

顾及道路网约束的商业设施空间点模式分析

王 腾¹ 王艳东^{1,2} 赵晓明¹ 付小康¹ 蒋波涛¹

1 武汉大学测绘遥感信息工程国家重点实验室,湖北 武汉,430079
2 地球空间信息技术协同创新中心,湖北 武汉,430079

摘 要:基于社交网络点评数据,考虑到城市的路网结构特征,采用道路网约束下的核密度估计方法研究城市商业设施的空间分布模式。同时,结合商业设施的顾客光临次数和评分对核密度估计进行加权,发现了城市商业设施顾客光临和顾客满意度在空间上的差异性,即在部分路段商业设施的分布、顾客光临分布以及顾客满意度分布不匹配。为了分析这种空间差异性,针对城市的路网结构特征以及社交网络点评数据特征,提出了道路网约束下的 G 统计量作为指标的空间自相关分析方法,通过量化分析发现了高值聚集和低值聚集的路段,反映了顾客光临和顾客满意度在空间上的分布模式,揭示了顾客满意度高或低的商业设施在空间上的聚集分布。这些分析结果为城市规划、商业设施布局、选址问题等提供了定量化的参考依据。

关键词:社交网络点评数据;商业设施;道路网约束;核密度估计;空间自相关

中图分类号:P208 **文献标志码:**A

在城市中,商业设施在经济增长与发展方面扮演着非常重要的角色,合理的商业设施布局有利于城市的经济发展,并且影响城市居民的生活质量。研究城市商业设施的空间分布模式对于城市资源的合理配置、商业设施的选址、城市经济的健康发展等有着重要的意义。

目前,商业设施的空间模式分析方法主要有样方统计、空间核密度估计、空间自相关等空间点模式分析方法。胡庆武等^[1]采用样方统计的方法对签到数据进行预处理,在空间自相关检验的基础上,提出了一种空间分析热点聚类方法来探索城市的热点和商圈。禹文豪等^[2]将核密度估计与空间自相关方法相结合,提出了一种包含量化统计指标的城市设施分布热点探测方法。但实际上,商业设施与人类的活动是紧密相关的,并且受制于道路网。Okabe 等^[3]详细论证了分析城市距离在 500 m 以内的两个对象时,基于欧氏距离和最短路径距离的空间分析结果存在明显的差别。因此,采用传统基于平面的空间分析方法来分析这些受制于道路网的商业设施得到的结论存在一定的偏见。为了应对这个问题,近年来,大量学者探索并发展了空间点模式分析方法(如道路网约

束条件下的核密度估计^[4-6]、泰森多边形^[7]、K 函数分析^[8-11]等)来分析那些被道路网约束的空间点事件。Okabe 等^[12]展示了利用平面核密度估计方法来估计道路网约束下的空间点事件的密度分布会导致有偏见的结论,提出了道路网约束下的核密度估计方法,并应用于交通事故分析。Rui 等^[13]考虑到商业设施的类型以及路网的等级,利用道路网约束下的核密度估计和 K 函数,分析了南京苏果超市的空间布局特征,并揭示了苏果超市与外国连锁超市之间的竞争关系。Ai 等^[14]提出了使用流的思想实现道路网约束下的泰森多边形,考虑到交通限制条件,能够实时准确地估计设施的服务范围。

现有商业设施的空间模式分析研究中,大部分都是着眼于商业设施的空间分布,很少涉及商业设施的顾客光临及满意度的空间分布模式研究。本文以大众点评网中采集的北京三环内美食类商业设施点评类数据为基础,采用道路网约束下的核密度估计方法研究北京三环内的美食类商业设施空间分布模式。同时,结合美食类商业设施的光临次数和评分,提出了一种基于道路网约束的加权核密度估计方法,发现了北京三环内美

