

DOI:10.13203/j.whugis20170383



文章编号:1671-8860(2018)03-0327-09

集成多源地理大数据感知城市空间分异格局

刘 瑜^{1,2} 詹朝晖² 朱 递² 柴彦威³ 马修军^{4,5} 邬 伦²

1 国土资源部城市监测与仿真重点实验室,广东 深圳,518034

2 北京大学遥感与地理信息系统研究所,北京,100871

3 北京大学城市与环境学院,北京,100871

4 北京大学信息科学技术学院,北京,100871

5 北京大学机器感知与智能教育部重点实验室,100871

摘 要:多源地理大数据为地理现象的分布格局、相互作用及动态演化提供了前所未有的社会感知手段。城市是人类活动最为集中的区域,产生了多种地理大数据,并支持对于城市空间的理解。城市内部的分异格局是城市研究和规划所要面对的重要议题,社会感知数据提供了从“人-地-静-动”4 个维度刻画城市分异格局的途径。梳理了不同类型大数据对于表达这 4 个维度特征的支持,并借鉴“生态位”模型,通过一个实例研究展示了集成多源数据量化城市空间分异特征的应用,最后讨论了相关的理论问题。

关键词:地理大数据;空间分异;社会感知;城市

中图法分类号:P208

文献标志码:A

近年来,随着信息通信技术的发展,尤其是移动通信技术的发展,人们获取和传输信息的能力大大加强。每个人都可以扮演传感器的角色,随时随地获取带有个体标记和时空语义信息的数据^[1-2]。这些数据是大数据的重要组成部分,为地理空间分析及其在城市、交通、公共卫生等领域的应用提供了重要的支撑手段。地理大数据的类型很多,其中获得普遍应用的数据包括移动手机数据、公交数据、出租车 GPS 轨迹数据、社交媒体数据等。这些数据在采集方式、空间分辨率、用户属性的表达能力、活动语义表达能力、轨迹完整性等方面存在差异,适合于不同地理信息的分析提取。

针对地理大数据特点,文献[3]提出了社会感知这一概念,并建立了基于大数据提取人的空间行为模式及表征空间分异的研究框架。根据该研究框架,尽管每个个体的行为较为随机,但是大数据所反映的海量人群行为却往往呈现出有规律的模式,这种模式与地理环境的特征,尤其是社会经济方面的特征,密切相关。因此,可以基于地理大数据,实现“由人及地”的反演过程,一定程度上弥补了传统遥感技术侧重于感知自然地理特征的不

足^[4]。

城市是人类活动最为集中的区域,也是各种地理大数据产生最为密集的区域,因此,目前地理大数据研究多针对特定城市开展。在多源地理大数据支持下,提取不同区域的居民行为模式,并与传统地理空间数据相结合,揭示城市内不同空间单元的特征,是一个重要研究方向。其中,早期的工作主要利用不同类型地理大数据所提取的活动量日变化曲线揭示城市用地功能^[5-7],之后的研究则考虑了社交媒体数据提取的语义特征城市内差异^[8],并与遥感数据相结合^[9-10],以从多个维度感知城市的空间结构。

在城市研究和城市规划和管理中,表达和理解城市内部的空间分异格局是一项基础工作。给定一个城市,在确定基本空间分析单元的基础上选择合适的特征指标,从而定量刻画每个单元的属属性,并在宏观尺度上表达城市的空间分异格局。在地理大数据广泛应用之前,多采用针对物理空间的静态指标;大数据的出现,为构建描述社会经济空间的动态指标提供了有效的支撑手段。因此,结合多源地理大数据以及传统空间数据,有助

收稿日期:2017-11-23

项目资助:国土资源部城市土地资源监测与仿真重点实验室开放基金(KF-2016-02-023);国家自然科学基金(41625003)。

第一作者:刘瑜,博士,教授,主要从事地理信息科学与地理大数据理论方法研究。liuyu@urban.pku.edu.cn

通讯作者:邬伦,博士,教授。lwu@urban.pku.edu.cn

于从“人”和“地”两个维度、“动”和“静”两个视角，全面感知城市空间分异格局。本文将构建刻画城市内分异格局的指标体系，并梳理常见地理大数据对于获取上述指标的支持能力，尤其是对于与城市居民空间行为有关的指标的支持能力，进而通过一个实例展示构建多指标刻画城市分异格局的意义。

1 城市空间分异格局的指标体系

为了定量刻画城市空间结构，本文从时间与要素两个维度构建城市特征指标体系。该体系中，纵轴为城市的要素维度，两端分别为人与地。人与地一直以来都是城市研究，乃至地理研究的主体要素。实现以人为本的城市规划和管理，必须探究城市居民与城市环境的耦合关系及演化规律。横轴为时间维度，关注在不同时间尺度下城市特征所展现出来的不同含义。根据时间维可以将时间尺度下城市体征系统划分为静态与动态。通过两条轴线的交叉，可以将城市特征指标划分为4个子系统。第一象限基于人的动态系统，它关注城市中人群的活动/移动特征，包括城市利用、居民活动空间与设施可达性等内容，展现出人

