

动态数据空间分析的不确定性问题 ——以城市中心识别为例

周新刚¹ 乐阳^{2,3} 叶嘉安¹ 王海军^{1,4} 仲腾¹

1 香港大学城市规划与设计系,香港,999077

2 深圳大学土木工程学院,广东 深圳,518060

3 深圳大学深圳市空间信息智能感知与服务实验室,广东 深圳,518060

4 武汉大学资源与环境科学学院,湖北 武汉,430079

摘要:动态数据在空间分析中存在不确定性问题。以手机定位数据为例来识别城市中心,探索可塑性面积单元问题和不确定的地理情境问题,发现群体活动强度的空间自相关程度受到采样区域划分方式和分析单元大小的影响,地理情境的时空动态变化也会带来不确定的地理情境问题。讨论了减轻不确定性的可能方法。

关键词:不确定的地理情境;可塑性面积单元问题;手机定位数据;空间自相关;核密度估计

中图分类号:P208

文献标志码:A

动态数据存在空间异质性、时间异质性和时空自相关,在空间分析操作中会带来不确定性问题^[1]。手机定位数据作为一种典型的动态数据,从个体活动出发,刻画群体活动模式^[2-3]。然而其空间分析操作中的不确定性问题却少有研究。本文以手机定位数据为例,通过空间分析来识别城市中心,并探索其中的可塑性面积单元问题(modifiable area unit problem, MAUP)和不确定的地理情境问题(uncertain geographic context problem, UGCoP)^[2]。

1 实验数据

本文中的手机定位数据是一家主要通信运营商提供的基于基站的周期性位置更新数据,时间为1 d。周期性位置更新数据与手机通话数据的时间分辨率不同。只要手机通信服务正常,为手机提供信号服务的手机基站就会每隔0.5~1 h给手机发送一个信号,所以某个时间段手机基站记录的定位数据的总数反映群体活动的密度分布。手机定位数据包括手机用户ID、时间及经

纬度要素,手机用户的ID已经经过匿名化处理。基于手机基站位置建立的泰森多边形服务区用来估算手机用户的空间位置。手机基站服务范围的空间分布见图1,包括5 870个基站。

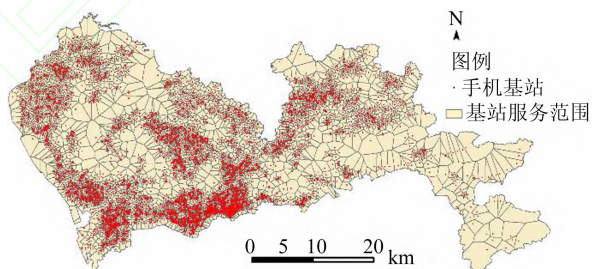


图1 手机基站服务范围

Fig. 1 Service Areas of Mobile Phone Towers

由于手机基站分布不均,根据基站建立的泰森多边形面积差异较大,而且并不是均质分布的,因此,在人口密集或人群活动频繁的地区,泰森多边形面积较小,群体活动密度非常高;而在郊区,手机基站服务范围较大,难以反映局部地区密度分布的细节,可能导致基于基站的分析存在误差甚至错误。选择合适的分析单元是相关研究的基础,也是常常被忽略的一个问题。本文以三种常

收稿日期:2014-01-21

项目来源:国家自然科学基金资助项目(41231171, 41171348);深圳市科技研发资金资助项目(JCYJ2012019111128765, JCYJ20130329144141856);CCF-腾讯犀牛鸟科研基金资助项目(CCF-TencentARG20130115)。

第一作者:周新刚,博士生。主要从事时空数据理论与方法研究。E-mail: xingang_zhou@connect.hku.hk

通讯作者:乐阳,博士,副教授。E-mail: yueyang@szu.edu.cn

用的分析单元——基站、交通小区(traffic analysis zone, TAZ)及格网为例,讨论不同分析单元对群体活动密度分布分析的影响,并识别城市中心。

2 城市中心的识别

城市中心不仅是居民就业活动的中心,也是购物和休闲活动的重要场所。城市中群体活动具有特定的模式,聚集到一定区域形成不同强度的活动中心。

把原始手机定位数据以 h 为单位进行划分,选取不同时间段的手机定位数据进行坐标系转换后,与行政区划数据进行配准。根据手机基站建立泰森多边形后,把群体活动分布聚集到基站服务范围,得到不同时间段的群体活动密度分布。