食类商业设施顾客光临和顾客满意度在空间上的差异性,即在部分路段商业设施的分布、顾客光临分布以及顾客满意度分布不匹配。为了分析这种空间差异性,本文提出了利用道路网约束下的 G 统计量作为指标进行量化分析,并揭示了高值聚集或者低值聚集的路段,反映了顾客光临和顾客满意度在空间上的分布模式,揭示了顾客满意度高或低的商业设施在空间上的聚集分布。这些分析结果可以为城市资源的合理配置、商业设施的选址等问题提供定量化的参考依据。

1 数据及其特征分析

随着大数据时代的到来,社会信息的传播方式发生了重大改变,特别是随着智能手机的普及,Facebook、Twitter、微博、点评等社交网络服务网站迅速发展,这为研究人类相关活动提供了重要的数据源。大众点评网(dianping.com)是国内最大的点评类社交网络服务网站,能够为顾客提供商户信息、消费点评及消费优惠等信息服务。每一位消费的顾客可以通过该网站对消费的商业设施进行点评,给出评分(1~5 分)以及评论。本文设计网页爬虫采集了北京三环内所有的美食类商业设施数据以及这些商业设施对应的点评数据,一共有 18 512 家美食类商业设施和 2 819 261 条点评数据。每条商业设施的数据主要包含了商业设施的 ID、名称、分类以及经纬度坐标,而每条点评数据主要包含了顾客的评分、评论、时间以及对应商业设施的 ID。通过商业设施的 ID,点评数据与商业设施数据被关联起来,每一条点评数据被理解为顾客的一次光临。

2 城市商业设施分布热点探测

核密度估计是求解给定样本点集合的分布密度函数的一种方法,它从数据样本本身出发,研究数据分布特征,属于非参数检验方法之一。与基于平面的核密度估计一样,基于路网约束的核密度估计依据核函数和带宽来计算商业设施在路网上的分布密度函数,估计函数如下:

$$D(s) = \sum_{i=1}^n \frac{1}{r} k\left(\frac{d_{is}}{r}\right) \tag{1}$$

式中, $D(s)$ 代表当前位置 s 的密度估计值; r 代表核密度估计的带宽; $k(\cdot)$ 代表一个服从距离衰减的核函数; d_{is} 代表当前位置 s 到第 i 个商业设施的最短路径距离; n 代表商业设施的个数。

对于核密度估计来说,带宽 r 和核函数 $k(\cdot)$ 是两个非常重要的参数。当带宽 r 较大时,核密度估计函数比较平滑,会忽略掉部分细节。当带宽 r 较小时,核密度估计函数比较突兀,难以反映整体趋势。一些研究表明,当在分析与城市经济活动相关的商业设施时,100~300 m 是比较理想的带宽值^[15],因为这个距离接近典型城市街区的距离,经常被用来作为街区行人步行可接受的最大范围^[12]。因此,本文选择 300 m 作为核密度估计的带宽。

相比带宽 r 而言,核函数 $k(\cdot)$ 的选择则没有那么敏感^[16],常见的核函数有高斯核函数、多项式核函数、样条核函数等。在本文的实验中,采用高斯核函数:

$$k\left(\frac{d_{is}}{r}\right) = \begin{cases} \sqrt{2\pi} \exp\left(-\frac{d_{is}^2}{2r^2}\right), & 0 < d_{is} < r \\ 0, & d_{is} > r \end{cases} \tag{2}$$

基于路网约束的核密度估计与基于平面的核密度估计最大的区别在于距离的定义。在基于平面核密度估计的基础上,基于路网约束的核密度估计考虑用两点之间的最短路径距离来代替两点之间的欧氏距离。实现基于路网约束的核密度估计的关键在于将路网划分为固定的长度作为研究的最小单元,同时用最短路径距离来定义带宽 r 。具体步骤为:

1) 将路网数据的每一条路段划分为固定长度(本文选用 10 m^[17])的基本单元,记为 lixel^[18],作为研究的最小单元。

2) 将每一个商业设施约束到最近的 lixel,并将这个商业设施到当前 lixel 的距离定义为 0,而将其他 lixel 到当前 lixel 的最短路径距离作为其他 lixel 到这个商业设施的距离。

