



网络约束下局部空间同位模式的扫描统计方法

郭艺文¹ 蔡建南¹ 陈袁芳¹ 邓敏¹ 赵斌¹

¹ 中南大学地理信息系,湖南 长沙,410083

摘要:空间同位模式挖掘旨在发现空间数据库中频繁发生在邻近位置的地理事件。由于空间异质性,地理事件在不同区域邻近出现的频繁程度亦存在差异,进而形成局部同位模式。现有局部同位模式挖掘方法多基于欧氏空间的平面假设,难以客观揭示网络空间(如城市道路)内地理事件间的局部同位规律,因此基于空间扫描统计思想,提出了一种网络约束下的局部同位模式挖掘方法。首先,发展了网络约束下的路径扩展方法,识别可能存在局部网络空间同位模式的候选路径;其次,基于网络约束下的二元泊松分布构建显著性检验的零模型,判别候选路径中局部网络空间同位模式的有效性。通过模拟实验与北京市出租车供需模式分析,发现该方法比现有方法得到的结果更精细、更客观,能够有效地挖掘网络约束下的局部同位模式。

关键词:空间数据挖掘;局部同位模式;网络约束;空间扫描统计;出租车供需模式

中图分类号:P208

文献标志码:A

现实情境中,不同类型的地理事件在位置上存在明显的空间关联,如宾馆附近易出现餐馆和便利店,出租车经常徘徊于交通站点等^[1]。此类频繁发生于邻近位置的多类事件集合被称为空间同位模式^[2-3]。近年来空间同位模式挖掘在公共安全、环境生态、移动商务及交通物流等领域有着广泛应用^[4-5]。由于地理环境存在空间异质性^[6-8],且人类活动多发生于道路网络附近,因此城市环境中空间同位模式亦呈现区域性分异及网络约束的特征,即多类地理事件仅频繁同现于城市路网的局部路径,形成局部网络空间同位模式,如在上班高峰期,乘客的出行需求对出租车的供应服务表现出较强的依赖关系,进而在住宅区附近的道路上形成局部网络空间同位模式。借助局部网络空间同位模式挖掘方法研究出租车供需模式可以为出租车的智能调度提供决策支持,为出租车行业发展以及城市道路的规划提供帮助。

根据研究区域的不同,现有空间同位模式挖掘方法可分为全局同位模式挖掘方法及局部同位模式挖掘方法。全局同位模式挖掘方法旨在全局范围内对候选同位模式的有效性进行评价,进而筛选出频繁(或显著)的空间同位模式,如

Shekhar等^[2]首先构建地理事件间的空间邻域关系,据此定义了同位模式频繁度量指标(即参与指数),根据预设的参与指数阈值识别同位模式;Huang等^[9]进一步结合空间统计量交叉 K 函数的统计含义,完善了参与指数的定义;后续研究多是在该框架基础上对计算效率的改进和拓展^[10-11];Barua等^[12]及Deng等^[1]针对上述方法依赖参与指数阈值设置的问题,分别提出了基于空间点过程的参数检验方法及基于模式重建的非参数检验方法,对候选模式的有效性进行显著性检验,识别统计显著的空间同位模式。

针对全局同位模式挖掘方法难以反映异质环境下地理事件间同位规律的问题,一些学者进行了更深入的研究,发展了一系列的局部同位模式挖掘方法。现有局部同位模式挖掘方法,先通过区域划分或空间聚类的策略识别局部同位模式的候选区域,再采用全局同位模式挖掘方法对候选区域内同位模式的有效性进行评价,如Celik等^[13]先借助四叉树的思想对研究空间进行分区,再采用全局模型识别每个分区中的局部同位模式;Eick等^[14]先使用划分聚类方法对事务型空间数据进行聚类,再提取局部同位模式的有效范围;Qian等^[15]先根据密度一致性准则,借助 k NN

收稿日期:2020-10-03

项目资助:国家重点研发计划(2018YFB1004603,2016YFB0502303);中南大学研究生自主探索创新项目(2020zzts174)。

第一作者:郭艺文,博士生,主要研究方向为时空关联模式挖掘方法及应用。yiweng.guo@csu.edu.cn

通讯作者:蔡建南,博士。jncai@outlook.com

(k -nearest neighbor)图将地理事件均匀分布的区域识别为候选区域,再在候选区域内提取频繁局部模式;蔡建南等^[16]将局部同位模式挖掘建模为针对全局非频繁模式实例的特殊空间聚类问题,并基于自适应聚类思想自动探测同位模式的频繁分布区域;Wang等^[17]发展了局部同位模式候选区域的启发式区域扩展方法,并利用假设检验方法识别显著的局部同位模式。

