

城市交通状态的空间依赖性和空间异质性特征分析¹

邹海翔 乐阳 李清泉

【摘要】城市交通状态具有显著的空间特征：一方面邻近路段之间存在着关联，另一方面不同区域之间又存在着差异，但是现有的研究很少对这种特征进行空间上的描述。本文从空间角度分析城市宏观交通特征，首先针对城市路网几何形态和交通数据的关系提出了适合于城市交通数据分析的空间模型；然后从长期的浮动车数据（FCD）中提取的路段行程速度作为描述交通状态的实验数据，分别采用空间自相关（spatial autocorrelation）和核密度估计（kernel density estimation）方法分析了路网行程速度的空间依赖性（spatial dependency）和空间异质性（spatial heterogeneity），并详细地分析了其在典型时段的空间分布特征。实验结果显示城市交通状态的关联性随着路段间距的增长而衰减，并且可以得出路段影响范围的量化分析结果；城市交通状态的空间分布特征随时间变化，而且受路网空间形态和功能的影响，这使得城市路网在自由空间中呈现具有集聚特征的局部路网，通过本文的方法可以划定这些局部路网的区域。

【关键词】交通地理信息系统；交通数据分析；空间分析；空间依赖性；空间异质性

1 引言

交通数据分析可以为交通规划和决策提供有用信息，因此一直都是交通领域的热点研究问题。交通领域中传统的分析方法是将交通数据看作时间序列数据，以研究单一路段交通数据的时序特征和模式为主要的研究目的^[1]。这种方法在高速公路、城市快速路等相对封闭道路的研究中取得了较好的成果。但是交通流在城市路网中具有较大的自由度，道路之间存在着相互的影响，因此城市管理者在决策过程中更为关心的是路网的整体运行情况，例如交通拥堵发生的地点和扩散范围，交通事故的影响范围等，而由于传统方法的研究对象很少涉及路网，因此无法很好的回答道路之间的关联性和交通状态影响范围等反映路网整体运行的问题。

近年来国内外有学者从道路关联性的角度对交通数据分析方法进行研究。目前出现的研究主要关注微观和宏观两个层面。其中微观层面是从时间序列的相似性或者先验知识量化多条在拓扑上“邻近”的路段之间的相关程度，并将其作为一种约束条件加入到微观交通控制中去，以达到数据降维^[2]、插补^[3]以及提高准确性^[4]的目的，这些方法本质上还是基于时序分析的思路，并没有真正描述交通数据的空间特征；而宏观层面主要是基于交通流理论来构建交通仿真模型^[5]，用数学模型来描述交通系统的运行规律，这类方法计算量大，建模复杂，而且城市交通系统是一个复杂的系统，这些模型很难顾及所有影响因素，因此缺乏可移植性。

¹注：本文由“空间信息智能感知与服务深圳市重点实验室（深圳大学）开放基金资助项目”资助完成。

随着近年来交通地理信息系统的发展（GIS-T），有学者提出GIS中的空间分析理论可以直接用于解决交通领域中的实际问题^[6-8]，因为从经验出发不难发现，城市交通拥堵现象具有一定的周期性和区域性：上下班时间的拥堵现象比平时要严重很多，而且重要路段交叉口附近是拥堵现象发生的热点区域，这和地理学第一定律^[9]中关于地理现象空间相关特征的描述是相符的。因此，可以把城市交通系统看作是一种复杂的空间过程，并且从空间角度来分析城市交通数据，这样就可以从本质上描述交通数据的整体特征，从而反映城市交通系统的整体运行情况。基于此，本文的研究目的就是空间分析理论最核心的空间依赖性（spatial dependency）和空间异质性（spatial heterogeneity）^[10]对城市交通数据进行分析，研究城市交通状态和路网空间形态之间的关系，从宏观上描述交通运行的空间特征。

2 基于路网结构的交通数据分析模型

在进行空间分析前首先要对研究对象进行抽象，得到具有均匀属性的空间对象。由于空间分析理论在对空间对象的空间分布特征进行分析时主要关注点状和面状的对象，而交通路网在几何形态上是由很多线状的路段所组成的，因此无法直接使用空间分析理论进行分析，所以本章针对城市路网的几何形态提出了一种城市路网的抽象方法，并建立了适合于城市交通数据空间分析的模型。