3) 根据带宽和核函数,对于每一个商业设施,确定距离该商业设施在搜索半径(即带宽 300 m)范围内的 lixel 对象,并根据核函数和该 lixel 到该商业设施的距离计算分布密度。

4) 依次遍历所有的商业设施,汇总得到每一个 lixel 的分布密度值,由于需要网络核密度估计,该方法被称为 N-KDE(network-kernel density estimation)。

采用上述方法,结合北京三环内美食类商业设施数据,通过计算得到了每一个 lixel 的分布密度值,如图 1 所示。依据密度值将所有的 lixel 分成 5 个等级,每个等级的 lixel 占总数量的 20%,分别用 5 种颜色表示,其中红色代表分布密度很大,蓝色代表分布密度很小。从图 1 中可以发现,

除了北京西站附近(图 1 中①)美食类商业设施分布密度较大以外,整体来说,北京三环内东部城区的商业设施分布密度比西部城区大。特别是在故宫(图 1 中②)周围的商业设施分布密度明显高于其他区域,原因是这块区域是北京的核心城区,几大商业中心以及旅游文化中心都分布于此。另外,在图 1 中③处出现了一片低值区域,主要是由于这块区域是陶然亭公园、天坛公园以及龙潭公园等几个城市公园所处的位置,因此,美食类商业设施的分布密度明显低于周围区域。

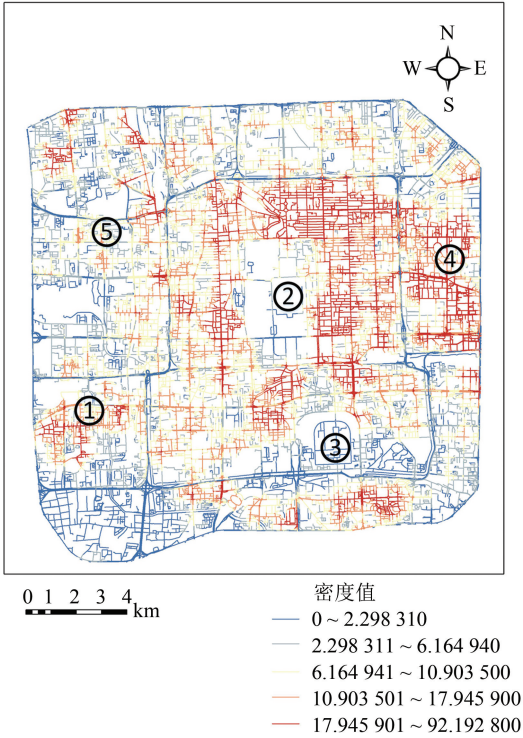


图 1 北京三环内美食类商业设施核密度分析
Fig.1 N-KDE Results for the Food Facilities Within the Third Ring Road of Beijing

除了研究商业设施的空间分布密度,还需探索商业设施的顾客光临和顾客满意度的空间分布。因此,在计算核密度时,考虑用商业设施的光临次数以及商业设施的平均评分进行加权,即:

$$D(s)=\sum_{i=1}^n\frac{1}{r}k\left(\frac{d_{is}}{r}\right)w_i\tag{3}$$

式中, w_i 表示第 i 个商业设施的光临次数或者平均评分。

分别用顾客光临次数和顾客满意度进行加权得到核密度估计结果,如图 2 所示。同样,依据核密度估计结果将所有的 lixel 分成 5 个等级,每个等级的 lixel 占总数量的 20%。结合图 1 和图 2 可以发现,不管是商业设施的空间分布密度(图 1),还是商业设施光临次数加权(图 2(a))或

者评分加权(图 2(b))的空间分布密度,故宫(图 1 和图 2 中的②)周围商业设施分布的核密度估计值都处于较高的水平,这说明这块区域商业设施的分布密度以及这些商业设施的顾客光临次数和顾客满意度都处于相当的水平。而对于图 1 和图 2 中的④,核密度估计结果存在很大的偏差。在图 1 和图 2(a)中,这块区域是属于相对的高值分布区,而在图 2(b)中,则是中值分布较多。这说明这块区域的美食类商业设施分布很多,顾客光临次数也处于平均水平,但顾客的整体满意度很低。