上述分析方法均基于欧氏空间的平面假设,采用欧氏距离度量不同事件间的距离,难以准确判断城市环境中网络约束事件间(如街头犯罪、交通事故等)的同位关系,从而导致挖掘结果出现偏差^[18]。近年来,一些学者通过定义网络划分或网络距离的策略探究了网络约束下的同位模式挖掘方法,田晶等^[19]通过划分路段的方式定义网络约束下的空间事务,进而借助统计模型推断同位模式的显著性;Yu^[20]以最短路径距离定义事件间的空间邻近关系,并发展加速方法以快速识别频繁的网络空间同位模式;Cai等^[21]基于事件在网络空间的分布特征构建假设检验模型,以此降低网络同位模式评价的主观性。上述方法无法揭示异质环境下网络约束事件间的局部同位规律。

本文提出了一种局部网络空间同位模式的扫描统计方法,并将该方法用于提取出租车供需失衡区域。

1 局部网络空间同位模式挖掘

1.1 研究策略

网络约束条件下的同位模式挖掘可为沿街设施选址、城市交通服务及产业分布等众多研究领域提供更深入、可靠的支持建议^[19-20]。现有欧氏空间的挖掘方法难以适应城市环境中地理事件的网络约束特性,可能造成局部同位模式的误判。如图1所示,实心符号表示两类事件的实例互相邻近,若不顾及网络约束对 A 、 B 两类事件实例间邻近关系的影响,则会造成局部区域中包含错误的模式实例。

本文基于扫描统计思想,将局部网络空间同位模式挖掘建模为网络约束下的二元空间扫描统计问题。首先,基于最短路径重新定义局部同位模式的兴趣度量指标与扫描统计量,将扫描单元沿路网进行扩展,识别使扫描统计量最大化的局部候选路径;其次,假设路网上不存在局部同位模式,即两类事件在全局路网上的关联强度基

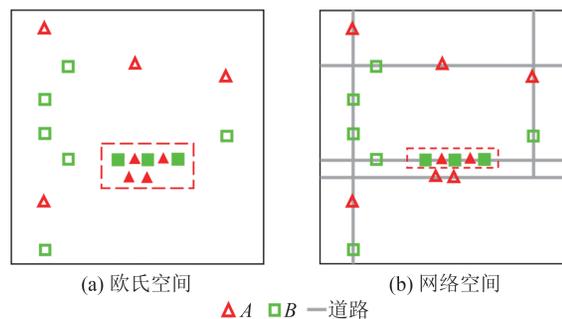


图1 欧氏空间和网络空间局部同位模式对比

Fig.1 Comparison of Regional Co-location Patterns in Euclidean Space and Network Space

本一致,基于网络约束下的二元泊松分布对该零假设进行建模,借助假设检验对候选路径上局部同位模式的有效性进行评价。

1.2 网络约束下路径扩展方法

空间扫描统计方法是识别地理事件聚集模式的重要手段,基本思想是计算研究区域中所有扫描窗口内事件聚集统计量,并筛选出具有最大统计量的局部区域,借助假设检验发现显著的聚集模式。传统的空间扫描统计方法仅针对欧氏空间内的一元空间数据,难以适用于本文的数据类型与研究任务。因此,本文定义了新的扫描统计量(即网络参与概率比),用于评价网络约束下不同地理事件间的邻近频繁程度,并将路径单元定义为基本扫描单元,发展了网络约束下的路径扩展策略,用于发现局部网络空间同位模式的候选路径。

给定研究区域内的全局路网 N , A 、 B 两类事件实例 $e(A)$ 与 $e(B)$,以及局部路径 L ,路径扩展过程中的同位模式兴趣度量指标以及扫描统计量的定义如下:

定义1:网络参与概率(NP)。局部路径 L 上 A 类事件周围 B 类事件的网络参与率定义为 L 上存在邻近的 B 类事件实例的 A 类事件实例数目与 L 上 A 类事件实例总数的比值,表达为:

$$NP(A \rightarrow B, L) = \frac{|D(e(A), e(B)) \leq d \wedge e(A) \text{ on } L|}{|e(A) \text{ on } L|} \quad (1)$$

