

基于POI与人口流动数据的城市功能识别与分区研究

姚文山

周口市国土空间规划编制研究中心 河南 周口 466000

摘要: 在城镇化迅猛推进的当下,精准识别城市功能区域并实施合理分区,对于城市规划的科学性、资源配置的高效性以及城市发展的可持续性具有不可忽视的重要意义。本文基于兴趣点(POI)数据与人口流动数据的城市功能识别与分区方法展开深入探讨。首先,详细阐述了基于这两种数据源的城市功能识别流程,涵盖数据预处理、特征提取、功能分类等关键步骤;然后,结合广西POI数据、人口流动数据进行案例探索分析;最后,系统探讨了基于识别结果的城市功能分区策略,以期对同类研究提供参考借鉴。

关键词: POI数据;人口流动数据;城市功能识别;城市功能分区

1 引言

城市是人类文明的结晶,内部功能区域多样,其合理布局与协同发展关乎城市运行效率及其可持续发展。准确识别与科学划分城市功能区域,能为城市规划科学决策提供依据,优化空间布局、提高资源利用率、缓解交通拥堵、改善生态环境、促进城市有序发展。传统方法依赖土地利用调查和问卷调查,存在成本高、更新慢、数据易受主观影响等局限。如今随着信息技术发展,大数据带来新契机,POI数据能直观反映城市空间结构和功能分布,人口流动数据体现区域交互与功能联系,基于二者研究可克服传统方法不足,为城市规划和管理提供更科学精准的支持。

2 基于POI与人口流动数据的城市功能识别流程

2.1 数据预处理

2.1.1 POI数据预处理

数据清洗: 原始的POI数据可能存在重复、错误和不完整等问题,需要进行数据清洗。例如,对于名称不规范、地理位置错误的POI点,可以通过人工审核或自动纠错算法进行修正或删除。对于重复的POI记录,可以根据地理位置和名称等信息进行去重处理。此外,还需要检查POI数据的属性信息是否完整,对于缺失的属性信息,可以通过数据挖掘或人工补充的方式进行完善。

分类编码: 根据城市功能的特点和研究需求,对POI数据进行分类编码。常见的分类方法包括按照设施类型(如商业、教育、医疗等)或按照功能强度(如大型商场、小型便利店)进行分类。例如,将商场、超市、餐厅等划分为商业类POI;将学校、培训机构等划分为教育类POI;将医院、诊所等划分为医疗类POI。在分类编码过程中,需要制定统一的分类标准和编码规则,确保数据的一致性和可比性。

空间插值: 对于分布稀疏的POI数据,可以采用空间插值方法进行补充,以提高数据的空间覆盖率和完整性。常用的空间插值方法包括反距离权重插值、克里金插值等。例如,在某些偏远地区或新建区域,POI数据可能较少,通过空间插值可以根据周边区域的POI分布情况,估算出这些区域的POI密度和类型分布,从而为城市功能识别提供更全面的数据支持。

2.1.2 人口流动数据预处理

数据清洗: 人口流动数据通常包含大量的噪声数据和异常值,如信号漂移、重复记录等,需要进行数据清洗。例如,对于手机信令数据中出现的短时间内位置异常跳变的情况,可能是由于信号干扰或设备故障导致的,可以通过滤波算法或阈值判断的方法进行过滤。对于重复的轨迹记录,可以根据用户ID和时间戳等信息进行去重处理^[1]。此外,还需要检查数据的完整性和准确性,对于缺失的时间戳或位置信息,可以通过数据插补或删除的方式进行修复。

轨迹提取: 从原始的人口流动数据中提取出有效的移动轨迹,包括起点、终点、时间戳等信息。例如,对于手机信令数据,可以通过分析用户的基站切换记录,提取出用户的移动轨迹。对于公交刷卡数据,可以根据乘客的上下车时间和站点信息,构建乘客的出行轨迹。在轨迹提取过程中,需要考虑数据的精度和可靠性,对于不完整的轨迹或异常轨迹,需要进行进一步的筛选和处理。

时间切片: 将人口流动数据按照一定的时间间隔(如小时、天)进行切片,以便分析不同时间段的人口流动特征。例如,将一天24小时划分为24个时间段,分别统计每个时间段内各区域之间的人口流动数量。通过时间切片,可以了解人口流动的日变化规律和季节变化

规律，为城市功能识别提供更细致的时间维度信息。

2.2 特征提取

2.2.1 基于POI数据的特征提取

数量特征：统计不同功能类别POI的数量，反映各区域不同功能的规模。例如，计算某区域内商场的数量，以衡量该区域的商业功能强度。可以通过设置不同的空间范围（如网格、行政区划等）来统计POI数量，得到各区域不同功能类别的POI数量分布。