结合图 1 和图 2 可以发现,图 1 和图 2 中的⑤也有一些变化。在图 1 和图 2(b)中,这个区域属于相对的中值区,而在图 2(a)中却出现了大量的高值。这说明这块区域的美食类商业设施分布不是很多,同时顾客的整体满意度也属于中等水平,但这些美食类商业设施的顾客光临次数却很多。事实上,这块区域周围集中分布了几所大学(如中央财经大学、北京交通大学、北京工商大学等),这个区域的美食类商业设施可能吸引了大量的大学生来这些商业设施消费,因此顾客光临次数较多。

3 城市商业设施顾客光临与顾客满意度空间自相关分析

§ 2 探讨了利用核密度分析城市内部路网约束下的商业设施分布密度,并且通过光临次数加权和评分加权发现了商业设施的顾客光临人数和顾客满意度在空间的差异性,但考虑到顾客光临次数和顾客满意度加权的核密度估计结果仍然是依赖于商业设施的分布密度,不能直观地反映顾客光临和顾客满意度在空间上的分布。因此,本文提出基于路网约束的局部 G 统计量作为评价指标的空间自相关分析方法,进一步量化分析顾客光临和顾客满意度与空间的相关关系。

局部 G 统计量是最常用的空间自相关局部指标之一,它是一种基于距离权矩阵的指标,能探测高值聚集和低值聚集^[19]。局部 G 统计量是给定某一距离范围内邻居位置上的观测值之和与所在位置上的观测值之和的比值,能够用来识别当前位置和周围邻居之间是高值聚集还是低值聚集。为了分析城市商业设施顾客光临与顾客满意度的空间自相关性,同时考虑道路网约束,将每一条道路路段作为研究对象,并且分别将约束到每一条道路路段上所有商业设施的平均光临次数以

