



武汉大学学报(信息科学版)

Geomatics and Information Science of Wuhan University

ISSN 1671-8860, CN 42-1676/TN

《武汉大学学报(信息科学版)》网络首发论文

题目: 网络约束下局部空间同位模式的扫描统计方法
作者: 郭艺文, 蔡建南, 陈袁芳, 邓敏, 赵斌
DOI: 10.13203/j.whugis20200177
网络首发日期: 2021-04-13
引用格式: 郭艺文, 蔡建南, 陈袁芳, 邓敏, 赵斌. 网络约束下局部空间同位模式的扫描统计方法. 武汉大学学报(信息科学版).
<https://doi.org/10.13203/j.whugis20200177>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式(包括网络呈现版式)排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊(光盘版)》电子杂志社有限公司签约, 在《中国学术期刊(网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊(网络版)》是国家新闻出版广电总局批准的网络连续型出版物(ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

DOI:10.13203/j.whugis20200177

引用格式：

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, 2021, DOI: 10.13203/j.whugis20200177 (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,2021, DOI: 10.13203/j.whugis20200177)

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。为此，本文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；进一步，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，发现本文方法比现有方法得到的结果更加精细且客观，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

文献标识码：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，例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、移动商务以及交通物流等领域有着广泛应用^[4,5]。然而，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。例如，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进一步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以及局部同

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。例如：Sherkhar 等人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^[2]。Huang 等人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^[9]。后续研究多是在该框架基础上对计算效率的改进和拓展^[10-12]。Barua 等人^[13]以及 Deng 等人^[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以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，一些学者进行了更为深入的研究，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首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，进而采用全局同位模式挖掘方法对

项目资助：国家重点研发计划“云计算和大数据”重点专项(2018YFB1004603)；国家重点研发计划(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，博士生。jiannan.cai@csu.edu.cn

候选区域内同位模式的有效性进行评价。例如, Celik 等人借助四叉树的思想对研究空间进行分区, 随后采用全局模型识别每个分区中的局部同位模式^[14]; Eick 等人使用划分聚类方法对事务型空间数据进行聚类, 从而提取局部同位模式的有效范围^[15]; Qian 等人根据密度一致性准则, 借助 kNN 图将地理事件均匀分布的区域识别为候选区域, 进而在候选区域内提取频繁局部模式^[16]; Song 等人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^[17]。

然而, 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^[18])的同位关系, 从而导致挖掘结果出现偏差^[19,20]。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^[21]; Yu 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^[22]; Cai 等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^[23]。然而, 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。为此, 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, 下面将对本文的研究策略与方法进行阐述。

1 基于扫描统计的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、城市交通服务以及产业分布等众多研究领域提供更深入、可靠的支持建议^[21,22]。然而, 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, 实心符号表示两个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。

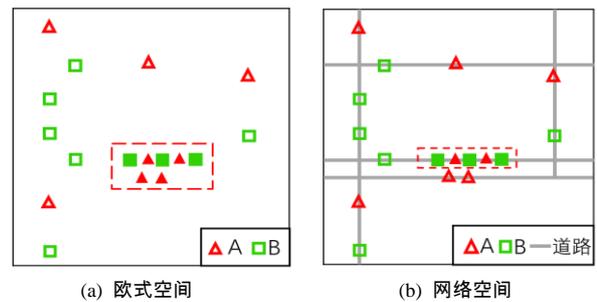


图 1 欧氏空间和网络空间局部同位模式对比

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

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。具体地, 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; 进一步, 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^[24], 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, 进而借助假设检验发现显著的聚集模式。然而, 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。为此, 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: 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：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：

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

例如，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， $NP(A \rightarrow B, L)=2/2$ ， $NR(A \rightarrow B, L)=7/3$ 。

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（如图 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)），随机选取与扫描单元邻接的 K 组路径单元，每组路径单元包含 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 CL ，并输出其对应的扫描统计量取值，记为 MNR_{obs} （见图 2 (d)）。