首先给出交叉口和路段的定义：

定义1：路网中的交叉口（intersection）采用节点-弧段模型中的节点（node）来表示，它是由两条以上具有拓扑邻接路段构成的交点；

定义2：路段（road link）是两个连续交叉口之间的弧段（edge）。

如图1所示，三张图都是节点-弧段模型下形成的路网结构图，但是根据定义1和定义2，第一张图中中间的节点由于只有两条邻接路段，因此只是普通节点而不是交叉口，而下方的弧段只和两个普通节点邻接，因此只是普通弧段而不是路段；图2是标准的交叉口结构，图3是标准的路段结构。

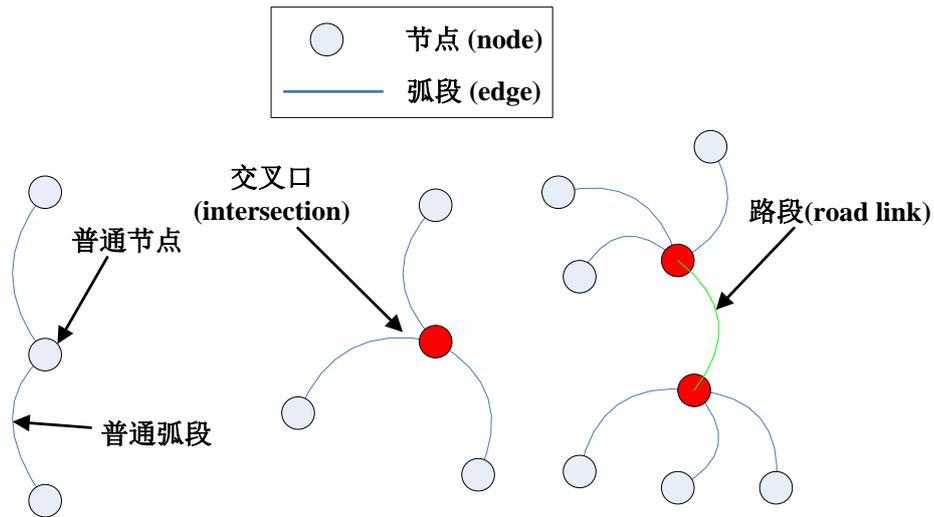


图1 交叉口和路段示意图

然后，假设路段交通状态在短时段 Δt （例如5分钟）内是均匀的，因此可以把路段按照 Δt 的粒度抽象为点状空间实体，而 Δt 内的各种交通数据（例如行程速度、流量等）作为“点状”路段的属性值。

定义3：在短时段 Δt 内，城市交通路网是所有路段的中点所构成的空间离散点，每个 Δt 内的交通数据是一组截面数据（cross-section data），因此长时期内的交通数据可以看作是许多截面数据的集合(见图2)。

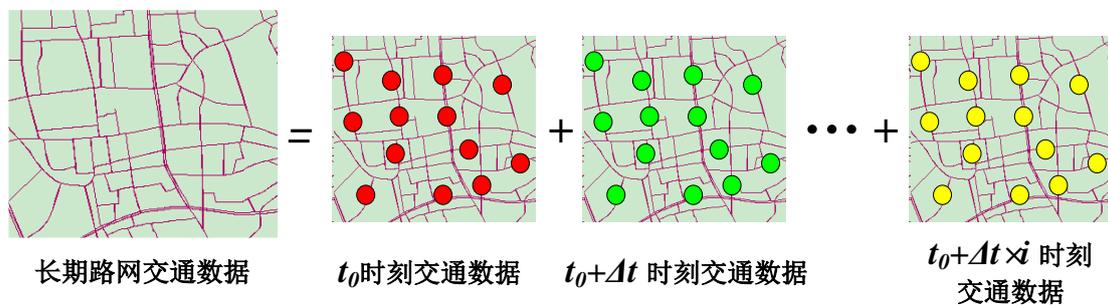


图2 将长时期交通数据按照某个短时段 Δt 划分为截面数据的集合

距离是空间统计学中衡量空间实体之间相关性的参数，其中自由空间中两点之间的距离一般采用欧式距离。但是在交通路网中，路段之间的可达性受到路网结构的约束，因此描述两条路段之间的距离不能采用欧式距离。由于路段的交通状态在短时间内不受路段长度的影响，而且交叉口的空间位置是固定的，所以本文采用从一条路段到另一条路段经过的交叉口数目来表示它们之间的距离。

