, type\_list{2}。
- 3) 通过写模块将所有数据输出到 Excel。

#### 3.2 POI 数据与地图关联

通过 ArcGIS for Desktop 实现 POI 数据和地图的关联。ArcGIS for Desktop 作为 GIS 中的主流产品,其特点是数据管理效率高、空间分析能力强、可视化程度高。ArcGIS 为大数据的管理、分析和可视化提供了有效的处理工具。ArcGIS for Desktop 主要由 ArcCatalog (管理空间数据)、ArcGlobe(大型三维场景处理与分析)、ArcMap(二维数据的处理、管理和分析)和 ArcScene(小型三维场景处理与分析)4 个部分构成。

通过 FME 数据合并模型对 POI 数据进行整合,导入到 ArcMap 中,按照经纬度属性添加 X, Y 坐标,生成位置信息,设置坐标系统,将 POI 地理信息数据可视化,得到整个研究区域的 POI 分布情况。POI 数据可视化结果如图 1 所示。

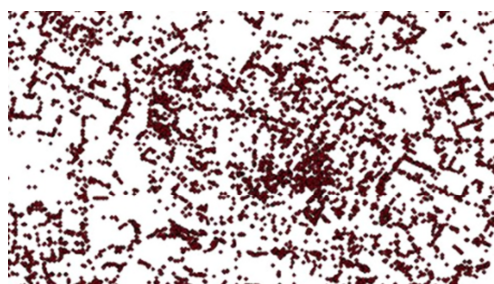


图1 POI 点数据空间分布示意

Fig. 1 Spatial distribution of POI point data

#### 3.3 POI 核密度分析

如果将每个 POI 点均作为一个功能单元,POI 密度越高,表明该地区城市功能越集中。为了分析 POI 的聚集特征,了解各类服务设施的空间分布状况,利用 ArcMap 10.2 的密度分析工具,对研究区域内的各类兴趣点进行分析。

##### 3.3.1 核密度分析原理

密度分析是通过输入数据的数量计算数据的集散程度,生成一个连续的密度分析面。通过计算得到密度,将每个点的密度值分布在研究区域上,最后得到每个栅格的像元值<sup>[13]</sup>。

密度分析中,最为重要的是核密度分析。核密度分析一般用于计算要素周围邻域的密度。既可以计算点要素的密度,也可以计算线要素的密度。核密度分析示意图如图 2 所示。

根据概率理论,核密度分析模型为<sup>[14-15]</sup>:

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n k\left(\frac{x - X_i}{h}\right). \quad (1)$$

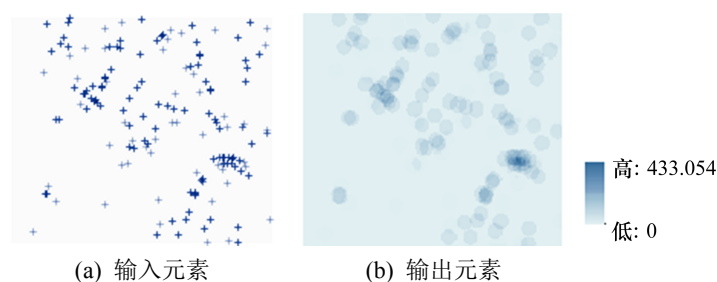


图 2 核密度分析输出要素示意

Fig. 2 Schematic diagram of output elements of kernel density analysis

式中:  $k()$  为核函数;  $h$  为带宽,  $h > 0$ ;  $x - X_i$  为估值点  $x$  到事件  $X_i$  处的距离。

核密度分析可以体现 POI 数据点的空间分布, 对点状数据进行分析尤为有效。通过核密度分析, 计算每个输出栅格像元周围点要素的密度, 生成 POI 数据点热力图。

核密度分析中, 需要设置输出像元大小和搜索半径 2 个参数。设置的搜索半径越大, 生成的密度栅格越平滑, 概化程度越高; 搜索半径值越小, 生成栅格显示的信息越详细。

### 3.3.2 交通小区划分

交通规划需要全面了解 POI 对各类交通源的吸引程度。由于不可能对每个 POI 单独进行研究, 因此, 分析 POI 时, 需要根据交通小区对 POI 进行合并处理。