式中, $D(e(A), e(B))$ 为 A 、 B 两类事件实例间的最短路径距离; d 为网络距离阈值。

定义2:网络参与概率比(NR)。网络空间同位模式的扫描统计量定义为 A 、 B 事件在局部路径 L 上与 L 外的网络参与概率比值,表达为:

$$NR(A \rightarrow B, L) = \frac{NP(A \rightarrow B, L)}{NP(A \rightarrow B, N - L)} \quad (2)$$

图 2(d)所示蓝色局部路径 L 上, A 类事件实例数目为 2, 其中邻近网络距离内存在 B 类事件实例的数目为 2; L 外 A 类事件实例数目为 7, 其中, 邻近距离存在 B 类事件实例的数目为 3。因此, $NP(A \rightarrow B, L) = 2/2$, $NR(A \rightarrow B, L) = 7/3$ 。

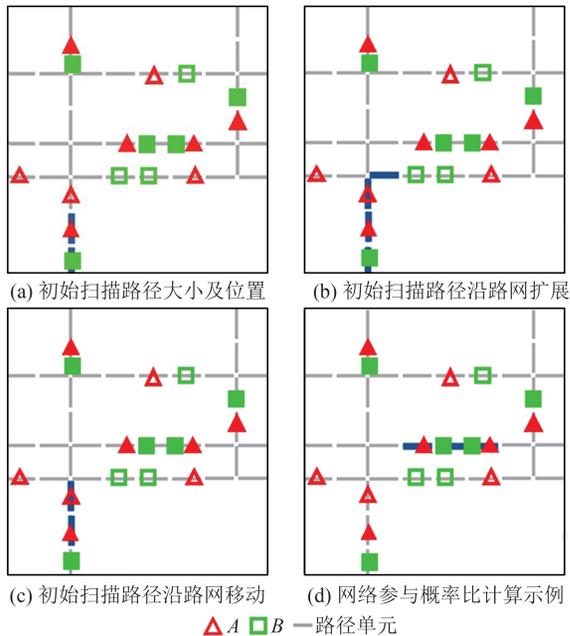


图 2 网络约束下的路径扩展方法步骤示意

Fig.2 Example of Network-Constrained Path Expansion Method

基于以上定义,网络约束下路径扩展方法的具体步骤描述如下:

1) 初始化扫描路径。对于全局路网 N 中的每个路段,将其等分为长度为 r 的路径单元;定义初始扫描路径及其初始位置(图 2(a)蓝色路径)。

2) 扩展扫描路径。将当前扫描路径进行路径扩展(见图 2(b)),随机选取与扫描路径邻接的 K 组路径单元,每组包含 M 个路径单元,选取其中使扫描统计量(即 $NR(A \rightarrow B, L)$)获得最大增量的一组路径单元进行扩展,得到扩展后的扫描路径。重复以上步骤,直至 $NR(A \rightarrow B, L)$ 取得最大值,并返回当前扫描路径。

3) 遍历扫描路径。移动初始扫描路径的位置(见图 2(c)),并返回步骤 2)。重复以上步骤,直至遍历路网中的所有可能位置,将具有最大扫描统计量的局部路径识别为该局部模式的候选路径 L_c ,并输出其对应的扫描统计量取值,记为 S_{obs} (见图 2(d))。

1.3 局部网络空间同位模式的显著性判别

为排除偶然发生的随机现象,本文采用蒙特卡罗模拟实验对候选路径内局部同位模式的显著性进行检验。假设研究区域内两类事件在全

局路网上具有一致的同现概率,即不存在具有显著高于期望水平的局部路径。为构建该零假设,本文借助二元泊松点过程^[22-23]的思想在网络空间内生成大量模拟数据,其步骤描述如下:

1) 利用期望值最大算法^[17,24]估计观测数据中 A 、 B 两类事件单独发生的概率 r_A 、 r_B 及联合发生的概率 r_{AB} 。

2) 在研究区域内的全局路网 N 上利用强度为 $r_A + r_B + r_{AB}$ 的一元泊松点过程生成随机分布的空间数据集 P 。

3) 对 P 中每个空间点的类型进行标记,被标记为独立发生的 A 类事件的比率为 $r_A / (r_A + r_B + r_{AB})$,被标记为独立发生的 B 类事件的比率为 $r_B / (r_A + r_B + r_{AB})$,被标记为与 B 类事件同现的 A 类事件比率为 $r_{AB} / (r_A + r_B + r_{AB})$ 。