2.1 以基站为基本分析单元

利用核密度函数^[4]将基于基站的活动强度转变为核密度表面,用以识别城市主中心和次中心。搜索半径根据平均综合误差最小确定^[5],选择 5 km 作为搜索半径。把某个上班时间段(10:00~11:00)的手机定位数据聚集到手机基站来估计工作时间群体活动的密度分布,如图 2 所示。

图 2 显示了在工作时间,罗湖和福田商务中心(红色部分)是群体活动的主中心,南山、宝安中心区、沙井工业区和龙华(黄色部分)是重要的就业次中心。

2.2 以交通小区为分析单元

交通小区是研究出行活动最常用的研究单元,把基于手机基站服务范围的群体活动密度分布转化为基于交通小区的活动密度分布,如图 3 所示,包括 491 个交通小区。其中颜色较深的红颜色交通小区是工作时间群体活动密度较高的区域,是城市就业活动中心。

2.3 以格网为分析单元

为了保证分析单元的大小和形状一致,选择 2 km 格网作为分析单元研究群体活动的密度分布,如图 4 所示,共有 496 个格网,与 TAZ 平均面积接近。从绿色到红色区域,群体活动密度分布逐级递增,城市中心区的活动密度高于郊区。

2.4 不同分析单元的结果比较

以交通小区和格网为分析单元得到的城市中心总体上与基于基站服务范围核密度估计的城市中心基本一致。但是城市中心在范围大小和识别方法上有不确定性,如果对于群体活动密度分布的空间关系(如空间自相关)作进一步分析,就会

发现截然不同的结果。

3 空间自相关分析的可塑性面积单元问题

根据邻里效应,群体活动强度不仅取决于该区域的土地利用,而且受到周边环境的影响,在空间上具有一定的相关性。空间自相关分析^[6]可以反映地理要素体现出来的空间格局是集聚、离散还是随机^[7]。

1) 以交通小区为分析单元,利用 ArcGIS 10.1 中的全局空间自相关分析模块计算群体活动密度分布的全局 Moran's *I* 指数。考虑到群体活动密度分布的距离衰减模型^[4],以反距离作为空间关系的概念化方法。选择不同的查找距离进行全局空间自相关分析,得到多组 Moran's *I* 指数和显著性检验统计量 *Z*-score。Moran's *I* 指数均大于 0,而指数 *Z*-score 远大于临界值 1.96(显著性检验水平为 5%),因此整体上群体活动密度分布呈现显著的正空间自相关。为了找出就业活动的“热点”和“冷点”区域,进一步进行 Anselin 局部空间自相关分析,结果如图 5 所示。可见,罗湖商务区和福田商务区表现出很强的“高-高”关联模式,是群体活动的热点区域。大部分中心区以外的群体活动密度分布为不显著自相关。

2) 为了保证分析单元的大小和形状一致,选择 2 km 格网作为分析单元进行空间自相关分析,结果如图 6 所示,可见与基于交通小区的分析结果(图 5)大为不同。

为了比较格网大小对群体活动密度的影响,选择 1 km 格网作为分析单元研究群体活动的密度分布,如图 7 所示,共有 1 984 个千米格网。

3) 进行 Anselin 局部空间自相关分析,结果如图 8 所示。其中“高-高”关联模式的区域(红色部分)是群体就业活动的热点区域,与图 6 略有差异。

以上结果表明,群体活动强度的空间自相关程度受到采样区域划分方式和分析单元大小的影响。这是一个典型的可塑性面积单元问题^[8],即空间分析结果的有效性随基本面积单元定义的不同而发生变化,这是产生空间分析结果不确定性的原因之一。所以,选择何种分析单元最能反映真实情况在当前大数据时代是一个更加值得关注的问题。必要时,需要利用多种源数据综合分析多种尺度和划分方法,以减少 MAUP 带来的不确定性、误差甚至错误解读。