### 3.4 POI 分区导出模型构建

为了将所有 POI 数据按照交通小区依次导出, 得到每个小区的各类 POI。传统的方法是划分好交通小区, 依次裁剪, 导出 POI 数据。但当数据量较大, 该方法费时费力, 效率较差。ArcMap 可提供一种自动化处理工具—ModelBuilder, 通过 ArcMap 的模型构建器, 轻松实现批处理工作。

ModelBuilder 是创建、编辑和管理模型的应用程序<sup>[16]</sup>。其原理是将一系列地理处理工具串联在一起的工作流, 它将其中一个工具的输出作为另一个工具的输入。可将 ModelBuilder 看作一种简单创建、可循环操作的可视化编程语言。进行模型构建时, 可以直接将工具箱的各种地理处理工具和需要处理的数据集拖动到模型构建器界面中。然后, 有序地把它们连接起来, 实现批处理工作。

模型构建器的优势: ①ModelBuilder 可以集合 ArcToolbox 各种工具, 进行创建, 并且自动化处理

所构建的工作流; ②结合使用 ModelBuilder 和脚本, 可将 ArcGIS 与其他应用程序进行集成; ③ModelBuilder 创建的自动化模型, 可以在 Python 脚本中使用。

本模型构建中, 将划分好的交通小区在 ArcMap 按字段依次导出, 添加裁剪工具, 输入要素为 POI 点数据, 裁剪要素为交通小区面数据, 得到各个交通小区的 POI 数量。

### 3.5 交通小区当量 POI 的计算

通过 ArcMap 中 ModelBuilder 构造, 批量导出模型, 将整体 POI 数据按照交通小区进行划分, 依次导出。交通小区内, 不同类型的 POI 对于交通规划的影响程度不一。通过结构熵权法, 确定各类 POI 对公交出行的影响权重, 依次对不同类别 POI 进行量化计算, 得到各个小区的当量 POI 数据, 为后续交通规划、公交线网优化<sup>[17]</sup>提供数据支撑。量化公式为:

$$D_i = \sum_{e=1}^E w_e \cdot d_{ie} \quad (2)$$

式中:  $D_i$  为交通小区  $i$  的当量 POI;  $E$  为获取的 POI 类型数;  $w_e$  为第  $e$  类 POI 的权重;  $d_{ie}$  为交通小区  $i$  第  $e$  类 POI 数量。

## 4 实例分析

以溆浦县为例, 通过 Geosharp 软件和高德地图, 获取城区各类 POI 数据。获取 POI 之前, 避免区域过大, 造成数据丢失, 对研究区域进行格网划分, 采用 FME 格网划分模型对溆浦县城区进行格网划分, 如图 3 所示。

通过 Geosharp, 由高德地图分区, 获取格网小区 POI。根据 FME 的数据合并模型, 将格网小区 POI 数据进行合并、数据清洗, 为 ArcGIS 建立 POI



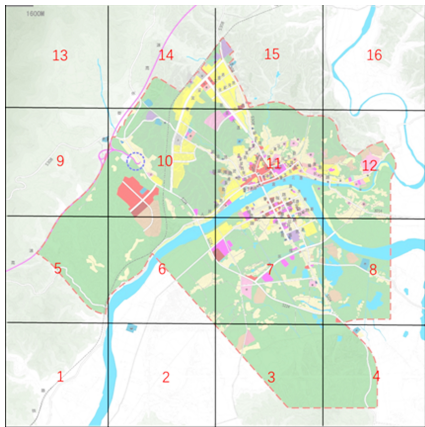


图 3 溆浦县网格划分示意

Fig. 3 Grid division diagram of the Xupu County

地理数据库进行前期数据准备。本试验获取了 12 个 POI 类型, 有效 POI 数量为 1 800 条, 见表 1。

先将 POI 数据导入 ArcGIS, 添加经纬度信息, 设置地理坐标系为 WGS 1984, 导入溆浦县中心城区用地现状图, 进行地理配准。再将 POI 数据加载到溆浦县中心城区用地现状图进行核密度分析。设置热力图颜色, 颜色越深, 表示 POI 数据比较密集; 颜色越浅, 表示 POI 聚集程度较为稀疏。溆浦县 POI 数据空间分布如图 4 所示。