密度特征：计算单位面积内POI的数量，反映各区域功能的密集程度。例如，计算每平方公里内学校的数量，评估该区域的教育资源分布密度。密度特征可以更直观地反映区域功能的集中程度，有助于识别城市功能的核心区域和边缘区域。

多样性特征：采用香农多样性指数等方法，计算各区域POI类型的多样性，反映区域功能的综合程度。香农多样性指数的计算公式为 $H = -\sum_{i=1}^n p_i \ln p_i$ ，其中 P_i 为第 i 类POI在该区域POI总数中所占的比例， n 为POI的类别数。多样性特征越高，说明该区域的功能越丰富多样，可能是一个综合性的功能区域。

空间分布特征：分析POI的空间分布模式，如集聚程度、分散程度等。可以通过空间自相关分析、核密度估计等方法来研究POI的空间分布特征。例如，通过核密度估计可以绘制出POI的密度分布图，直观地展示POI的集聚区域和稀疏区域，为城市功能识别提供空间维度的信息支持。

2.2.2 基于人口流动数据的特征提取

流量特征：统计不同时间段内各区域之间的人口流动数量，反映区域之间的交互强度^[2]。例如，计算工作日早高峰时段从居住区域A到工作区域B的人口流动数量。流量特征可以反映城市不同区域之间的功能联系和通勤需求，有助于识别城市的主要通勤走廊和功能节点。

停留时间特征：分析人口在不同区域的停留时间，反映区域对人口的吸引力。例如，计算某商业区域顾客的平均停留时间，评估该区域的商业活力。可以通过设置停留时间阈值，将停留时间较长的区域识别为具有较强吸引力的功能区域。

流动方向特征：研究人口流动的主要方向，揭示城市功能的空间联系模式。例如，通过分析人口流动数据，发现城市中存在从外围居住区向中心商务区的主要流动方向，这反映了城市功能的核心-边缘结构。流动方向特征可以为城市交通规划和功能布局优化提供重要参考。

时间序列特征：分析人口流动数量随时间的变化规律，反映城市功能的动态变化。例如，通过绘制人口流

动数量的时间序列图，可以观察到工作日和节假日人口流动的差异，以及不同季节人口流动的变化趋势。时间序列特征有助于了解城市功能的季节性变化和周期性规律，为城市规划和管理提供动态的信息支持。

2.3 功能分类

2.3.1 构建分类模型

采用机器学习算法，如支持向量机（SVM）、随机森林（RF）、神经网络等，构建城市功能分类模型。将提取的POI特征和人口流动特征作为模型的输入变量，以已知功能的区域作为训练样本，对模型进行训练和优化。在构建分类模型时，需要对数据进行归一化处理，将不同量纲的特征值转换为统一的尺度，以提高模型的训练效果。

2.3.2 模型评估与验证

使用交叉验证等方法对分类模型的性能进行评估，选择准确率、召回率、F1值等指标来衡量模型的性能。准确率是指模型正确分类的样本数占总样本数的比例；召回率是指模型正确分类的正样本数占实际正样本数的比例；F1值是准确率和召回率的调和平均数，综合考虑了模型的精确性和召回率。

2.3.3 功能分类结果

将训练好的分类模型应用于整个城市区域，对每个区域进行功能分类，得到城市不同功能区域的初步识别结果。常见的城市功能类型包括居住功能区、商业功能区、工业功能区、公共服务功能区、休闲娱乐功能区等。在功能分类过程中，可能会存在一些分类不明确区域，需要结合实际情况和专家知识进行进一步的判断和修正。

3 广西案例探索

3.1 基于POI数据的广西公共服务功能分析

通过网络爬虫技术，基于百度地图爬取广西壮族自治区全域共22148个公共服务设施POI点，归并为医疗、教育、社会福利、文化、体育5大类基础公共服务设施，其中教育设施细分高等教育、职业教育、中小学3类。然后采用核密度分析技术探讨各类公服设施分布的热点区域，结果显示：层次体系上呈现“主核+次核”布局，南宁是公共服务设施分布最为集中的区域，其次柳州、桂林、玉林是公共服务设施分布较为集中布局的区域。空间分布存在差异性，桂东南公共服务设施多而稠密，桂西北公共服务设施少而分散，与人口分布基本一致。