在城市中的动态过程。第二象限为基于人的静态系统，即城市人口系统，包含居住、就业等内容。第三象限为基于地的静态系统，即关注城市建成环境，主要通过土地利用、交通设施布局方面体现；第四象限为基于地的动态系统，即城市运行系统，交通、设施利用、市政运行等是其主要内容。在城特征指标体系中，4个部分之间并不是完全割裂的，而是相互影响，从而形成完整的系统。

2 集成多源地理大数据量化城市特征指标

根据图1所示的特征指标体系，大数据主要提供了对于人的动态特征的感知表达能力。在城市研究中，主要有两条研究路径以构建人群异质性和地理空间环境异质性的耦合关系：(1)根据个体行为特征的相似性^[11]或社交关系联系的强度^[12]，对人群进行分类从而处理人群异质性，进而确定不同类别人群在空间的分布；(2)对城市空间进行划分，然后量化不同空间单元内的人群行为特征，并对空间单元进行分类。由于个体粒度数据获取难度较大、属性维数较少等不足^[4]，目前的研究多采用第二条技术路线，主要包括以下几个方向。

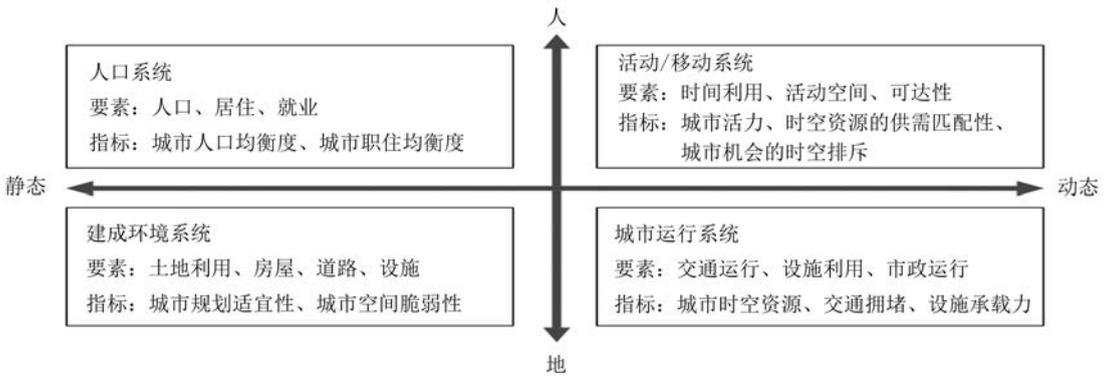


图1 城市特征指标体系
Fig.1 Index System of Urban Features

1)不同类型地理大数据反映了相应活动(如手机通话、搭乘出租车)事件点的空间分布，假定在不同位置的人群在一段时间内进行某类活动的比例相近，则可以用大数据所提取的点分布作为人口分布的估计^[13-14]。如图2所示，利用百度移动应用的定位数据，可以展示北京市的人口空间分布模式。在此基础上，考虑到城市不同用地单元在一天内不同时段的人数变化特征存在差异，并且呈现较强的日周期变化规律，可以基于不同单元的活动量变化曲线对城市用地进行分类^[5-7]。

2)随着社交媒体的广泛应用以及人工智能技

术的发展，人们可以获取大量的具有时空标签的文本及照片数据，并分析其中蕴含的语义信息，这种语义信息反映了人们对于城市内不同地理单元关联的认知和感受。基于文本的工作如利用Twitter数据分别开展的幸福情感和抑郁情感的度量及空间分布研究^[15-16]，文献^[8]利用Twitter数据揭示伦敦市人口与地理环境的分异格局。而在基于照片的研究中，利用谷歌(Google)街景图像以及照片共享网站的图像分析不同城市环境给人的感受差异成为最近的一个研究热点^[17-18]。

3)地理大数据可以量化地理单元之间的空间

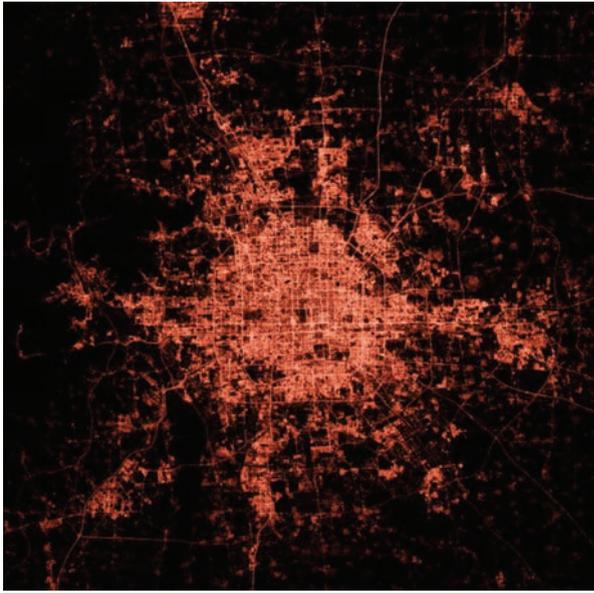


图 2 基于百度定位数据所估计的北京市人口空间分布
Fig.2 Baidu Positioning Data for Population
Distribution of Beijing