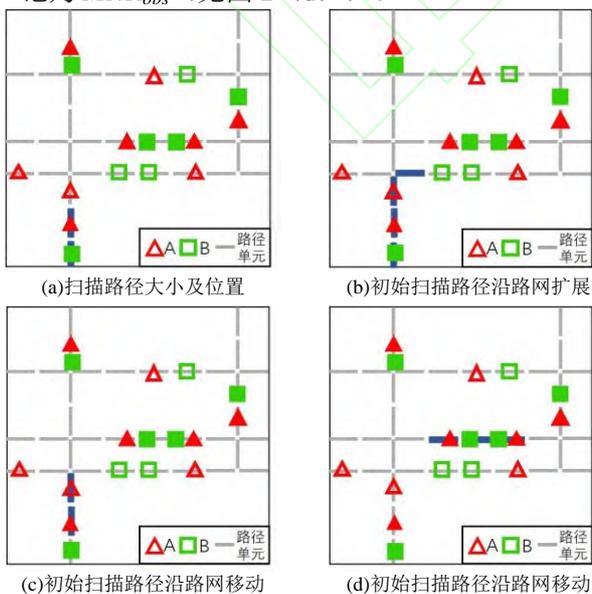


图 2 网络约束下的路径扩展方法步骤示意

Fig.2 Example of Network-Constrained Path Expansion Method

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。首先，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^[25,26]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^[17,27]估计观测数据中 A 、 B 两类事件的单独发生的概率 r_A 、 r_B ，以及联合发生的概率 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 $r_{AB} / (r_A + r_B + r_{AB})$ ；

进而，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 2.2 节所述的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 MNR_{sim}^i 。局部网络空间同位模式候选路径 CL 的 p 值定义为零假设下 MNR_{sim}^i 大于等于观测值 MNR_{obs} 的概率，表达为：

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

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，认为该路径 CL 上局部同位模式的发生并非偶然事件，并将 CL 识别为局部同位模式的显著路径；否则，接收零假设，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，本文首先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^[15]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。如图 3 所示，设计了两组包含预设局部网络同位模

式的模拟数据集：

模拟数据 1 (SD1): 如图 3 (a) 所示，预设有一个含有局部网络同位模式的路径（阴影区域）以及一个非局部网络同位模式的干扰区域（虚线区域）。

模拟数据 2 (SD2): 如图 3 (b) 所示，SD2 预设一个含有局部网络同位模式的路径（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，具体分析如下：

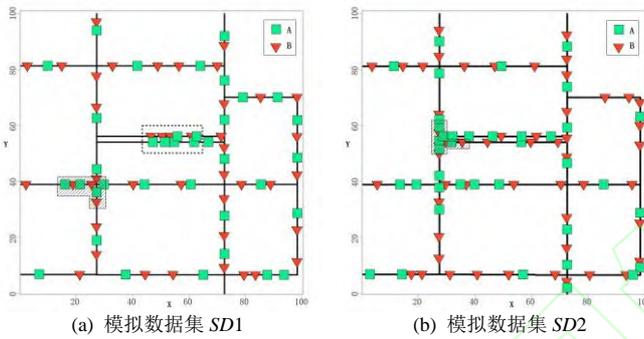


图 3 模拟数据集与预设模式
Fig.3 Simulated Dataset and Preset Patterns

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网络同位模式，避开了分布于不同道路的事件实例，能够精细地定位出现网络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它 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。