选取扫描统计量网络参与概率比作为显著性判别的检验统计量。对于每一组模拟数据集,均采用路径扩展方法找到具有最大扫描统计量的局部路径及其对应的扫描统计量取值,记为 S_{sim}^i 。局部网络空间同位模式候选路径 L_c 的 p 值定义为零假设下 S_{sim}^i 大于等于观测值 S_{obs} 的概率,表达为:

$$p = \frac{|\sum_{i=1}^I S_{sim}^i \geq S_{obs}| + 1}{I + 1} \quad (3)$$

式中, I 为模拟数据集数目。给定显著性水平 α (通常设为 0.05), 如果 p 值小于等于 α , 则拒绝零假设, 认为路径 L_c 上局部同位模式的发生并非偶然事件, 并将 L_c 识别为局部同位模式的显著路径; 否则, 接受零假设, 认为研究区域内不存在局部网络空间同位模式。

2 实验分析与应用

2.1 模拟实验与比较

为验证方法的有效性,本文通过包含预设局部网络空间同位模式的模拟实验,将本文方法与现有欧氏空间显著局部同位模式的挖掘方法^[17]进行对比。实验中,将两种方法的邻近距离阈值 d 设为 5, 扫描单元组数 K 设为 5, 每组包含单元个数 M 设为 2, 模拟数据集数目设为 999。两组包含预设局部同位模式的模拟数据集(simulated dataset, SD)如图 3 所示。

模拟数据集 1: 如图 3(a) 所示, SD_1 预设有一个含有局部网络空间同位模式的路径(阴影区域)及一个非局部网络同位模式的干扰区域(虚

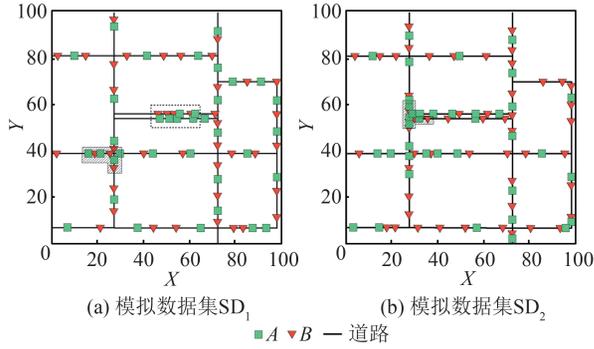


图3 模拟数据集与预设模式

Fig.3 Simulated Dataset and Preset Patterns

线区域)。

模拟数据集2:如图3(b)所示,SD₂预设有一个含有局部网络空间同位模式的路径(阴影区域),且在预设路径周围设置了若干干扰实例。

采用两种方法在两组模拟数据中挖掘到的局部同位模式如图4和图5所示。可以发现,本文方法能够有效地顾及网络约束对局部同位模式探测的影响,识别结果与预设模式一致,而现有方法缺乏建模网络约束的能力,进而可能造成结果的误判。