交互强度。除了采用网络科学方法构造空间交互网络以识别城市空间结构外^[19-20],空间交互也反映了一个地理单元的二阶特征。对于一个地理单元而言,其一阶特征是指自身具有的属性,如基于大数据所感知的活动时间变化曲线等,而二阶特征则是基于其与其他地理单元空间交互模式所表征的特征,它反映了在城市流空间中一个地理单元扮演的角色。因此,综合考虑一阶特征和二阶特征,可以更好地刻画城市空间分异^[21]。

上述 3 个方面的研究主要是基于单一地理大数据开展。正如文献[3]在提出社会感知概念时所指出的,集成遥感和社会感知数据,有助于人们更为全面地理解地理环境,目前已有较多研究通过结合遥感和社交媒体数据等新型地理大数据,进行城市用地分类^[9]、城市的多中心结构识别等研究。

值得指出的是,对于上述城市特征的刻画,采用常规数据采集手段(如出行调查)亦可获取,并且精度更高。大数据手段在具有成本低、样本量大的优点的同时,在代表性、精度等方面存在不足^[22],因此在实践中,一方面要重视大数据的数据质量问题,另一方面也要研究大数据和小数据的集成^[4],实现地理大数据支持下的“指标替代”计算方法,即利用大数据,经过合适的分析方法,得到可以替代传统方法获得指标的近似并加以替代,从而构建大数据支持的城市管理和规划流程。

3 实例研究

当前,我国正经历世界历史上规模最大、速度最快的城镇化发展。城市就业岗位有限,许多人自谋生计,如路边摆摊、非法营运等,这类活动给城市秩序、宜居城市建设等都带来挑战。城市专门组织了城管执法队伍加强城市管理,产生了一系列的城管执法事件。城管执法事件的时空点分布与城市环境存在一定关联,反映了城市脏、乱、差区域的分布。如何基于城管执法时间点的空间分布特征,构建其与相关城市特征之间的关联,进而理解其空间分异格局,是城市社会感知研究中的重要议题。

3.1 原理

生态位模型在广义上通常指所有根据生态位理论建立的数学统计或推测模型。其中,生态位是指一个物种能够生存和繁衍的所有条件的总和^[23]。目前,生态位模型一般被狭义地定义为以已知样本点(如野外调查或标本记录等)为基础,分析物种在生态位空间或环境空间中的特征,进而研究物种的环境耐受能力。生态位模型在预测物种分布、气候变化对物种分布的影响等方面都有广泛的应用。城市可以视为一个有机体,城市中不同的公共设施、商业设施,以及城市居民活动都对应于不同的城市空间环境;反言之,城市内部的地理空间分异也为不同设施及活动提供了相应的“生态位”。因此,人们同样可以在地理大数据支持下,构建基于“人-地-静-动”4个维度的城市环境变量的生态位模型,分析和预测设施及活动在城市空间的分布格局。

本文将利用 2015 年北京市城管执法事件在城市的已知分布特征,以相应区域的出租车上下车点数据、兴趣点(points of interest, POI)数据、交通便利条件等刻画城市环境特征,进而发现城管执法事件在城市其他区域潜在发生的可能性。

3.2 数据

因变量为城管执法数据,故对北京市城市管理综合行政执法局 2015 年立案的 711 起城市执法案件进行预处理,执法案件类型包括非法营运、无照经营、违法广告等,其具体内容如表 1 所示。利用地理编码手段将违法事实中的案件地点转换为经纬度坐标,得到城管执法事件在城市的时空点分布。本文认为城管执法数据并不能发现所有

的城市违法事件,因此必须利用合适的机器学习方法,找到已知城管执法事件存在的主要影响因子,并对城市中未被发现的城市违法事件进行预测。为了计算方便,将北京市六环范围内划分为

11 664个 500 m×500 m 的网格作为分析单元,其中存在城管执法事件的网格数为 277 个,执法事件数目由 1 到 35 不等,执法事件的空间分布如图 3 所示。