3.2 基于人口流动数据的广西职住平衡分析

通过清洗广西手机信令数据，将其属性链接入县级行政区网格，然后根据所在区县进行统计，得到每个区

县的居住用户总数和就业用户总数,相除后得到区县的职住比。结果显示:广西所有区县的职住比平均值0.99,中位数0.97,整体职住比良好。南宁-柳州-桂林轴线上的区县单位职住比整体较高。职住比大于1的区县单位共49个,集中于南宁-柳州-桂林轴线;其他多为地级市市辖区,反映出轴线上经济活动活跃,提供就业岗位多;职住比最高是前10个的区县单位中4个位于柳州,3个位于桂林;东兴、凭祥两处边境城市也有着较高职住比。桂东地区贺州、梧州、贵港、玉林四市的区县单位职住比整体偏低。绝大部分区县的职住比小于0.9。原因是本地居民多选择外出务工,留守非劳动力居住人口较多。

4 基于识别结果的城市功能分区策略

4.1 分区原则

城市功能分区需遵循多项原则。功能相似性原则要求把功能类型相近区域划为同一分区,如合并多个商业功能区为商业中心分区,可集中布局商业设施,增强集聚效应与竞争力,优化城市功能布局,提升运行效率。空间连续性原则强调考虑城市空间结构连续性,保持功能分区空间完整,避免破碎化,像划分居住功能分区时尽量将相邻居住区域划为同一分区,利于维护城市整体形象与空间秩序。交通便利性原则需结合城市交通网络,确保各功能分区交通联系便捷,如重要功能节点与交通干道相连,方便人员物资流动,缓解拥堵、促进经济发展。

4.2 分区方法

4.2.1 基于聚类分析的分区方法

采用K-means、层次聚类等聚类算法,对城市功能识别结果进行聚类分析。根据聚类结果,将功能相似的区域划分为同一功能分区。在聚类过程中,可以根据实际情况调整聚类参数。例如,通过K-means聚类算法,可以将城市功能区域划分为若干个簇,每个簇代表一个功能分区^[3]。聚类分析方法能够自动发现城市功能的内在结构和分布规律,但可能会受到初始聚类中心和聚类参数的影响。

4.2.2 基于空间分析的分区方法

利用GIS软件的空间分析功能,如缓冲区分析、叠加分析等,结合城市道路、河流等自然和人工边界,对城市功能区域进行划分。例如,以主要交通干道为边界,将城市划分为不同的功能片区;以河流为自然边界,划分出生态保护区等功能分区。空间分析方法能够充分考

虑城市的地理环境和空间结构,但可能会受到边界划分的主观性影响。

4.3 分区优化

4.3.1 功能平衡优化

检查各功能分区内部的功能平衡情况,避免出现功能单一或功能过度集中的问题。例如,在居住功能分区中适当增加公共服务设施和休闲娱乐设施,提高居民的生活质量;在商业功能分区中合理布局不同类型的商业业态,满足居民的多样化消费需求^[4]。功能平衡优化有助于提高功能分区的综合服务能力和吸引力。

4.3.2 交通衔接优化

加强各功能分区之间的交通衔接,优化交通网络布局,提高交通运行效率。例如,在功能分区之间规划建设快速路、轨道交通等交通设施,缩短区域之间的通行时间;在功能分区内部完善慢行交通系统,提高居民出行的便利性。交通衔接优化有助于促进城市各功能分区之间的协同发展。

结语

本文基于POI与人口流动数据提出科学、系统的城市功能识别与分区方法,经数据预处理、特征提取等步骤实现准确识别与合理分区。该方法能克服传统方法不足,为城市规划管理提供科学依据。POI与人口流动数据结合可为功能识别提供丰富信息,基于机器学习的功能分类模型准确性和泛化能力强,综合聚类与空间分析分区能兼顾多方面原则。不过,研究仍有不足,未来需融合社交媒体、遥感影像等多源数据,建立动态监测更新机制,开展多尺度研究,并优化模型引入深度学习算法、考虑空间异质性等,以提升研究效果。

参考文献

- [1]韩世聆.整合POI和建筑足迹的城市功能区识别研究[D].上海师范大学,2025.
- [2]曹文琦,毋亭.基于POI数据的城市功能分区识别研究——以福建省主要城区为例[J].房地产世界,2022,(22):13-16.
- [3]李满娥,缪淑佳,王小龙.基于POI数据的城市功能区识别方法研究[J].测绘与空间地理信息,2025,48(01):133-135.
- [4]孙明英,王灵芝,钟巧莹,等.基于POI数据的城市功能识别与分布特征研究——以长春市为例[J].西南大学学报(自然科学版),2025,47(01):189-199.