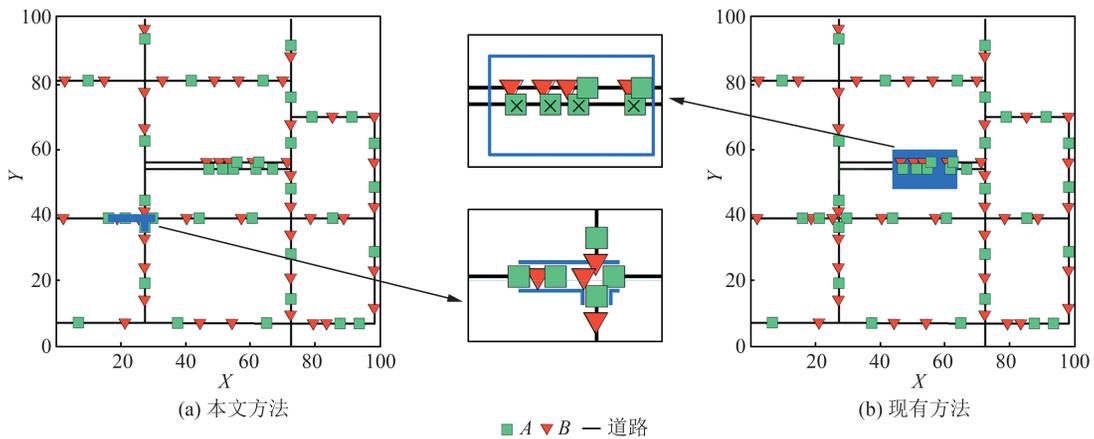


图4 SD₁模拟实验结果

Fig.4 SD₁ Simulated Experiment

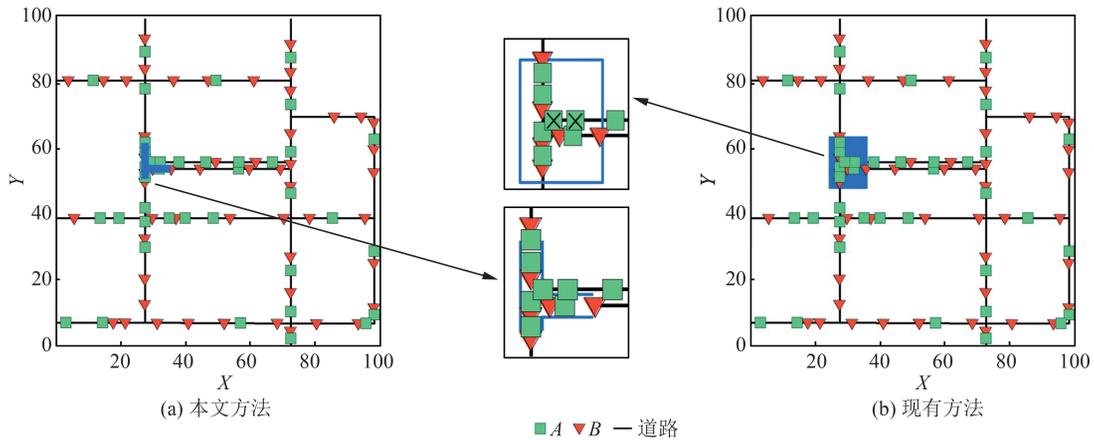


图5 SD₂模拟实验结果

Fig.5 SD₂ Simulated Experiment

SD₁结果分析:如图4所示,本文方法准确得到了预设局部同位模式的路径。现有方法识别的局部区域中,尽管两类事件实例间欧氏距离较近,具有较高的欧氏参与概率比(值为2.58),但事件实例大多分布于不同的道路(如图4(b)中“x”标注的A类事件实例与区域中的B类事件实例),网络参与概率比很低(值为0.86)。现有方法构建实例邻域时没有顾及网络空间中事件实例受到路网约束这一特性,采用欧氏距离定义事件实例间的邻近关系,高估了两类事件的同现频率。

SD₂结果分析:如图5所示,两种方法得到的局部同位模式位置相近,但本文方法通过扫描邻近路段的方式挖掘局部同位模式,避开了分布于不同道路的事件实例,能够精细地定位出现同位模式的局部路径,其网络参与概率比为2.43。相比之下,现有方法识别的区域更大,其中包含了在欧氏空间内与B类事件实例相互邻近的其他A类事件实例(如图5(b)中“x”标注的A类事件实例),但两类事件实例间的网络距离大于所设的邻近距离阈值,不存在网络约束下相互邻近的同

现关系。尽管该区域的欧氏参与概率比为 2,但其网络参与概率比仅为 1.63,小于本文所识别路径的概率比。

2.2 实际实验与分析

采用出租车供需数据进行应用分析,出租车供需分析对于促进出租车资源的优化配置和城市交通系统的协调运作具有重要意义^[25]。研究区域包含北京市五环内 6 个城区(东城区、西城区、朝阳区、海淀区、丰台区和大兴区),实验数据为北京市 28 165 辆出租车工作日 2015-11-04T08:30、

16:00 及 18:00 的轨迹数据,从中分别提取各城区不同时刻的出行需求点(即乘客上车点)和出租车供应点(即出租车空载点)。实验中,以出租车需求点定义为式(2)中的 A 类事件,出租车供应点定义为式(2)中的 B 类事件,邻近距离阈值设为 120 m,采用本文方法在 6 个区域中分别提取具有最大扫描统计量的局部网络空间同位模式,在其分布路径中出租车的供应对需求表现出高于预期的显著空间依赖性,将该路径识别为出租车供过于求的供需失衡模式,见图 6。