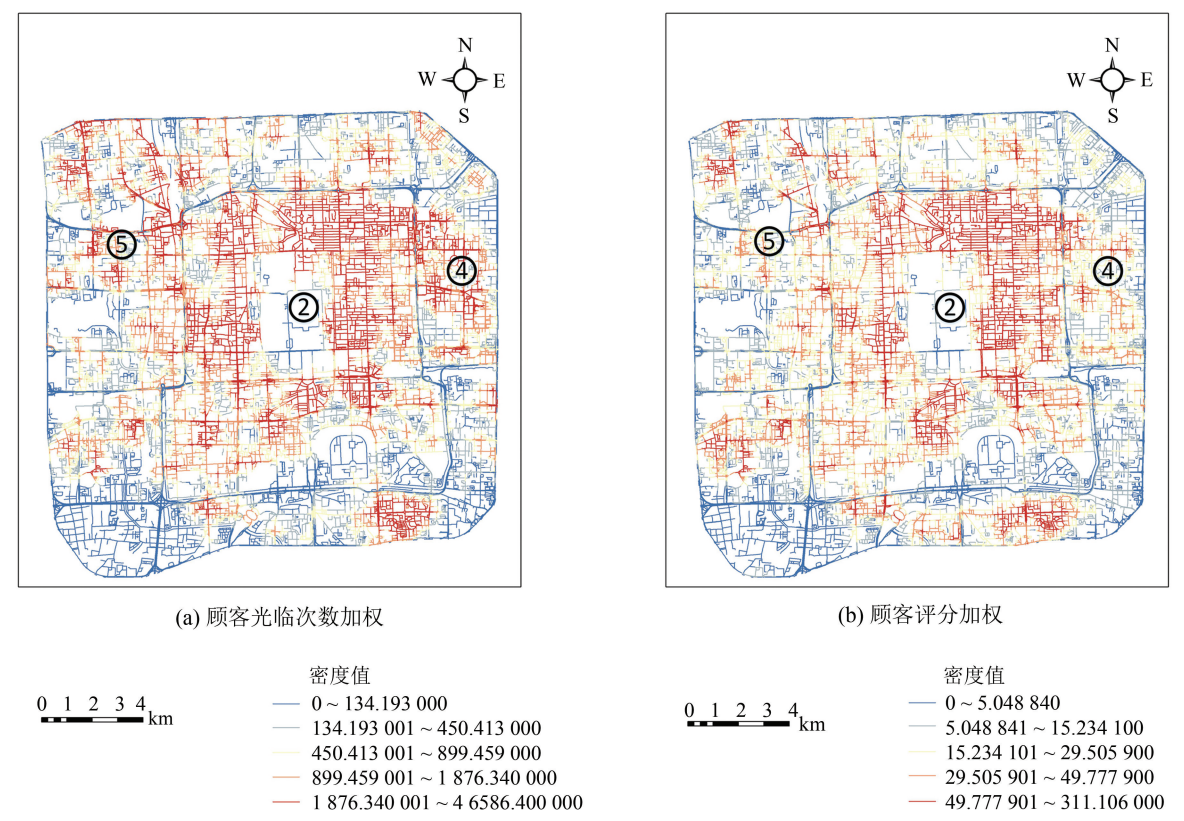


图 2 北京三环内美食类商业设施加权核密度分析

Fig.2 Weighted N-KDE Results for the Food Facilities Within the Third Ring Road of Beijing

及平均评分作为观测值,没有商业设施分布的道路对象将被去掉,不参与分析。计算公式为:

$$G_i = \frac{\sum_{j=0, j \neq i}^n w_{ij} x_j}{\sum_{j=0, j \neq i}^n x_j} \tag{4}$$

式中, G_i 代表当前道路路段*i*的*G*统计量; x_j 代表道路路段*j*对应的所有商业设施的平均光临次数或者平均评分; w_{ij} 代表道路路段*i*和道路路段*j*之间的空间权重关系; n 为道路路段数量。

本文采用基于距离的权重矩阵,同样选择 300 m 作为距离阈值,则有:

$$w_{ij} = \begin{cases} 1, & d_{ij} < 300 \text{ m} \\ 0, & \text{其他} \end{cases} \tag{5}$$

式中, d_{ij} 为道路路段*i*和道路路段*j*之间的距离。假设不存在空间自相关, G_i 服从渐进正态分布^[20],那么*G*统计量的*Z*-统计检验可以表示为:

$$Z(G_i) = \frac{G_i - E(G_i)}{\sqrt{\text{var}(G_i)}} \tag{6}$$

式中, $E(G_i)$ 表示*G*的期望; $\text{var}(G_i)$ 表示*G*的方差。

如果 $Z(G_i)$ 值为正,且非常显著,则表明位置*i*周围的值相对比较大(高于均值),出现高值空间聚集。相反,如果 $Z(G_i)$ 值为负,且非常显

著,则表明位置*i*周围的值相对比较小(低于均值),出现低值空间聚集。

分析结果如图 3 所示,其中 $Z(G_i)$ 得分大于 2.58,表示高值空间聚集,显著性水平为 0.01,置信度为 99%; $Z(G_i)$ 得分大于 1.96,表示高值空间聚集,显著性水平为 0.05,置信度为 95%; $Z(G_i)$ 得分大于 1.65,表示高值空间聚集,显著性水平为 0.1,置信度为 90%; $Z(G_i)$ 得分小于 -2.58,表示低值空间聚集,显著性水平为 0.01,置信度为 99%; $Z(G_i)$ 得分小于 -1.96,表示低值空间聚集,显著性水平为 0.05,置信度为 95%; $Z(G_i)$ 得分小于 -1.65,表示低值空间聚集,显著性水平为 0.1,置信度为 90%。

由图 3(a)可以发现,故宫以北的后海酒吧街(图 3(a)中①)以及后海以东的鼓楼大街沿线(图 3(a)中②)出现了大量的高值空间聚集,说明这些区域的顾客整体满意度比较高;而在北京西站(图 3(a)中③)、北京站(图 3(a)中④)、北京工人体育场(图 3(a)中⑤)和阜成门外(图 3(a)中⑥)等区域,则出现了低值空间聚集,说明这些区域的顾客整体满意度比较低。图 3(b)则反映了顾客光临与空间的相关性,可以发现顾客光临在空间上没有明显的低值聚集;而在西单北大街(图 3(b)中