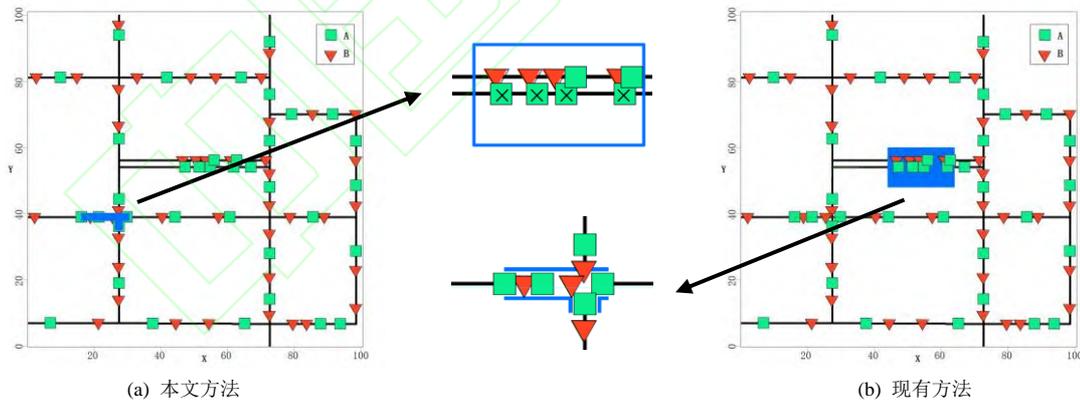
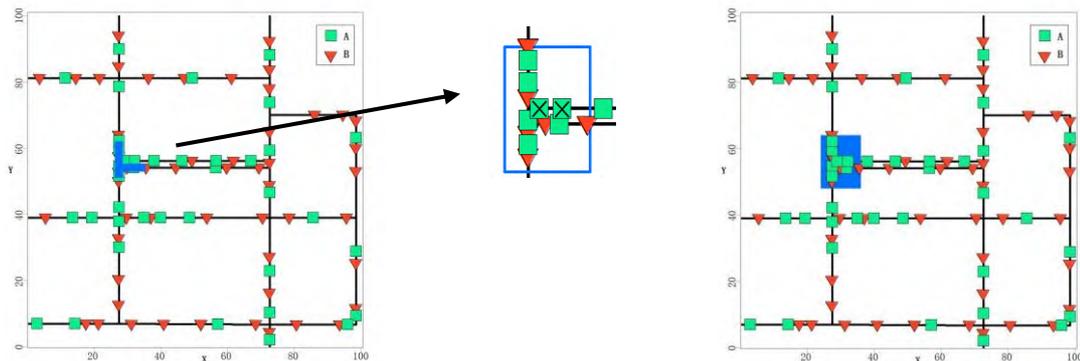
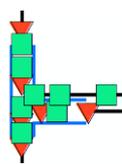


图 4 SD1 模拟实验结果
Fig.4 the First Simulated Experiment



(a) 本文方法



(b) 现有方法

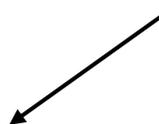


图 5 SD2 模拟实验结果
Fig.5 the Second Simulated Experiment

2.2 实际实验与分析

进一步,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^[28,29]。研究区域包含北京市五环内六个城区(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 2015 年 11 月 4 日(工作日)上午 8 时 30 分、下午 4 时以及下午 6 时的轨迹数据,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,采用本文方法在六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,进一步将该路径识别为出租车供过于求的供需失衡模式。

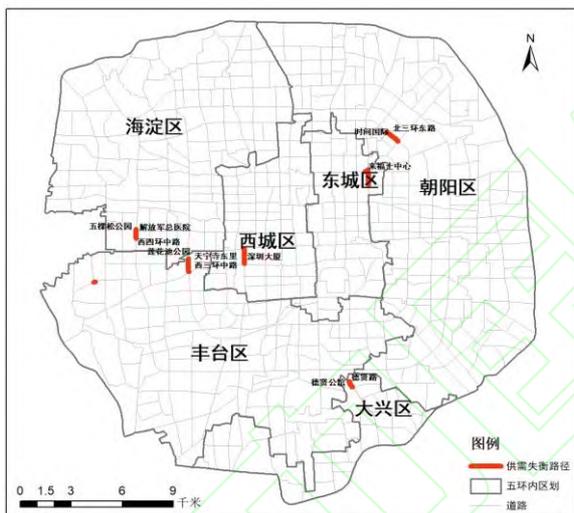


图 6 上午 8 时 30 分不同城区具有最大扫描统计量的显著局部网络同位模式

Fig.6 Significant Regional Network Co-location Pattern with Maximum Scan Statistic at 8:30 a.m.