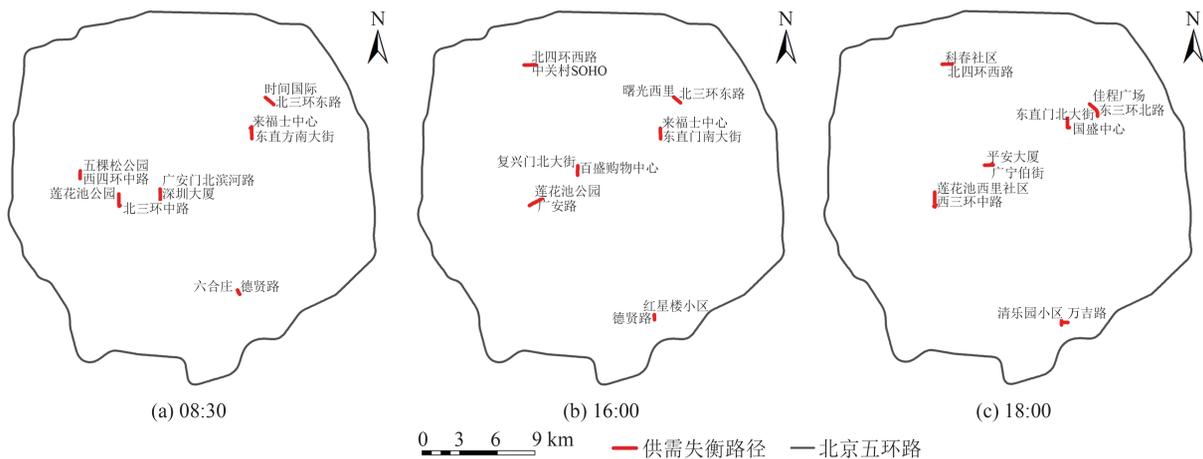


图 6 不同城区具有最大扫描统计量的显著局部网络同位模式

Fig.6 Significant Regional Network Co-location Pattern with Maximum Scan Statistic in Different Districts

08:30 出租车供需失衡路径挖掘结果如图 6(a) 所示,6 个城区的供需失衡路径大致分布在大型的休闲娱乐场所附近,包括东直门南大街来福士中心附近、广安门北滨河路深圳大厦附近、北三环东路时间国际附近、西四环中路五棵松公园附近、西三环中路莲花池公园附近及德贤路六合庄附近。16:00 出租车供需失衡路径挖掘结果图 6(b) 所示,包括东直门南大街来福士中心附近、复兴门北顺南路百盛购物中心附近、北三环东路曙光西里附近、北四环西路中关村 SOHO 附近、广安路莲花池公园附近及德贤路红星楼小区附近,各城区供需失衡路径的周边环境存在较大差别:东城区及西城区的供需失衡路径分布在商业中心附近,朝阳区、丰台区、大兴区的供需失衡路径分布在住宅社区周边,而海淀区内的供需失衡路径则与中关村的科技园相邻。18:00 出租车供需失衡路径挖掘结果如图 6(c) 所示,包括东直门北大街国盛中心附近、广宁伯街平安大厦、东三环北路佳程广场附近、北四环西路科春社区附近、西三环中路莲花池西里社区附近以及万吉路清乐园小区附近,除东城区、西城区及朝阳区内供需失衡路径在大型商场附近外,其余城区内的供需失衡路径邻近住宅小区。可见早晚高峰

时的商业娱乐中心及大型住宅区附近是出租车供应过量的多发区域,可能是因为上述区域人流量较大,出租车在此处逗留徘徊导致出租车空车过多,而在日常工作时间(如 16:00),出租车供应过量的分布则差别较大。因此,建议出租车司机可在白天多前往需求较大的住宅区或办公商厦等场所,而在傍晚尽量不在住宅区停留。此外,通过对比同一城区在不同时刻的供需失衡路径发现,5 个城区在 16:00 以及 18:00 发生出租车供需失衡路径位置均较为邻近;丰台区内 3 个时刻的供需失衡路径均位于经济活动更频繁、人流量更大的城区北部;东城区内的供需失衡路径则基本发生在东直门附近。

3 结 语

针对现有局部同位模式挖掘方法未顾及网络事件受路网约束的问题,本文基于扫描统计的思想,在候选路径探测以及显著性检验过程中充分考虑事件的网络约束特性,提出了局部网络空间同位模式的扫描统计方法。首先,通过网络约束下的路径扩展方法探测局部同位模式的候选路径;其次,以网络约束下空间点分布模型构建

零假设,甄别具有统计显著性的局部网络空间同位模式的路径。采用模拟数据与现有方法进行对比分析,发现本文方法可以有效避免由欧氏距离邻近但网络距离偏远的事件实例导致的两类事件同现频率的错误估计问题,进而可以在网络空间内更加准确地定位两类事件具有显著同位关系的分布路径。运用本文方法分析了北京市不同城区在不同时空的出租车供需失衡模式,可有效辅助出租车在城市道路网络内的资源配置。