定义4：两条路段之间的距离是从起始路段到终止路段所经过的最少交叉口数目 N 。如图3所示， T 表示从路段A到路段B经过的交叉口数目最少的路径，其中经过的交叉口数目为37，因此从路段A到路段B的距离为37。



图3 路段间距离的示意图

3 实验数据

浮动车由于覆盖广的特点已经成为目前最常用的一种采集大规模路网交通数据的手段，因此在本文的研究中为了尽可能多的采集路段的交通数据，采用浮动车数据（Floating Car Data, FCD）作为实验的原始数据。

3.1 城市路网行程速度数据

为了保证分析的准确度，需要采用长期且大量的原始数据，而且要满足覆盖率的要求^[11]。因此本文选取了南昌市浮动车系统采集的FCD作为原始数据，采集时间为2008年7月14日到7月20日以及2009年1月1日到1月14号，共计三个星期，其中包括了七个节假日。南昌市浮动车系统采用的频率按照每隔100米和400米两种模式来发送信号，平均发送时间间隔为40秒，覆盖了市区90%以上的路段。

行程速度作为一种典型的交通数据，可以反映道路的交通运行状态^[12]，而且在以前研究中，笔者在以前的研究中提出了基于车辆轨迹的路段行程速度提取方法^[13]，这种方法可以从FCD中提取出路段在固定时段内的行程速度值，因此本文的实验采用交通领域常用的反映路段实时运行状态的5分钟作为时间粒度，以5分钟的路段平均行程速度作为交通数据样本。另外，由于凌晨0点到6点的浮动车覆盖率很低，而且这个时段车流量极少，对城市整体交通而言没有太大的研究价值，因此剔除掉这个时段的数据。这样一条路段一天中理论上可以采

集到216个样本 $S=\{S_1, S_2, \dots, S_{216}\}$ 。

3.2 城市路网数据

本文选择南昌市的市区路网作为实验的路网数据，其中包括4359条路段，按照道路功能可以分为快速路，主干路，次干路和支路（如图4所示）。其中，部分路段由于覆盖率的问题导致行程速度值缺失严重，影响了分析的精度，因此在实验过程中剔除了行程速度缺失率高于25%的路段。