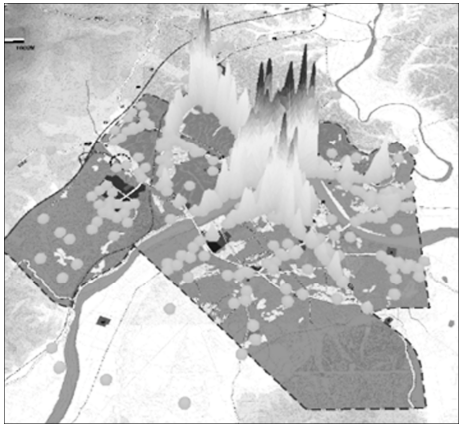


图 4 溆浦县 POI 空间分布热力

Fig. 4 Thermal map of of POI in the Xupu County

根据溆浦县城区道路现状和城市规划分区情况, 将该区域划分为 22 个交通小区, 小区划分示意图如图 5 所示。

通过 ArcMap 的 ModelBuilder 构造, 批量导出模型, 将所有 POI 数据按照划分的交通小区依次导出, 得到每个小区的 POI 类别和数量, 通过权重进行量化见表 2。

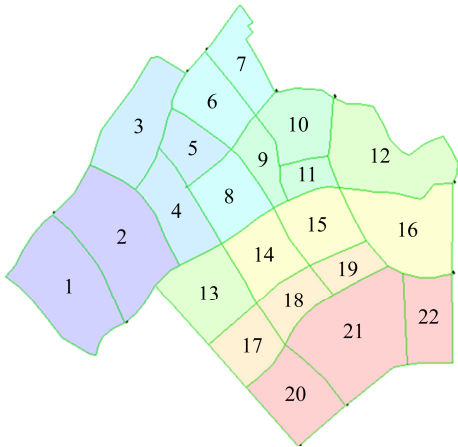


图 5 溆浦县交通小区划分示意

Fig. 5 Division diagram of traffic district of the Xupu County

表 2 量化后交通小区 POI 数量

| Table 2 Number of POI in traffic area after quantification |        |                    |      |        |                    |
|------------------------------------------------------------|--------|--------------------|------|--------|--------------------|
| 小区编号                                                       | 当量 POI | 面积/km <sup>2</sup> | 小区编号 | 当量 POI | 面积/km <sup>2</sup> |
| 1                                                          | 0.14   | 2.08               | 12   | 12.22  | 1.89               |
| 2                                                          | 0.80   | 2.44               | 13   | 0.45   | 1.37               |
| 3                                                          | 0.78   | 1.79               | 14   | 2.25   | 1.03               |
| 4                                                          | 0.88   | 1.11               | 15   | 25.83  | 1.09               |
| 5                                                          | 1.75   | 0.75               | 16   | 2.23   | 1.84               |
| 6                                                          | 4.52   | 1.07               | 17   | 0.31   | 0.84               |
| 7                                                          | 18.32  | 0.89               | 18   | 0.75   | 0.69               |
| 8                                                          | 5.27   | 1.04               | 19   | 0.93   | 0.52               |
| 9                                                          | 7.84   | 0.61               | 20   | 0.03   | 1.25               |
| 10                                                         | 10.10  | 1.07               | 21   | 2.48   | 2.55               |
| 11                                                         | 25.91  | 0.44               | 22   | 0.31   | 1.05               |

5 结论

通过 FME 构造格网坐标提取模型, 对研究区域进行格网划分, 再采用 Geosharp 软件获取 POI 数据, 最后基于 ArcGIS 模型构建分区导出模型, 将 POI 数据分区导出, 并进行量化处理。通过采用格网坐标提取模型和分区导出模型进行自动化处理, 再导出各类 POI 数据。该方法比原有裁剪导出操作更节省时间, 也更有效率。

参考文献(References):

[1] Mummidi L N, Krumm J. Discovering points of interest from users' map annotations[J].GeoJournal, 2008,72(3-4):215-227.