进一步研究将集中于以下两个方面:(1)本文方法只返回具有最大扫描统计量的区域,虽然用户可以根据需求指定网络局部同位模式的数量,但发展局部网络空间同位模式多个显著路径的探测方法仍是未来探究的方向;(2)显著性检验阶段利用二元泊松分布构建零模型,而现实中数据的分布情况是未知的,需要进一步降低零模型构建对先验假设的依赖。

参 考 文 献

- [1] Deng M, He Z J, Liu Q L, et al. Multi-scale Approach to Mining Significant Spatial Co-location Patterns [J]. *Transactions in GIS*, 2017, 21 (5) : 1023-1039
- [2] Shekhar S, Huang Y. Discovering Spatial Colocation Patterns: A Summary of Results [C]//International Symposium on Spatial and Temporal Databases, Redondo Beach, USA, 2001
- [3] Bian Fuling, Wan You. A Novel Spatial Co-location Pattern Mining Algorithm Based on k -Nearest Feature Relationship [J]. *Geomatics and Information Science of Wuhan University*, 2009, 34 (3): 331-334 (边馥苓, 万幼. k -邻近空间关系下的空间同位模式挖掘算法[J]. 武汉大学学报·信息科学版, 2009, 34(3): 331-334)
- [4] Qian Feng. Research on Mining of Co-location Patterns [D]. Hangzhou: Zhejiang University, 2012 (钱烽. 同位模式挖掘研究[D]. 杭州: 浙江大学, 2012)
- [5] Chen Yuanfang, Cai Jiannan, Liu Qiliang, et al. A Nonparametric Test-Based Approach for Mining Spatio-Temporal Co-occurrence Patterns of Urban Crimes [J]. *Geomatics and Information Science of Wuhan University*, 2019, 44(12): 1883-1892 (陈袁芳, 蔡建南, 刘启亮, 等. 城市犯罪时空同现模式的非参数检验方法[J]. 武汉大学学报·信息科学版, 2019, 44(12): 1883-1892)
- [6] Kolasa J, Pickett S T A. Ecological Heterogeneity [M]. New York, USA :Springer, 1991
- [7] Goodchild M F. The Fundamental Laws of GIScience [C]// Invited Talk at University Consortium for Geographic Information Science, Santa Barbara, USA, 2003
- [8] Sha Zongyao, Li Xiaolei. Algorithm of Mining Spatial Association Data Under Spatially Heterogeneous Environment [J]. *Geomatics and Information Science of Wuhan University*, 2009, 34 (12) : 1480-1484 (沙宗尧, 李晓雷. 异质环境下的空间关联规则挖掘[J]. 武汉大学学报·信息科学版, 2009, 34(12): 1480-1484)
- [9] Huang Y, Shekhar S, Xiong H. Discovering Colocation Patterns from Spatial Data Sets: A General Approach [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2004, 16(12): 1472-1485
- [10] Yoo J S, Shekhar S, Smith J, et al. A Partial Join Approach for Mining Co-location Patterns [C]//The 12th Annual ACM International Workshop on Geographic Information Systems, Washington, USA, 2004
- [11] Yoo J S, Shekhar S. A Joinless Approach for Mining Spatial Colocation Patterns [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2006, 18 (10): 1323-1337
- [12] Barua S, Sander J. Mining Statistically Significant Co-location and Segregation Patterns [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2014, 26(5): 1185-1199
- [13] Celik M, Kang J M, Shekhar S. Zonal Co-location Pattern Discovery with Dynamic Parameters [C]//The 7th IEEE International Conference on Data Mining, Omaha, USA, 2007
- [14] Eick C F, Parmar R, Ding W, et al. Finding Regional Co-location Patterns for Sets of Continuous Variables in Spatial Datasets [C]//The 16th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, Irvine, USA, 2008
- [15] Qian F, Chiew K, He Q M, et al. Mining Regional Co-location Patterns with k NNG [J]. *Journal of Intelligent Information Systems*, 2014, 42(3): 485-505
- [16] Cai Jiannan, Liu Qiliang, Xu Feng, et al. An Adaptive Method for Mining Hierarchical Spatial Co-location Patterns [J]. *Acta Geodaetica et Cartographica Sinica*, 2016, 45(4): 475-485 (蔡建南, 刘启亮, 徐枫, 等. 多层次空间同位模式自适应挖掘方法[J]. 测绘学报, 2016, 45(4): 475-485)
- [17] Wang S, Huang Y, Wang X S. Regional Co-locations of Arbitrary Shapes [M]//Advances in Spatial and Temporal Databases, Berlin, Heidelberg: Springer, 2013
- [18] Xie Z X, Yan J. Kernel Density Estimation of Traf-