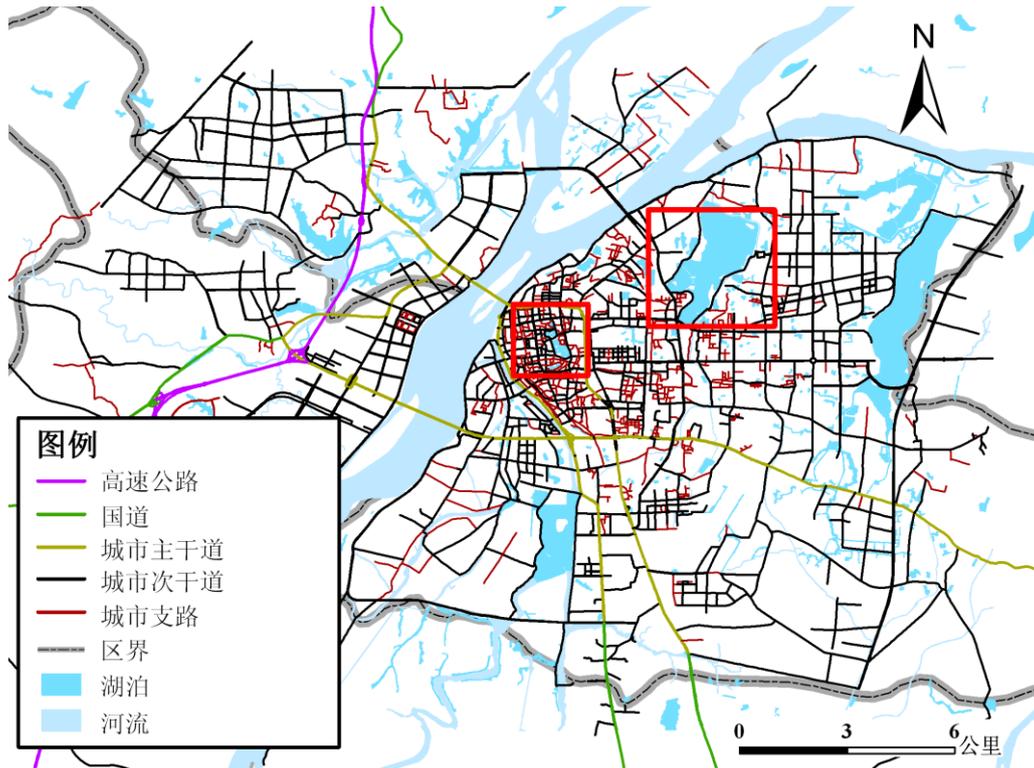


图4 南昌市路网数据

3.3 城市交通的典型时段划分

以天为单位的交通数据在时序上存在多模式，相同模式的交通数据具有相似性，而不同模式的交通数据具有显著的差异性，为了分析城市交通在不同模式的空间特征，有必要对其进行时段划分。

本文首先计算2008年7月14日（周一）到7月20日（周日）所有路段5分钟行程速度的平均值，以此作为路网5分钟的整体速度，然后对早上7点到晚上23点间的路网整体速度两两计算Pearson相关系数（如图5所示）。

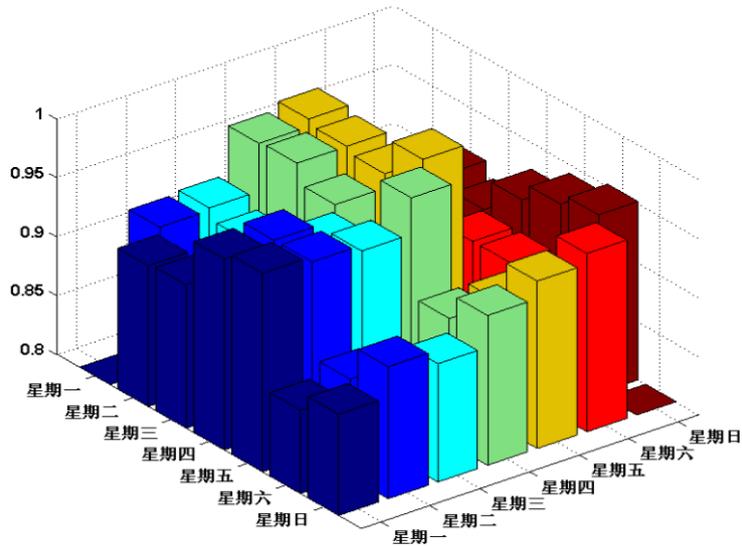


图5 路网行程速度的Pearson相关系数比较

实验结果显示工作日和非工作日（节假日）的路网整体交通状态具有显著的差异，但是工作日之间和非工作日之间又存在着相似性。

另外，本文选择每天早上8:00~9:00作为早高峰时段，下午15:00~16:00作为平峰时段，晚上18:00~19:00作为晚高峰时段，对路段每天在不同时段的速度分布情况进行了分析（如图6所示），实验结果显示工作日具有明显的“三峰”特征，而非工作日的不明显。