表 1 城管执法案件示例数据

Tab.1 Sample Records of Law Enforcement Events by City Inspectors

案件类型	部门名称	执法日期	经度/(°)	纬度/(°)
无照经营	西城区城管执法局	2016-01-26	116.370	39.892
非法营运	丰台区城管执法局	2016-01-06	116.393	39.861
违法广告	石景山区城管执法局	2016-01-24	116.232	39.915
擅自摆摊设点	东城区城管执法局	2016-01-25	116.419	39.875
擅自设置地锁	丰台区城管执法局	2016-01-14	116.368	39.871
未按规定清运垃圾渣土	朝阳区城管执法局	2016-01-18	116.568	39.916

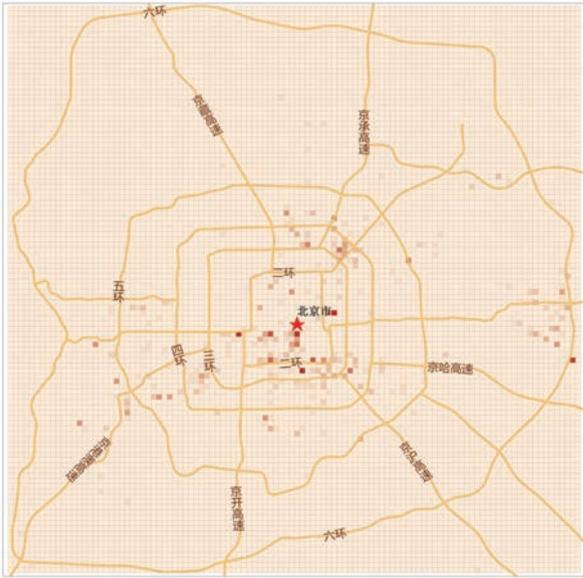


图 3 城管执法事件在北京市的空间分布

Fig.3 Spatial Distribution of Law Enforcement Events in Beijing

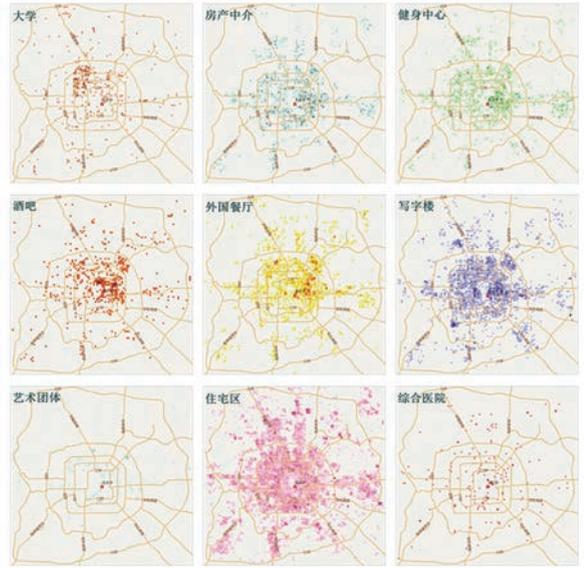


图 4 不同类型 POI 点在城市中的空间分布

Fig.4 Spatial Distribution of Various Types of POI

为了解释城管执法事件关联的城市环境,从 POI 数据、出租车上下车点数据、地铁出行距离等多个维度构建城市人类活动模型。城市中 POI 点的分布能够很好地反映城市结构及功能分区^[11],是通过城市特征指标体系中“地”的维度来描绘城市。本文采用的 POI 数据描绘了城市设施的静态特征,包括住宅区、写字楼、学校、医院、商场等,共计 84 种类型,407 971 个点(见图 4)。出租车数据可以作为计算城市居民出行特征的“替代指标”,反映了人在城市中的活动^[6,24]。对一天内的出租车上下车点数据进行分析,是从城市特征指标体系中“人”和“动”两个维度来描绘单元特征。地铁出行距离可以反映城市各区域的交通可达性,是从城市特征指标体系中“地”和“动”两个维度来描绘城市。

对 2013-05-13 - 2013-05-19 一周时间内,

17 000 辆出租车轨迹数据进行预处理,汇总得到工作日和休息日每小时网格内的上下车点均值,共计 96 个图层。对于每个分析单元,其上下车次数的日变化曲线反映了城市用地功能的特征^[6],图 5 展示了国贸区域的活动曲线。

为了刻画“地+动”方面的特征,研究采用到最近地铁站点的曼哈顿距离来刻画相应地理单元的空间可达性,该变量命名为 MHT,其空间分布如图 6 所示。

将上述数据汇总到数据表,一共得到 181 个解释变量,包括 POI(84)、出租车数据(96)和 MHT。利用机器学习方法构造 0-1 分类模型,其因变量是一个单元格是否存在城管执法事件。值得指出的是,在实践中,可以引入其他类型数据用于刻画每个分析单元“人-地-静-动”维度的特征,如社交媒体签到分布及主题特征、照片语义等,从而提高分类和预测精度,由于本文实例主要目的