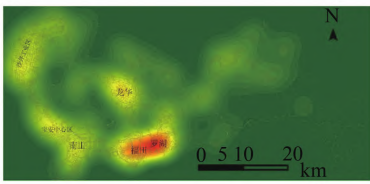


图 2 工作时间群体活动密度分布

Fig. 2 Residents' Activity Distributions at Working Hours

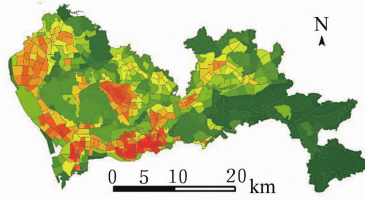


图 3 基于交通小区的群体活动密度分布

Fig. 3 Activity Density Distribution Based on TAZ

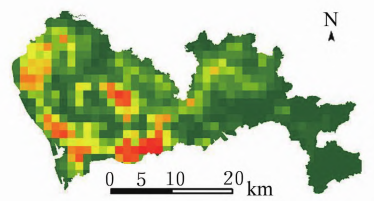


图 4 基于网格的群体活动密度分布(2 km)

Fig. 4 Activity Density Distribution Based on Grids (2 km)

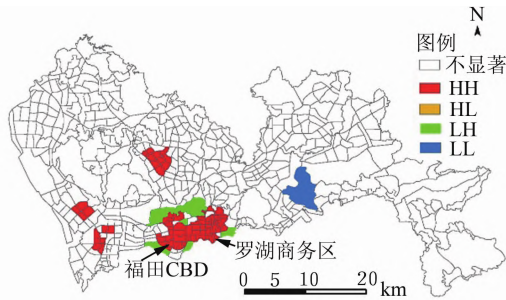


图 5 局部自相关分析结果(交通小区)

Fig. 5 Result of Local Autocorrelation (TAZ)

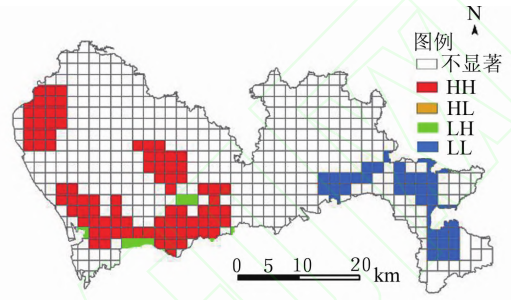


图 6 局部自相关分析结果(2 km)

Fig. 6 Result of Local Autocorrelation(2 km)

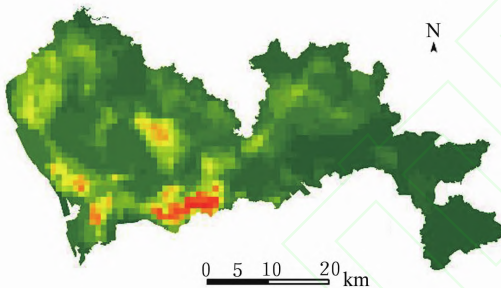


图 7 基于网格的群体活动密度分布(1 km)

Fig. 7 Activity Density Distribution Based on Grids (1 km)

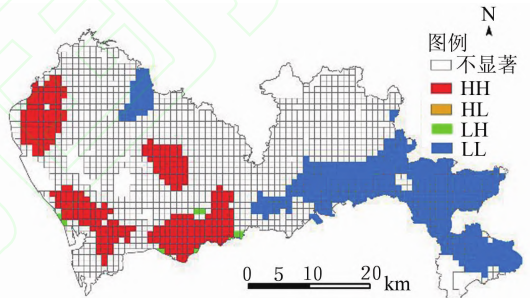


图 8 局部自相关分析结果(1 km)

Fig. 8 Result of Local Autocorrelation (1 km)

4 城市中心识别的地理情境不确定性问题

即使克服了 MAUP 问题,所划分区域内地理情境变量的变化也会带来不确定问题。仍以城市中心识别为例,中心区的范围除了受到分析单元的划分和尺度的影响,还会受到分析单元内人群活动和社会经济等因子变化的影响。Kwan 将此问题定义为 UGCoP^[9],即地理环境不确定性和情景的不确定性问题。

笔者认为,表述为地理情境的不确定性问题也许更为恰当。因为“context”是指地理分析单元内的与人类移动和行为相关的变量,既反映了地理单元的地理和物理环境,也包含了人类活动

相关的因素。计算机领域将“context-aware computing”翻译为“情境(情景)感知计算”,因为“情境(context)是由人在何处、在何时、作(doing)了什么所组成的”^[10]。

由于受到时间、空间以及行为等因素的影响,这些人类活动相关的变量具有显著的时空可变性,使得研究单元的情境和内容也具有了不确定性。以城市中心的识别为例,区域内的多数人群不是处于静止状态的,人群的移动直接影响到区域的活动强度。一些区域是全天活跃状态,一些是白天活跃,另一些可能只是晚上。如图 9 所示为工作时间段(10:00~11:00)和晚上睡觉时间段(4:00~5:00)的手机定位数据聚集到手机基站用核密度估计的居民分布密度图,可见分析区域内地理情境(群体活动强度)的昼夜变化。在晚上睡