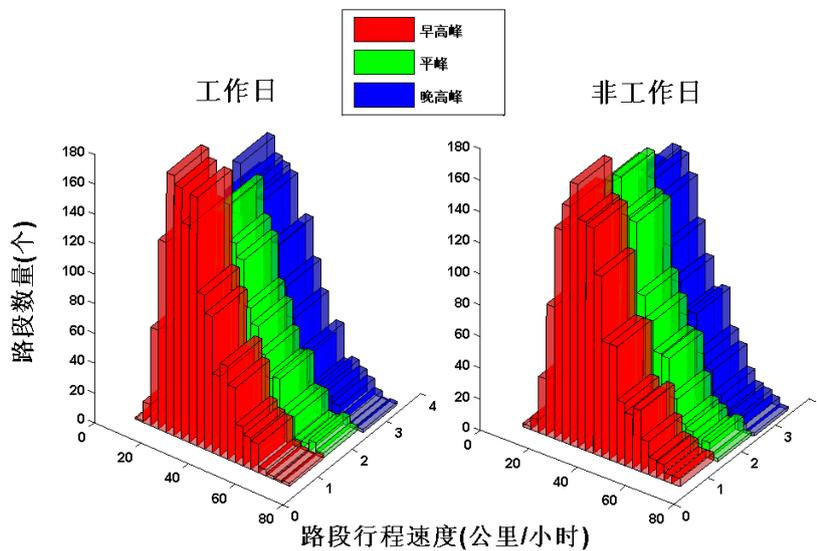


图6 工作日和非工作日的路段行程速度分布

根据上述分析结果，在本文的实验中城市路网行程速度的分析划分为工作日早高峰，平峰，晚高峰以及非工作日四个典型时段。

4 城市路网行程速度的空间分析结果及特征

4.1 空间依赖性分析

在空间分析理论中，空间依赖性用来反映空间对象某种属性的空间模式^[10]。空间自相关分析（spatial autocorrelation analysis）是度量空间依赖性的手段，常用的方式是通过全局（Global）或者局部（Local）的空间统计量来量化空间自相关，从而反映空间对象的空间依赖程度。其中，全局统计量用来判断空间对象整体的空间自相关程度，而局部统计量反映空间对象在空间中的集聚程度。

通过分析整体路网的空间依赖性，可以发现城市交通运行中路段之间的空间关联程度，这有利于描述路段交通状态对整体路网交通状态的空间影响力。本文采用经典的全局Moran I统计量作为指标。全局Moran I的定义如下^[14]：

$$I(d) = \frac{\sum_i \sum_{j \neq i} w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{S^2 \sum_i \sum_{j \neq i} w_{ij}}$$

其中， S^2 ， x_i 和 x_j 分别表示属性X的方差以及位置i和位置j上的观测值， S^2 是 $\{x_i\}$ 的均值。 w_{ij} 是空间权重矩阵（Spatial Weight Matrix, SWM），它用数值来量化空间实体之间的近邻程度，根据不同的应用有许多不同的空间权重矩阵定义方法，其中比较常用的是二进制SWM和基于距离的SWM。Moran I值一般在-1到1之间，正值表示属性具有正空间自相关性，负值表示具有负空间自相关性，其绝对值越大表示空间自相关程度越强。

另外，有的学者提出了Z score统计量对Moran I进行显著性检验，在零假设为Moran I服从正态分布的情况下，Z score的形式为^[9]：

$$Z(d) = \frac{I(d) - E(I)}{\sqrt{VAR(I)}}$$

其中， $E(I)$ 是Moran I的期望， $VAR(I)$ 是Moran I的方差。当 $|Z| > 1.96$ 时拒绝零假设，表示空间对象之间存在着空间自相关性，否则不存在。

在本文的实验中，为了计算城市交通数据的全局Moran I，笔者以路段行程速度作为属性X。首先采用Anselin提出的方法^[15]来构建高阶SWM：

$$W = \sum_{d=1}^D W_d$$

其中， d 表示阶数。然后按照定义4分别构建7个不同阶数的SWM（1-7阶），其中阶数也可以称为连接层次，表示两路段“邻近”的标准，越低阶的SWM表示邻近标准越苛刻，其邻近路段也越少；其在空间上表现为从某路段出发，按照拓扑层次关系搜索到的所有路段的集合。图7和表1分别显示了在典型时段交通数据全局Moran I和其Z score的计算结果。