是为了展示研究思路和方法,因此没有考虑更多 解释变量。

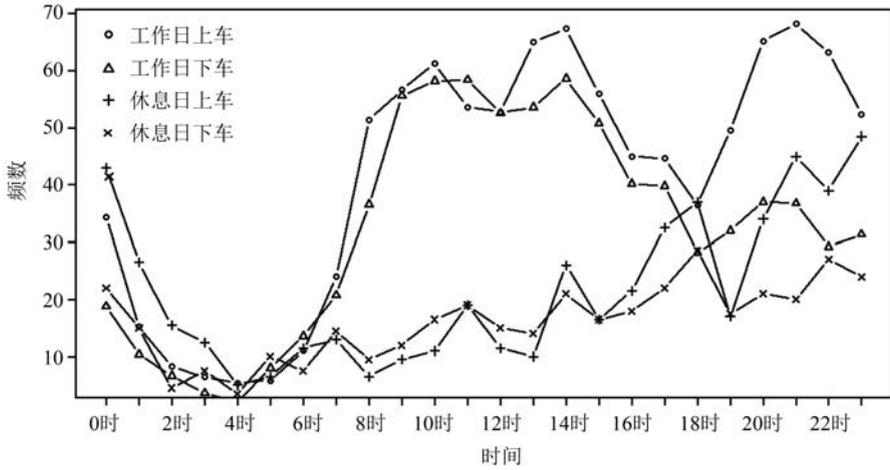


图 5 国贸区域出租车上下车每小时均值

Fig.5 Hourly Average Number of Taxi Pick-up and Drop-off in Guomao Area

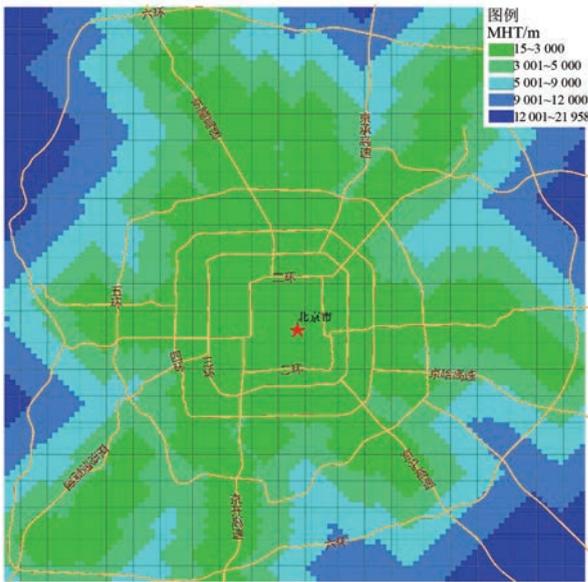


图 6 城市各区域距最近地铁站点的距离

Fig.6 Spatial Displacements Between Locations and Their Nearest Subway Stations

3.3 方法

随机森林是一种基于分类树的算法,即基于分类树组合成随机森林,并在变量的使用和数据的使用上进行随机化,生成多棵分类树,通过投票得到最终的预测结果。它具有很高的预测准确率,对异常值和噪声具有很好的容忍度,且不容易出现过拟合^[25]。考虑到随机森林方法的优势,本文采用它构建城管执法事件与城市环境特征之间的关联模型。

3.4 结果

对区域内是否存在城市违法事件进行预测。

城市违法事件的有或无,受到人类城市活动的影响。因变量为城管执法事件,解释因子则需要从 181 个变量中进行选择。通过特征选择可以去除大量冗余和不相关特征信息的干扰,有效降低特征空间的层数,从而提高对目标函数的预测性能。本文采用随机森林的 Boruta 算法对模型的特征打分,获得重要性较高的特征变量。如图 7 所示,最终选择排名前 14 的特征,包括 MHT、4 个 POI 点及 9 个出租车上下车变量,进而基于选取的解释变量,预测研究区内所有网格存在城市违法事件的概率。

为从实验得到的概率中选取最佳的临界值来识别网格区域内是否会发生城管执法事件,本文基于预测概率与已知城管事件的标签,将概率设定出多个不同的临界值,生成精度回归曲线(precision-recall curve, PR)如图 8(a)所示。PR 曲线是准确率和召回率的点连成的线,可以指导确定分类阈值。本文希望准确率高,即预测到发生城管事件区域都是准确的,因此假阳性对应的单元区域即为潜在可能发生城管执法事件的区域,通常城市环境较差。根据图 8(a),取受试者工作特征曲线(receiver operating characteristic curve, ROC)最佳临界值 0.063 时,准确率为 0.514。将最佳临界值调高到 0.1,此时的准确率为 0.887, Kappa 系数为 0.906。图 8(b)展示了预测结果的空间分布,其中红色单元(假阳性区域)表示潜在在城市违法事件发生的网格单元。