觉时间,居民分布较分散,到上班时间聚集到城市中心,所以城市活动中心的界线也是随时间变化的。

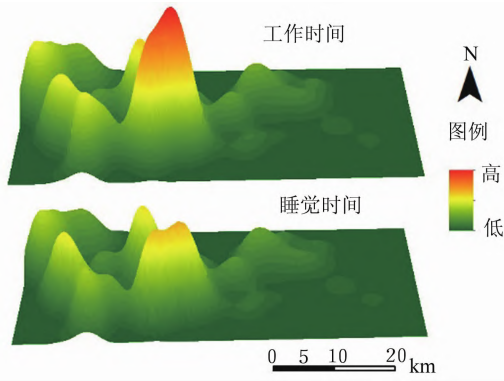


图9 不同时间段群体活动密度分布

ig. 9 Activity Density Distributions at Different Periods

选择典型主中心(华强北)和次中心(科技园)比较在不同时间段的群体活动强度,其空间位置如图10所示。分别计算并比较华强北和科技园在不同时间段的活动强度分布,如图11所示。

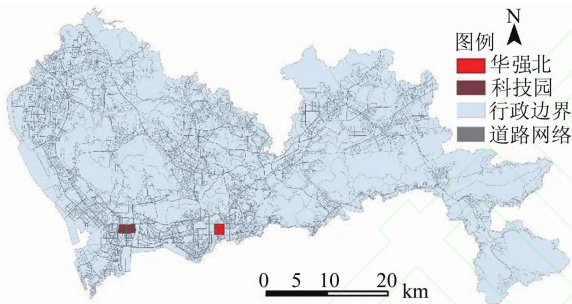


图10 典型主中心和次中心(华强北和科技园)

Fig. 10 Typical Center and Sub-center (Huaqiangbei and Science Park)

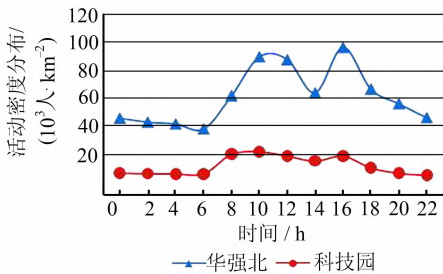


图11 活动强度随时间变化

Fig. 11 Activity Density Varies with Time

由图11可见,华强北的活动密度分布比科技园高很多,而且会随时间变化。华强北的群体活动密度分布呈“M”状,上午工作时段的活动强度迅速增加,中午下降后,下午又达到很高的活动强度,峰值接近10⁴人/km²。这是因为华强北商业、办公和居住的混合度、开发强度很高。在工作

时间,大量的工作人员聚集从事就业活动,下班后仍然有很多人进行购物、休闲等活动。科技园的群体活动密度分布在工作时间变化较平缓,比华强北的活动密度小很多,昼夜变化没有华强北显著,因为科技园以写字楼为主。所以,如果研究城市中心的范围划分是以全天活动强度的平均值来衡量的,则华强北的主中心地位和科技园的次中心地位的区别将大为缩小。

这个地理情境的不确定性问题在城市中心区的识别上较为简单,但是涉及其他对地理情境更为敏感的空间分析领域,地理情境变量的选择和动态变化所带来的不确定性是不容忽视的^[11]。

5 结 语

本文以手机定位数据为例识别城市中心,探索动态数据空间分析中存在的 uncertainty 问题。

1) 群体活动强度的空间自相关程度受到可塑性面积单元问题的影响。选择不同的区划方式和分析单元大小会得到不同的空间分析结果。所以在利用动态数据进行空间分析时,需要综合比较多种区划方式和分析单元,以减少 MAUP 带来的不确定性。

2) 除了 MAUP 问题,地理情境的时空动态变化也会给动态数据的空间分析带来不确定的地理情境问题。城市中心的活动强度在不同时间的密度分布差异很大,而且受到邻里效应的影响,所以城市中心的识别会受到地理情境变量的选择和动态变化的影响,存在不确定的地理情境问题。有必要利用多源时空数据对个体活动空间进行时空分析,以减轻不确定的地理情境问题给宏观动态数据分析带来的不确定性。