上午 8 时 30 分出租车供需失衡路径挖掘结果如图 6 所示,六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、西二环深圳大厦附近、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以及德贤路德贤公馆附近,如莲花池公园、五棵松公园、东直门等;下午 4 时出租车供需失衡路径挖掘结果图 7 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



图 7 下午 4 时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.7 Significant Regional Network Co-location Pattern with Maximum Scan Statistic at 4:00 p.m.

近、北四环西路中关村 SOHO 附近、广安路莲花池公园附近以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;下午 6 时出租车供需失衡路径挖掘结果图 8 所示,包括东直门北大街国盛中心附近、广宁伯街平安大厦及投资广场附近、东三环佳程广场附近、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、西城区以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。

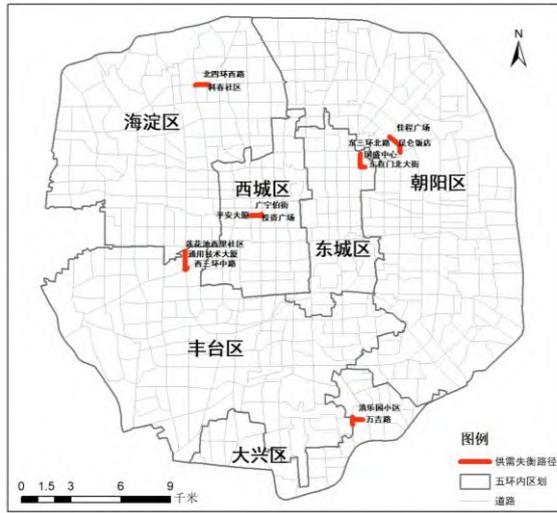


图8 下午6时不同城区具有最大扫描统计量的显著局部网络同位模式

Fig.8 Significant Regional Network Co-location Pattern with Maximum Scan Statistic at 6:00 p.m.

由此可见，早晚高峰时的商业娱乐中心以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（如下午4时）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，五个城区在下午4时以及下午6时发生出租车供需失衡路径位置均较为邻近；丰台区内三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 总结与讨论

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，进一步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 (References)

- [1] Deng M, He Z, Liu Q, 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. Co-location Rules Mining: A Summary of Results[C]. Proceedings of the 7th Spatio-temporal Symposium on Databases. 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, Deng Min, Zhang Xueying. 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, 邓敏, 张雪英. 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]. Springer New York, 1991.
- [7] Goodchild M F. The fundamental laws of GIScience[C].

- Invited talk at University Consortium for Geographic Information Science, University of California, Santa Barbara, 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 & 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]. Proceedings of the 12th annual ACM international workshop on Geographic information systems. ACM, 2004: 241-249.
- [11] Yoo J S, Shekhar S. A Joinless Approach for Mining Spatial Colocation Patterns[M]. IEEE Educational Activities Department, 2006.
- [12] Xiao X, Xie X, Luo Q, et al. Density Based Co-location Pattern Discovery[C]. Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems. ACM, 2008: 29.
- [13] Barua S, Sander J. Mining Statistically Significant Co-location and Segregation Patterns[J]. IEEE Transactions on Knowledge & Data Engineering, 2014, 26(5): 1185-1199.
- [14] Celik M, Kang J M, Shekhar S. Zonal Co-location Pattern Discovery with Dynamic Parameters[J]. Seventh IEEE International Conference on Data Mining, 2007: 433-438.
- [15] Eick C F, Parmar R, Ding W, et al. Finding Regional Co-location Patterns for Sets of Continuous Variables in Spatial Datasets[C]. ACM Sigspatial International Conference on Advances in Geographic Information Systems. ACM, 2008:20-193.
- [16] Qian F, Chiew K, He Q, et al. Mining Regional Co-location Patterns with kNNG[J]. Journal of Intelligent Information Systems, 2014, 42(3): 485-505.
- [17] Wang S, Huang Y, Wang X S. Regional Co-locations of Arbitrary Shapes[C]. International Conference on Advances in Spatial and Temporal Databases. 2013: 19-37.
- [18] Okabe A, Sugihara K. Spatial Analysis along Networks: Statistical and Computational Methods[M]. Wiley, 2012.
- [19] Xie Z, Yan J. Kernel Density Estimation of traffic accidents in a network space[J]. Computers Environment & Urban Systems, 2008, 32(5): 396-406.
- [20] Shiode S. Street-level Spatial Scan Statistic and STAC for Analysing Street Crime Concentrations. [J]. Transactions in GIS, 2011, 15(3): 365-383.
- [21] 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.)
- [22] Yu W. Spatial Co-location Pattern Mining for Location-based Services in Road Networks[J]. Expert Systems with Applications, 2016, 46: 324-335.
- [23] Cai J, Deng M, Liu Q, He Z, et al. Nonparametric Significance Test for Discovery of Network-Constrained Spatial Colocation Patterns[J]. Geographical analysis, 2019, 51(1): 3-22.
- [24] M. Kulldorff. 1997. A spatial scan statistic[J]. Communications in Statistics: Theory and Methods 26(6), 1481-1496.
- [25] Paloheimo J E. A Spatial Bivariate Poisson Distribution[J]. Biometrika, 1972, 59(2): 489-492.
- [26] Babaei A, Jabbari B. Distance Distribution of Bivariate Poisson Network Nodes[J]. IEEE Communications Letters, 2010, 14(9): 0-850.
- [27] Dimitris Karlis. An EM Algorithm for Multivariate Poisson Distribution and Related Models[J]. Journal of Applied Statistics, 2010, 30(1): 63-77.
- [28] Shuai Zhaohui. Study on Allocation of Taxi Resources in Jiaying City[D]. Shanghai: Shanghai Jiao Tong University, 2011. (帅朝晖. 城市出租车资源配置研究——以浙江嘉兴为例[D]. 上海: 上海交通大学, 2011.)
- [29] Feng Xiaomei. Study on the Development Scale of Taxi Under the Balance of Supply and Demand[D]. Chengdu: Southwest Jiaotong University, 2010. (冯