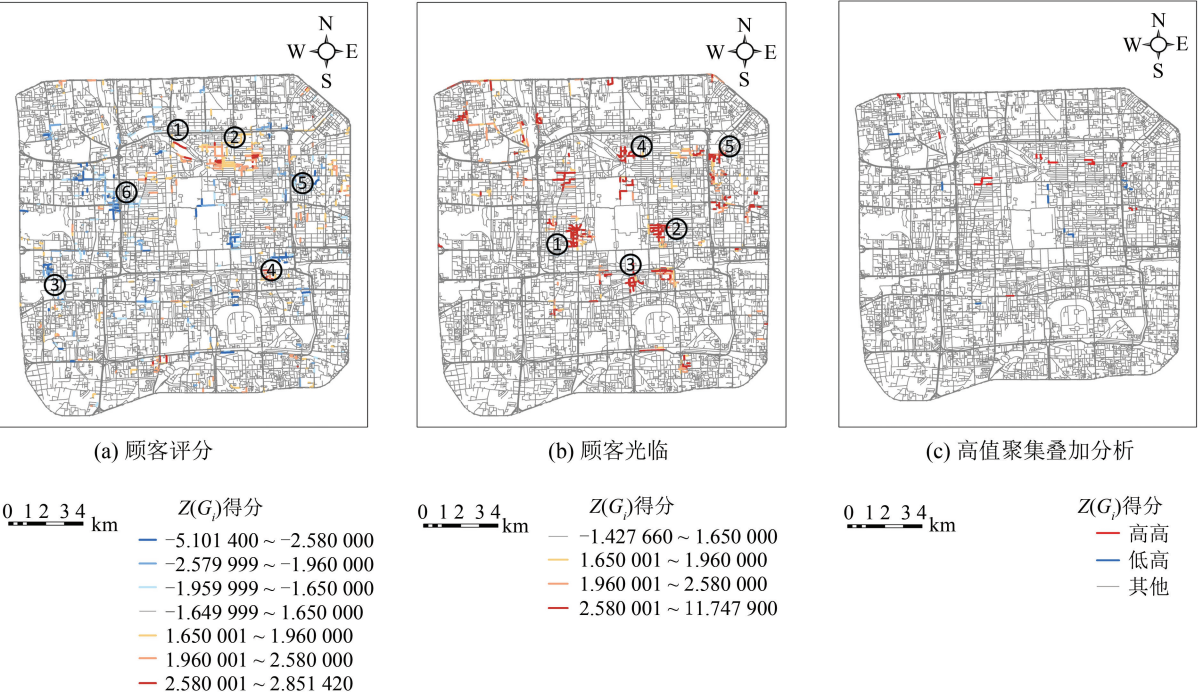


图 3 北京三环内路网约束下的美食类商业设施顾客光临与顾客满意度空间自相关分析

Fig.3 Network-Constrained Spatial Autocorrelation Analysis Between Customer's Coming and Satisfaction for the Food Facilities Within the Third Ring Road of Beijing

①)、王府井(图 3(b)中②)、前门大街(图 3(b)中③)、鼓楼(图 3(b)中④)和东直门(图 3(b)中⑤)等区域出现了显著的高值空间聚集,说明这些地方的美食类商业设施的消费人群数量很大,分布在这些路段的美食类商业设施的顾客光临次数比较多,仍然有增加美食类商业设施的发展空间。

为了探索商业设施顾客光临与顾客满意度空间自相关分析结果之间的关系,将顾客评分与顾客光临高值聚集的路段进行叠加分析,得到了 24 条顾客评分和顾客光临皆高值聚集的路段(图 3(c)中红色路段),表明这些路段的美食类商业设施可能处于一种良性发展的状态,顾客的整体满意度比较高,同时吸引了大量顾客。另外,将顾客评分低值聚集的路段与顾客光临高值聚集的路段进行叠加分析,得到了 13 条顾客评分低值聚集且顾客光临高值聚集的路段(图 3(c)中蓝色路段),说明这些路段的美食类商业设施有进一步提高的空间,顾客整体满意度比较低,但由于这些路段周围的潜在消费人群比较大,因此顾客的光临人数也比较大。相关部门应该加强监管,提高这些路段美食类商业设施的服务水平,保障消费者的权益。