参 考 文 献

[1] Shi Wenzhong, Chen Jiangping, Zhan Qingming, et al. Reliable Spatial Analysis[J]. *Geomatics and Information Science of Wuhan University*, 2012, 37(8): 883-887(史文中, 陈江平, 詹庆明, 等. 可靠性空间分析初探[J]. *武汉大学学报·信息科学版*, 2012, 37(8): 883-887)

[2] Long Ying, Shen Zhenjiang, Mao Qizhi. Retrieving Individual Attributes from Aggregate Dataset for Urban Micro-simulation: A Preliminary Exploration [J]. *Acta Geographica Sinica*, 2011, 66(3): 416-426(龙瀛, 沈振江, 毛其智. 城市系统微观模拟中的个体数据获取新方法[J]. *地理学报*, 2011, 66(3): 416-426)

- [3] Liu Yu, Xiao Yu, Gao Song, et al. A Review of Human Mobility Research Based on Location Aware Devices[J]. *Geography and Geo-Information Science*, 2011, 27(4): 8-13(刘瑜, 肖昱, 高松, 等. 基于位置感知设备的人类移动研究综述[J]. *地理与地理信息科学*, 2011, 27(4): 8-13)
- [4] Wang F, Guldmann J M. Simulating Urban Population Density with a Gravity-based Model[J]. *Socio-economic Planning Sciences*, 1996, 30(4): 245-256
- [5] Wand M, Jones M, Smoothing K. Monographs on Statistics and Applied Probability [M]. Florida: Chapman & Hall/CRC, 1995
- [6] Goodchild M F. Spatial Autocorrelation (CATMOG 47)[M]. Norwich: Geo Books, 1986
- [7] Wang Handong, Yue Yang, Li Yuguang, et al. Spatial Correlation Analysis of Attractiveness of Commercial Facilities[J]. *Geomatics and Information Science of Wuhan University*, 2011, 36(9): 1 102-1 106(王汉东, 乐阳, 李宇光, 等. 城市商业服务设施吸引力的空间相关性分析[J]. *武汉大学学报·信息科学版*, 2011, 36(9): 1 102-1 106)
- [8] Openshaw S. The Modifiable Areal Unit Problem, Concepts and Techniques in Modern Geography [M]. Norwich: Geo Books, 1984
- [9] Kwan M P. The Uncertain Geographic Context Problem[J]. *Annals of the Association of American Geographers*, 2012, 102(5): 958-968
- [10] Erickson F, Shultz J. When is a Context? Some Issues and Methods in the Analysis of Social Competence[M]. Green J L, Wallat C. Ethnography and Language in Educational Settings. Norwood, NJ: Ablex Publishing, 1981
- [11] Kwan M P. How GIS can Help Address the Uncertain Geographic Context Problem in Social Science Research[J]. *Annals of GIS*, 2012, 18(4): 245-255

Uncertainty in Spatial Analysis of Dynamic Data —Identifying City Center

ZHOU Xingang¹ YUE Yang^{2,3} YEH Anthony Gar On¹ WANG Haijun^{1,4} ZHONG Teng¹

1 Department of Urban Planning and Design, The University of Hong Kong, Hongkong 999077, China

2 Department of Transportation Engineering, College of Civil Engineering, Shenzhen University, Shenzhen 518060, China

3 Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China

4 School of Resource and Environmental Sciences, Wuhan University, Wuhan 430079, China

Abstract: The spatial analysis of dynamic data faces uncertainty problems. This paper utilizes mobile phone positioning data as the example to identify the city center, and to explore the modifiable areal unit problem and the uncertain geographic context problem. We find that the spatial autocorrelation of activity density distribution is influenced by the zoning scheme and areal unit scale, while complex spatial and temporal dynamics also generate uncertainty. We also discuss possible means to mitigate the uncertainty problems.

Key words: uncertain geographic context problem; modifiable area unit problem; mobile phone positioning data; spatial autocorrelation; kernel density estimation

First author: ZHOU Xingang, PhD candidate, specializes in spatial-temporal data theory and analysis. E-mail: xingang_zhou@connect.hku.hk

Corresponding author: YUE Yang, PhD, associate professor. E-mail: yueyang@szu.edu.cn

Foundation support: The National Science Foundation of China, Nos. 41231171, 41171348; Shenzhen Scientific Research and Development Funding Program, Nos. JCYJ20121019111128765, JCYJ20130329144141856; CCF-Tencent ARG20130115.