- fic Accidents in a Network Space [J]. *Computers, Environment and Urban Systems*, 2008, 32 (5) : 396-406
- [19] Tian Jing, Wang Yiheng, Yan Fen, et al. A New Method for Mining Co-location Patterns Between Network Spatial Phenomena[J]. *Geomatics and Information Science of Wuhan University*, 2015, 40 (5): 652-660 (田晶, 王一恒, 颜芬, 等. 一种网络空间现象同位模式挖掘的新方法[J]. 武汉大学学报·信息科学版, 2015, 40(5): 652-660)
- [20] Yu W H. Spatial Co-location Pattern Mining for Location-Based Services in Road Networks[J]. *Expert Systems with Applications*, 2016, 46: 324-335
- [21] Cai J N, Deng M, Liu Q L, et al. Nonparametric Significance Test for Discovery of Network - Constrained Spatial Co-location Patterns[J]. *Geographical Analysis*, 2019, 51(1): 3-22
- [22] Paloheimo J E. A Spatial Bivariate Poisson Distribution[J]. *Biometrika*, 1972, 59(2): 489-492
- [23] Babaei A, Jabbari B. Distance Distribution of Bivariate Poisson Network Nodes[J]. *IEEE Communications Letters*, 2010, 14(9): 848-850
- [24] Karlis D. An EM Algorithm for Multivariate Poisson Distribution and Related Models[J]. *Journal of Applied Statistics*, 2003, 30(1): 63-77
- [25] Feng Xiaomei. Study on the Development Scale of Taxi Under the Balance of Supply and Demand[D]. Chengdu: Southwest Jiaotong University, 2010 (冯晓梅. 供需平衡状态下的出租车发展规模研究[D]. 成都: 西南交通大学, 2010)

Spatial Scan Statistic Method for Discovering Regional Network Co-location Patterns

GUO Yiwen¹ CAI Jiannan¹ CHEN Yuanfang¹ DENG Min¹ ZHAO Bin¹

¹ Department of Geo-Informatics, Central South University, Changsha 410083, China

Abstract: Objectives: Currently, most methods for mining regional co-location patterns focus on the planar geospatial space which can hardly support the analysis on the network space such as urban roads. Therefore, a regional network co-location pattern mining method is proposed based on spatial scan statistics. **Methods:** A network-constrained path expansion method is developed to detect the candidate paths where co-location patterns could occur. These candidates are further validated using significance tests, where the null model is constructed using a network-constrained bivariate Poisson distribution. **Results:** Experiment results of simulated data and taxi datasets show that the proposed method is more effective for discovering regional co-location patterns on the network space than a baseline method. Compared with traditional methods for discovering regional co-location patterns, our proposed method fully considers the network constraint properties of events, in both phases of the candidate path detection and the significance testing. **Conclusions:** Our method can effectively assist the resource allocation of taxis within the urban road network in Beijing by analyzing the taxi supply-demand patterns in different urban districts.

Key words: spatial data mining; regional co-location patterns; network-constrained; spatial scan statistics; taxi supply-demand patterns

First author: GUO Yiwen, PhD candidate, specializes in spatial-temporal association patterns mining and applications. E-mail: yiwen.guo@csu.edu.cn

Corresponding author: CAI Jiannan, PhD. E-mail: jncai@outlook.com

Foundation support: The National Key Research and Development of China(2018YFB1004603, 2016YFB0502303); the Postgraduate Research and Innovation Foundation of Central South University(2020zzts174).

引文格式: GUO Yiwen, CAI Jiannan, CHEN Yuanfang, et al. Spatial Scan Statistic Method for Discovering Regional Network Co-location Patterns[J]. *Geomatics and Information Science of Wuhan University*, 2022, 47(9): 1383-1389. DOI:10.13203/j.whugis20200177 (郭艺文, 蔡建南, 陈袁芳, 等. 网络约束下局部空间同位模式的扫描统计方法[J]. 武汉大学学报·信息科学版, 2022, 47(9): 1383-1389. DOI:10.13203/j.whugis20200177)