3.5 验证

为了验证模型预测结果,根据六环区域内各

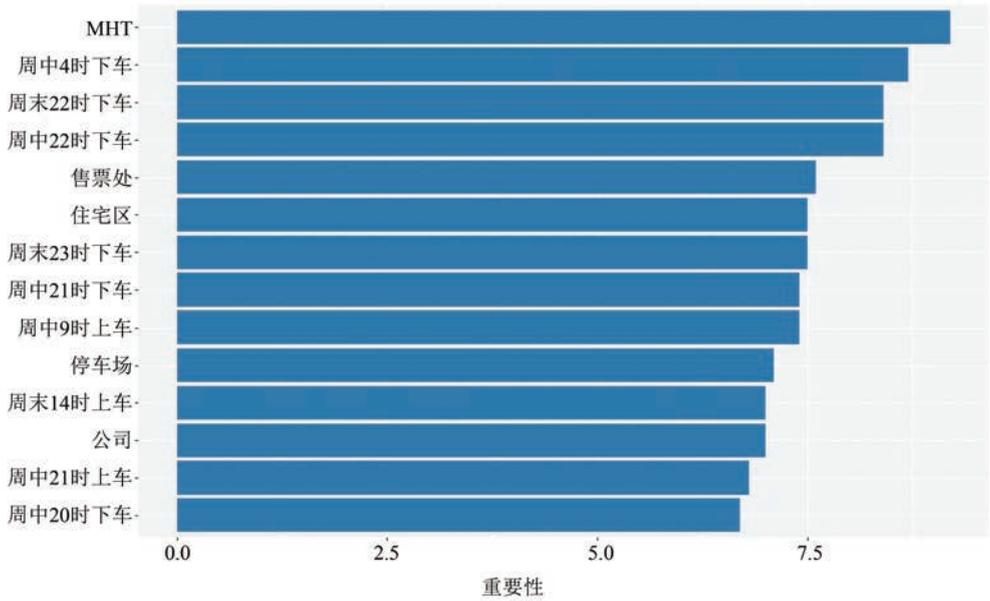
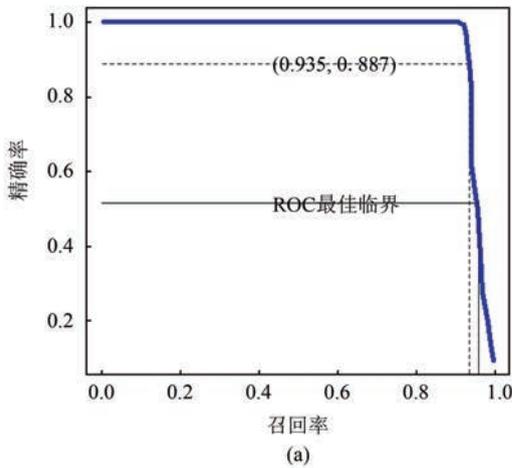
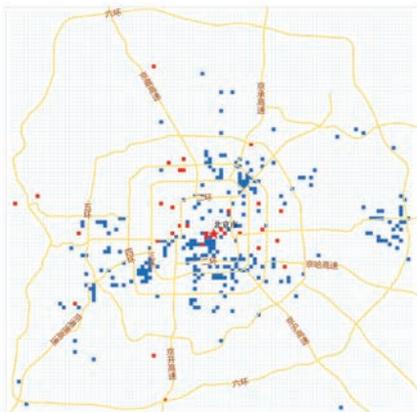


图7 解释变量重要性排序

Fig.7 Rank of Importance as Explanatory Variables



(a)



(b)

图8 基于随机森林算法预测结果的PR曲线以及城管事件点的空间分布

Fig.8 PR Curve Based on Random Forest Prediction Results and Spatial Distribution of Law Enforcement Events by City Inspectors

网格存在城管执法事件的概率,取百度街景照片进行验证(见表2),发现预测与城市真实环境比较吻合,即城管执法事件高发区域与城市脏乱差区域相对应。

4 讨论与结论

从个体尺度的时空数据出发理解地理空间时,不可避免地需要空间聚合的操作,也就是将个体数据映射到预定义好的规则或不规则的空间单元之上。目前大部分研究使用面状单元(泰森多边形、格网、行政区等)来生成聚合层面的数据集,然而很多分析单元的定义是与实际研究场景不契合的。将采样数据从小计数地理单元聚合成更大单元的数据处理问题,即是地理学著名的可塑性面积单元问题(modifiable areal unit problem, MAUP)^[26-27]。部分研究对时空数据的预处理首先是对城市空间进行等距离的规则格网划分。例如,在研究城市内部人类群体的空间分布或交互模式时,常见的划分城市空间的格网大小为250 m、500 m和1 000 m等尺度^[28-29],然后利用多源地理大数据所反映的人类移动性特征来研究格网单元内的空间分布或单元间的交互及差异性。另外,一些研究基于不规则的空间划分方式来研究城市空间内的人类移动模式。例如运用手机基站构建泰森多边形来分析手机通话数据中体现出的嵌入空间网络社区分割以及城市人口职住分离等现象^[12]。另外也有研究指出,城市空间内的物理