4 结 语

商业设施在城市生活中有着重要的地位,会影响城市居民的生活质量,研究商业设施的空间布局以及空间点模式有着重要的意义。本文以社交网络点评数据为基础,顾及城市路网结构特征,设计了基于道路网约束的核密度估计的城市商业设施分布热点探测方法,揭示了商业设施的空间分布状况以及顾客光临和顾客满意度在空间上的差异性,并提出了道路网约束下的 G 统计量作为量化分析指标的空间自相关分析方法来探索顾客光临和顾客满意度在空间上的分布模式。这些分析结果对城市资源的合理配置、商业设施的规划等有着重要的参考价值。

本文研究也存在一些问题:①社交网络点评数据的代表性问题,社交网络的使用者往往是以年轻人为主,并不能代表整个社会群体;②社交网络点评数据也可能包含一些噪音信息,可能存在网络刷单等行为,需要对数据进行清洗;③本文研究没有考虑道路的限制条件,如道路宽度、道路方向等。这些问题将有待下一步深入研究。

参 考 文 献

- [1] Hu Qingwu, Wang Ming, LI Qingquan. Urban Hotspot and Commercial Area Exploration with Check-in Data[J]. *Acta Geodaetica et Cartographica Sinica*, 2014, 43(3):314-321(胡庆武, 王明, 李清泉. 利用位置签到数据探索城市热点与商圈[J]. 测绘学报, 2014, 43(3):314-321)
- [2] Yu Wenhao, Ai Tinghua, Yang Min, et al. Detecting “Hot Spots” of Facility POIs Based on Kernel Density Estimation and Spatial Autocorrelation Technique[J]. *Geomatics and Information Science of Wuhan University*, 2016, 41(2): 221-227(禹文豪, 艾廷华, 杨敏, 等. 利用核密度与空间自相关进行城市设施兴趣点分布热点探测[J]. 武汉大学学报·信息科学版, 2016, 41(2):221-227)
- [3] Okabe A, Satoh T, Furuta T, et al. Generalized Network Voronoi Diagrams: Concepts, Computational Methods and Applications[J]. *International Journal of Geographical Information Science*, 2007, 22(9):965-994
- [4] Li Q, Zhang T, Wang H, et al. Dynamic Accessibility Mapping Using Floating Car Data: A Network-Constrained Density Estimation Approach[J]. *Journal of Transport Geography*, 2011, 19(3): 379-393
- [5] Tang L, Kan Z, Zhang X, et al. A Network Kernel Density Estimation for Linear Features in Space-Time Analysis of Big Trace Data[J]. *International Journal of Geographical Information Science*, 2015, 30(9):1-21
- [6] Jiang Botao, Wang Yandong, Ye Xinyue. Detecting Development Pattern of Urban Business Facilities Using Reviews Data[J]. *Acta Geodaetica et Cartographica Sinica*, 2015, 44(9):1 022-1 028(蒋波涛, 王艳东, 叶信岳. 使用点评数据探测城市商业服务设施的发展规律[J]. 测绘学报, 2015, 44(9): 1 022-1 028)
- [7] She B, Zhu X, Ye X, et al. Weighted Network Voronoi Diagrams for Local Spatial Analysis[J]. *Computers Environment & Urban Systems*, 2015, 52:70-80
- [8] Yamada I, Thill J C. Comparison of Planar and Network K-Functions in Traffic Accident Analysis [J]. *Journal of Transport Geography*, 2004, 12(2):149-158
- [9] Garrocho C, Lobato J A A, Soto T L C. Calculating Intraurban Agglomeration of Economic Units with Planar and Network K-Functions[J]. *Urban Geography*, 2013, 34(2):261-286
- [10] Lamb D S, Downs J A, Lee C. The Network K-Function in Context: Examining the Effects of Network Structure on the Network K-Function [J]. *Transactions in GIS*, 2016, 20(3):448-460
- [11] Xu X, Peng Z. The K-Function Analysis of Space-Time Point Pattern on Road Network[C]. The 19th International Conference on Geoinformatics, Shanghai, China, 2011
- [12] 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
- [13] Rui Y, Yang Z, Qian T, et al. Network-Constrained and Category-Based Point Pattern Analysis for Suguo Retail Stores in Nanjing, China[J]. *International Journal of Geographical Information Science*, 2015, 30(2):1-14
- [14] Ai T, Yu W, He Y. Generation of Constrained Network Voronoi Diagram Using Linear Tessellation and Expansion Method[J]. *Computers Environment & Urban Systems*, 2015, 51:83-96
- [15] Porter M D, Reich B J. Evaluating Temporally Weighted Kernel Density Methods for Predicting the Next Event Location in a Series[J]. *Annals of GIS*, 2012, 18(3):225-240
- [16] Bil M, Andrášik R, Janoška Z. Identification of Hazardous Road Locations of Traffic Accidents by Means of Kernel Density Estimation and Cluster Significance Evaluation[J]. *Accident Analysis & Prevention*, 2013, 55(3):265-273
- [17] Nie K, Wang Z, Du Q, et al. A Network-Constrained Integrated Method for Detecting Spatial Cluster and Risk Location of Traffic Crash: A Case Study from Wuhan, China [J]. *Sustainability*, 2015, 7(7):2 662-2 677
- [18] Xie Z, Yan J. Kernel Density Estimation of Traffic Accidents in a Network Space[J]. *Computers Environment & Urban Systems*, 2008, 32(5):396-406
- [19] Ord J K, Getis A. Local Spatial Autocorrelation Statistics: Distributional Issues and an Application [J]. *Geographical Analysis*, 1995, 27(4): 286-306
- [20] Getis A, Ord J K. The Analysis of Spatial Association by Use of Distance Statistics[J]. *Geographical Analysis*, 1992, 24(3): 189-206