[2] Xie Z, Yan J. Kernel density estimation of traffic

- accidents in a network space[J].Computers, Environment & Urban Systems,2008,32(5):396-406.
- [3] Lian D, Zhao C, Xie X, et al. GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation[C]//ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM,2014:831-840.
- [4] 李伟,张涵之.基于 POI 的厦门岛常规公交线路可达性评价[J].城市建筑,2018(8):112-115.(LI Wei, ZHANG Han-zhi. On conventional public transport route reachability evaluation based on POI Xiamen island[J]. Urbanism and Architecture,218(8):12-115.(in Chinese))
- [5] 于浩川.城市公交线路规划系统的设计与实现[D].北京:北京交通大学,2016.(YU Hao-chuan, Design and implementation of city transportation planning system[D]. Beijing:Beijing Jiaotong University,2016.(in Chinese))
- [6] 周春辉,朱欣焰,苏科华,等.基于 LBS 的兴趣点查询与更新机制研究[J].微计算机信息,2009(19):143-145.(ZHOU Chun-hui, ZHU Xin-yan, SU Ke-hua, et al. Research on POI query and update mechanism based on LBS[J]. Microcomputer Information, 2009(19): 143-145. (in Chinese))
- [7] 许泽宁,高晓路.基于电子地图兴趣点的城市建成区边界识别方法[J].地理学报,2016,71(6):928-939.(XU Ze-ning,GAO Xiao-lu.A novel method for identifying the boundary of urban built-up areas with POI data[J].Acta Geographica Sinica,2016,71(6):928-939.(in Chinese))
- [8] 王腾,李佳.基于 POI 数据的城市商业中心识别与集聚特征分析——以乌兰察布市集宁区为例[J].住宅与房地产,2018(2):215-216.(WANG Teng, LI Jia. Identification and clustering characteristics of urban commercial centers based on POI data—Taking Jining district of Ulanqab city as an example[J].Housing and Real Estate,2018(2):215-216.(in Chinese))
- [9] 房佳萱.基于 POI 的深圳市南山区步行可达性评价研究[D].哈尔滨:哈尔滨工业大学,2017.(FANG Jia-xuan. Research of walking accessibility based on POI in nanshan district[D].Harbin:Harbin Institute of Technology, 2017.(in Chinese))
- [10] 黄磊,许大璐,任福,等.一种微博 POI 签到数据的四叉树格网获取方法[J].地理信息世界,2018,25(2):92-96.(HUANG Lei, XU Da-lu, REN Fu, et al. Acquisition of microblog POI check-in data based on quadtree dynamic grid[J]. Geographic World, 2018, 25(2): 92-96. (in Chinese))
- [11] 裴莲莲,唐建智,毕小硕.多源空间大数据的获取及在城市规划中的应用[J].地理信息世界,2019,26(1):13-17.(PEI Lian-lian, TANG Jian-zhi, BI Xiao-shuo. The acquisition of multi-source spatial data and its application to urban planning[J]. Geographic World, 2019, 26(1): 13-17.(in Chinese))
- [12] 周浩,王永峰,蔡欢.基于 FME 和 ArcEngine 的基础空间数据出入库系统设计与实现[J].测绘与空间地理信息,2018,41(12):112-114.(ZHOU Hao, WANG Yong-feng, CAI Huan.Design and implementation of basic spatial data warehouse system based on FME and Arcengine[J]. Geomatics & Spatial Information Technology, 2018, 41(12):112-114.(in Chinese))
- [13] 牟乃夏.ArcGIS 10 地理信息系统教程:从初学到精通[M].北京:测绘出版社,2012.(MU Nai-xia.ArcGIS 10 GIS course: From beginning to master[M]. Beijing: Mapping Press,2012.(in Chinese))
- [14] Okabe A,Satoh T,Sugihara K.A kernel density estimation method for networks, its computational method and a GIS-based tool[J].International Journal of Geographical Information Science,2009,23(1):7-32.
- [15] Dehnad K.Density estimation for statistics and data analysis[J].Technometrics,1987,29(4):495-495.
- [16] 林璐,马丽华,陈延博,等.利用 ArcGIS 模型构建器实现矢量数据整合的方法[J].地理空间信息,2018,16(4):62-63,10.(LIN Lu,MA Li-hua,CHEN Yan-bo,et al. Vector-data integration method based on ArcGIS model builder[J].Geospatial Information,2018,16(4):62-63,10. (in Chinese))
- [17] 朱飘,龙科军.轨道交通与常规公交竞合关系一体化研究[J].交通科学与工程,2019,35(1):94-100.(ZHU Piao, LONG Ke-jun, Research on the integration of rail transit and conventional bus co-opetition[J].Journal of Transport Science and Engineering, 2019, 35(1): 94-100. (in Chinese))