表 2 预测不同概率发生城管事件点的单元对应百度街景图片

Tab.2 Corresponding Baidu Street View Images for Urban Areas with Different Occurring Probabilities of Law Enforcement Events

概率	街景				
0.72					
0.54					
0.36					
0.18					
0					

移动受到道路网络的约束,街道是人们观察和理解城市意象的主导认知元素,因此将街道作为捕捉城市动态和理解城市分异格局的基本单元,在定量城市科学中也具有其独特的价值^[30]。另外,采用自下而上的聚合方式,以模拟的方式生成空间分区,从而发现地理分析单元,是一种相对更加自然的尺度发现方法,但仍然受到相关模拟参数的主观影响。总体而言,集成多源地理大数据感知城市空间分异格局时,研究单元的选择是缺乏规范的标准和参考的。不同空间单元对城市格局理解造成的影响有待跨尺度和跨单元的定量实证研究进一步探索。

表达和理解城市内部的空间分异格局是一项基础工作。多源地理大数据为地理现象的分布格局、相互作用及动态演化提供了前所未有的社会感知手段。城市是人类活动最为集中的区域,产生了多种地理大数据,并支持对于城市城市空间的理解。城市内部的分异格局是城市研究和规划所要面对的重要议题,社会感知数据提供了从“人-地-静-动”4个维度刻画城市分异格局的途径。本文梳理了不同类型大数据对于表达这4个

维度特征的支持,并通过一个实例研究展示了集成多源数据量化城市空间分异特征的应用,最后讨论了相关的理论问题。本文从时间与要素两个维度构建了刻画城市内分异格局的指标体系,梳理了常见地理大数据对于获取上述指标的支持能力,并通过一个实例展示构建多指标刻画城市分异格局的意义。结果表明,结合多源地理大数据以及传统空间数据,运用社会感知思想,有助于全面地感知城市空间分异格局。

参 考 文 献

[1] Goodchild M F. Citizens as Sensors: The World of Volunteered Geography[J]. *Geo Journal*, 2007, 69 (4): 211-221

[2] Miller H J, Goodchild M F. Data-Driven Geography [J].*Geo Journal*, 2015, 80(4): 449-461

[3] Liu Y, Liu X, Gao S, et al. Social Sensing: A New Approach to Understanding our Socioeconomic Environments [J]. *Annals of the Association of American Geographers*, 2015, 105(3): 512-530

[4] Liu Yu. Revisiting Several Basic Geographical Concepts: A Social Sensing Perspective [J]. *Acta Geo-*