Network-Constrained Spatial Point Pattern Analysis for Commercial Facilities

WANG Teng¹ WANG Yandong^{1,2} ZHAO Xiaoming¹ FU Xiaokang¹ JIANG Botao¹

1 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing,
Wuhan University, Wuhan 430079, China

2 Collaborative Innovation Center of Geospatial Technology, Wuhan 430079, China

Abstract: Based on social network reviews data and considering the structural features of the road network, we apply network kernel density estimation to study the spatial distribution pattern of commercial facilities. Meanwhile, the number and satisfaction of customers corresponding to the commercial facilities are used as a weighted index for network kernel density estimation, which reveals the spatial difference in consumer satisfaction of the commercial facilities. In some road sections, the spatial distribution of commercial facilities does not match the spatial distribution of the number and satisfaction of customers. In order to further analyze the spatial difference in consumer satisfaction of the commercial facilities, aimed at the structural features of the road network and the characteristics of the social network reviews data, this paper proposes network-constrained Getis-Ord G_i^* for quantitative analysis. The high-high road segment and low-low road segment are revealed, which reflects the spatial distribution patterns of the number and satisfaction of customers, and demonstrates the aggregated distribution of the satisfied or unsatisfied commercial facilities. These results will provide a quantitative reference for urban planning, commercial facilities layout, and location problem.

Key words: social network reviews data; commercial facilities; network-constrained; kernel density estimation; spatial autocorrelation

First author: WANG Teng, PhD, specializes in spatio-temporal data mining. E-mail: wangtengvas@whu.edu.cn
Corresponding author: WANG Yandong, professor. E-mail: ydwang@whu.edu.cn
Foundation support: The National Key Research and Development Program of China, No. 2016YFB0501403; the National Natural Science Foundation of China, No. 41271399; China Special Fund for Surveying, Mapping and Geoinformation Research in the Public Interest, No. 201512015.