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

*Guo Yiwen*¹ *Cai Jiannan*¹ *Chen Yuanfang*¹ *Deng Min*¹ *Zhao Bin*¹

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

Abstract: Spatial co-location pattern discovery aims to mining subsets of geographic features which are frequently located in close proximity. Due to spatial heterogeneity, the frequency of different features being co-located usually varies across the space, thus forming the regional co-location patterns. Currently, most methods for discovering regional co-location patterns focus on planar geospatial spaces which can hardly support the corresponding analysis on network space such as urban roads. Therefore, this paper proposes a regional network co-location pattern mining method based on spatial scan statistics. First, 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 a significance test, where the null model is constructed using a network-constrained bivariate Poisson distribution. Experiments using 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.

Keywords: 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 candidate. E-mail: jiannan.cai@csu.edu.cn

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

网络首发:

标题: 网络约束下局部空间同位模式的扫描统计方法

作者: 郭艺文,蔡建南,陈袁芳,邓敏,赵斌

收稿日期: 2020-11-02

DOI:10.13203/j.whugis20200177

引用格式:

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, 2021, DOI: 10.13203/j.whugis20200177(Gou 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,2021, DOI: 10.13203/j.whugis20200177)

网络首发文章内容和格式与正式出版会有细微差别, 请以正式出版文件为准!

您感兴趣的其他相关论文:

显著局部空间同位模式自动探测方法

徐枫, 蔡建南, 刘启亮, 何占军, 邓敏

武汉大学学报 信息科学版, 2018, 43(10): 1538-1545

<http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20170008>

城市犯罪时空现模式的非参数检验方法

陈袁芳, 蔡建南, 刘启亮, 邓敏, 张雪英

武汉大学学报·信息科学版, 2019, 44(12): 1883-1892

<http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20180112>

基于 Spark 计算框架的路网核密度估计并行算法

郭宇达, 朱欣焰, 芮维, 余冰

武汉大学学报 信息科学版, 2020, 45(2): 289-295

<http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20180473>

使用时序出租车轨迹识别多层次城市功能结构

姚尧, 张亚涛, 关庆锋, 麦可, 张金宝

武汉大学学报 信息科学版, 2019, 44(6): 875-884

<http://ch.whu.edu.cn/cn/article/doi/10.13203/j.whugis20170111>