- graphica Sinica*, 2016, 71(4):564-575 (刘瑜. 社会感知视角下的若干人文地理学基本问题再思考[J]. 地理学报, 2016, 71(4):564-575)
- [5] Ratti C, Frenchman D, Pulselli R M, et al. Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis[J]. *Environment and Planning B: Planning and Design*, 2006, 33(5): 727-748
- [6] Liu Y, Wang F, Xiao Y, et al. Urban Land Uses and Traffic 'Source Sink Areas': Evidence from GPS-enabled Taxi Data in Shanghai [J]. *Landscape and Urban Planning*, 2012, 106(1): 73-87
- [7] Pei T, Sobolevsky S, Ratti C, et al. A New Insight into Land Use Classification Based on Aggregated Mobile Phone Data [J]. *International Journal of Geographical Information Science*, 2014, 28(9): 1 988-2 007
- [8] Lansley G, Longley P A. The Geography of Twitter Topics in London[J]. *Computers, Environment and Urban Systems*, 2016, 58: 85-96
- [9] Jendryke M, Balz T, McClure S C, et al. Putting People in the Picture: Combining Big Location-based Social Media Data and Remote Sensing Imagery for Enhanced Contextual Urban Information in Shanghai[J]. *Computers, Environment and Urban Systems*, 2017, 62: 99-112
- [10] Liu X, He J, Yao Y, et al. Classifying Urban Land Use by Integrating Remote Sensing and Social Media Data[J]. *International Journal of Geographical Information Science*, 2017, 31(8): 1 675-1 696
- [11] Long Y, Shen Z. Geospatial Analysis to Support Urban Planning in Beijing [M]. Switzerland: GeoJournal Library, 2015
- [12] Shi L, Chi G, Liu X, et al. Human Mobility Patterns in Different Communities: A Mobile Phone Data-Based Social Network Approach[J]. *Annals of GIS*, 2015, 21(1): 15-26
- [13] Kang C, Liu Y, Ma X, et al. Towards Estimating Urban Population Distributions from Mobile Call Data[J]. *Journal of Urban Technology*. 2012, 19(4): 3-21
- [14] Deville P, Linard C, Martin S, et al. Dynamic Population Mapping Using Mobile Phone Data[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(45): 15 888-15 893
- [15] Mitchell L, Frank M, Harris, K, et al. The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place [J]. *PLoS ONE*, 2013, DOI:10.1371/journal.pone.0064417
- [16] Yang W, Mu L. GIS Analysis of Depression among Twitter Users[J]. *Applied Geography*, 2015, 60: 217-223
- [17] Liu L, Zhou B, Zhao J, et al. C-IMAGE: City Cognitive Mapping through Geo-tagged Photos[J]. *GeoJournal*, 2016, 81:817-861
- [18] Li X, Zhang C, Li W. Building Block Level Urban Land-use Information Retrieval Based on Google Street View Images[J]. *GIScience & Remote Sensing*, 2017, 54(6): 819-835
- [19] Zhong C, Arisona S M, Huang X, et al. Detecting the Dynamics of Urban Structure through Spatial Network Analysis[J]. *International Journal of Geographical Information Science*, 2014, 28(11): 2 178-2 199
- [20] Liu X, Gong L, Gong Y, et al. Revealing Travel Patterns and City Structure with Taxi Trip Data[J]. *Journal of Transport Geography*, 2015, 43: 78-90
- [21] Liu X, Kang C, Gong L, et al. Incorporating Spatial Interaction Patterns in Classifying and Understanding Urban Land Use [J]. *International Journal of Geographical Information Science*, 2016, 30(2): 334-350
- [22] Zhao Z, Shaw S L, Xu Y, et al. Understanding the Bias of Call Detail Records in Human Mobility Research[J]. *International Journal of Geographical Information Science*, 2016, 30(9):1 738-1 762
- [23] Grinnell J. Presence and Absence of Animals[M]. Berkeley: University of California Chronicle Press, 1928
- [24] Cai H, Zhan X, Zhu J, et al. Understanding Taxi Travel Patterns [J]. *Physica A: Statistical Mechanics and its Applications*, 2016, 457: 590-597
- [25] Breiman L. Random Forests[J]. *Machine Learning*, 2001, 45(1): 5-32
- [26] Openshaw S. The Modifiable Areal Unit Problem [M]. Norwick: Geo Books, 1983
- [27] Armhein C. Searching for the Elusive Aggregation Effect: Evidence from Statistical Simulations [J]. *Environment & Planning A*, 1995, 27(1):105-119
- [28] Reades J, Calabrese F, Ratti, C. Eigenplaces: Analysing Cities Using the Space-Time Structure of the Mobile Phone Network[J]. *Environment and Planning B: Planning and Design*, 2009, 36: 824-836
- [29] Toole J L, Ulm M, Gonzale Z, et al. Inferring Land Use from Mobile Phone Activity [C]. The ACM SIGKDD International Workshop on Urban Computing, Beijing, 2012
- [30] Zhu D, Wang N, Wu L, et al. Street as a Big Geo-Data Assembly and Analysis Unit in Urban Studies: A Case Study Using Beijing Taxi Data[J]. *Applied Geography*, 2017, 86: 152-164

Incorporating Multi-source Big Geo-data to Sense Spatial Heterogeneity Patterns in an Urban Space

LIU Yu^{1,2} ZHAN Zhaohui² ZHU Di² CHAI Yanwei³ MA Xiujun^{4,5} WU Lun²

1 Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Land and Resources, Shenzhen 518034, China

2 Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China

3 College of Urban and Environmental Sciences, Peking University, Beijing 100871, China

4 School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China

5 Key Laboratory of Machine Perception, Peking University, Beijing 100871, China

Abstract: Multi-source big geo-data provides us an unprecedented opportunity to investigate geographic phenomena from perspective of their spatial distribution patterns, spatial interactions and dynamic evolution. Cities are the most concentrated areas of human activities and thus massive amount of geographic big data have been produced to improve our understanding of urban spaces. The spatial heterogeneity patterns in cities is an essential topic in geographic research and urban planning. Social sensing offers an analytical framework to characterize urban spatial heterogeneity from four dimensions: human, environment, statics and dynamics. This paper summarizes the contributions of different types of big geo-data in characterizing urban features. Borrowing the concept of "niche model" from ecological studies, a case study is introduced to demonstrate the quantification of spatial heterogeneity patterns in urban space incorporating multi-source big geo-data. Theoretical issues such as unit selection are also discussed to address some related problems.

Key words: big geo-data; spatial heterogeneity; social sensing; urbanspace

First author: LIU Yu, PhD, professor, specializes in GIScience and big geo-data. E-mail: liuyu@urban.pku.edu.cn

Corresponding author: WU Lun, PhD, professor. E-mail: lwu@urban.pku.edu.cn

Foundation support: The Open Fund of Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Land and Resources, No. KF-2016-02-023; the National Natural Science Foundation of China, No. 41625003.