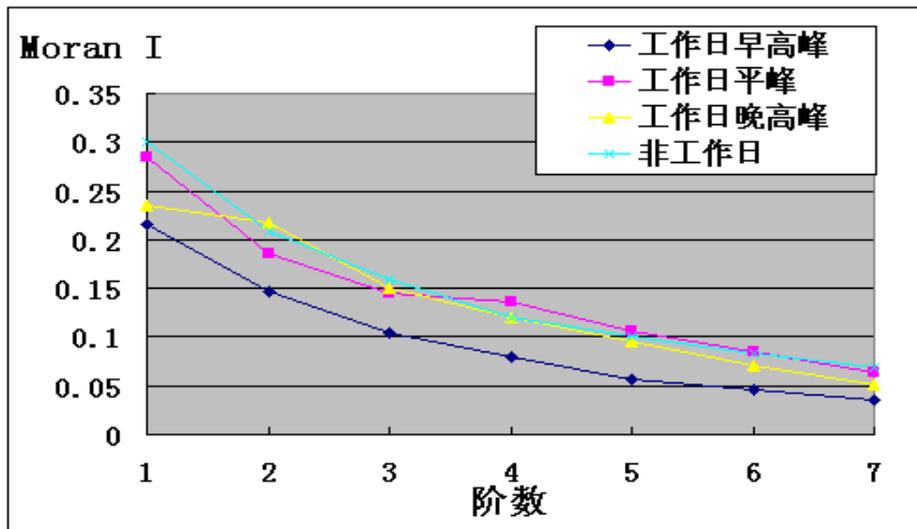


图7 城市路网行程速度在典型时段的全局Moran I

从表1中可以发现，在四个时段内的Z score值都远远大于1.96，这表明城市交通数据存在着显著的空间自相关性；但是从图1的曲线中发现，Moran I值随着阶数的增长在不断衰减，并且从5阶开始趋向于0 (<0.1)，这说明路段之间的相关性随着距离的增长在不断减弱，并且增长到一定距离之后将会逐渐消失，例如本文的实验结果为5阶，这说明单一路段的影响范围大约是5个交叉口之内。上述的分析结果和人们的先验知识也是相符合的：邻近路段的交通状态存在着相似性，但是当超过一定范围之后路段间交通状态的差异也会随之增大。通过上述的分析可以量化地分析城市路网交通状态的相关程度，这对于交通事件的影响范围和变化规律的研究具有重要的意义。

表1 城市路网行程速度全局Moran I的Z score检验结果

阶数	早高峰(工)	平峰(工)	晚高峰(工)	非工作日
1	5.4087	6.1831	5.2770	7.3655
2	6.3077	7.0030	7.6301	8.7308
3	6.6787	7.7005	7.9844	9.5592
4	6.9198	8.8484	8.4722	9.9139
5	6.8115	9.0619	8.7090	10.400
6	6.9370	7.7217	6.6267	8.7085
7	6.8628	6.7254	5.6499	8.3672

4.2 空间异质性分析

空间异质性在空间分析理论中用来反映空间对象分布的不稳定程度。对空间对象来说，

空间异质性和空间依赖性同时存在的，这表现为在整体上存在着关联性，但是又存在着不均匀的局部集聚现象^[6]。

通过对城市交通空间异质性的分析可以从宏观上发现城市交通状态在空间中的影响程度和范围，结合交通状态的时态特征还可以用于描述城市交通在一定时期内的演变规律。在本文的实验中引入了核密度估计（Kernel Density Estimation, KDE）方法来分析城市交通的异质性，这是一种在空间分析领域常用的趋势面分析（Trend surface analysis）方法，其优点是能够用连续函数来描述空间对象的分布特征，具有较好的可视化效果，比较适合宏观分析。KDE方法的核心公式如下^[16]：

$$f(s) = \sum_{i=1}^n \frac{1}{\pi r^2} k\left(\frac{d_{is}}{r}\right)$$

其中， $f(s)$ 表示位置 s 处的密度值， r 是搜索半径（带宽），只有在 r 范围内的空间对象才能被用于计算 $f(s)$ ， k 是到 s 距离为 d_{is} 的点 i 的权重值，通常用自变量为 d_{is} 和 r 比值的函数来表示（核函数），目前常用的核函数包括Gaussian, Quartic和 Minimum variance function，但是以前的研究中发现带宽的选择比核函数的选择要重要^[16]。在本文的实验中为了能够尽可能细化分析结果，所以选择的带宽为500米（在市中心500米范围内至少会出现5条路段，可以充分保证分析的精度）。

图8显示了基于KDE方法构建的工作日早高峰行程速度的密度趋势面，不同颜色表示不同的密度，其中颜色越深表示越低的行程速度。从分析结果可以发现：（1）行程速度在空间中的密度分布具有明显的不均匀性，这说明行程速度的分布存在局部集聚的现象；（2）从密度的区域可以判断不同交通状态的空间分布情况，这有利于从宏观上直观地发现各种交通状态的影响范围。另外，为了更进一步的分析，笔者针对南昌市商业区较集中的区域（A）和居民区较集中的区域（B）进行了行程速度分布情况的比较，从图9的结果中可以发现，同一时段内交通状态的空间分布受到区域功能的影响：商业区（A）在早高峰时段路段的行程速度在20~30km/h段分布比较集中，而居民区在30~40km/h段分布比较集中，这说明商业区在早高峰时段的交通状况比居民区要差，这主要受城市居民出行的影响，但同时也和商业区密集的路网结构有关。

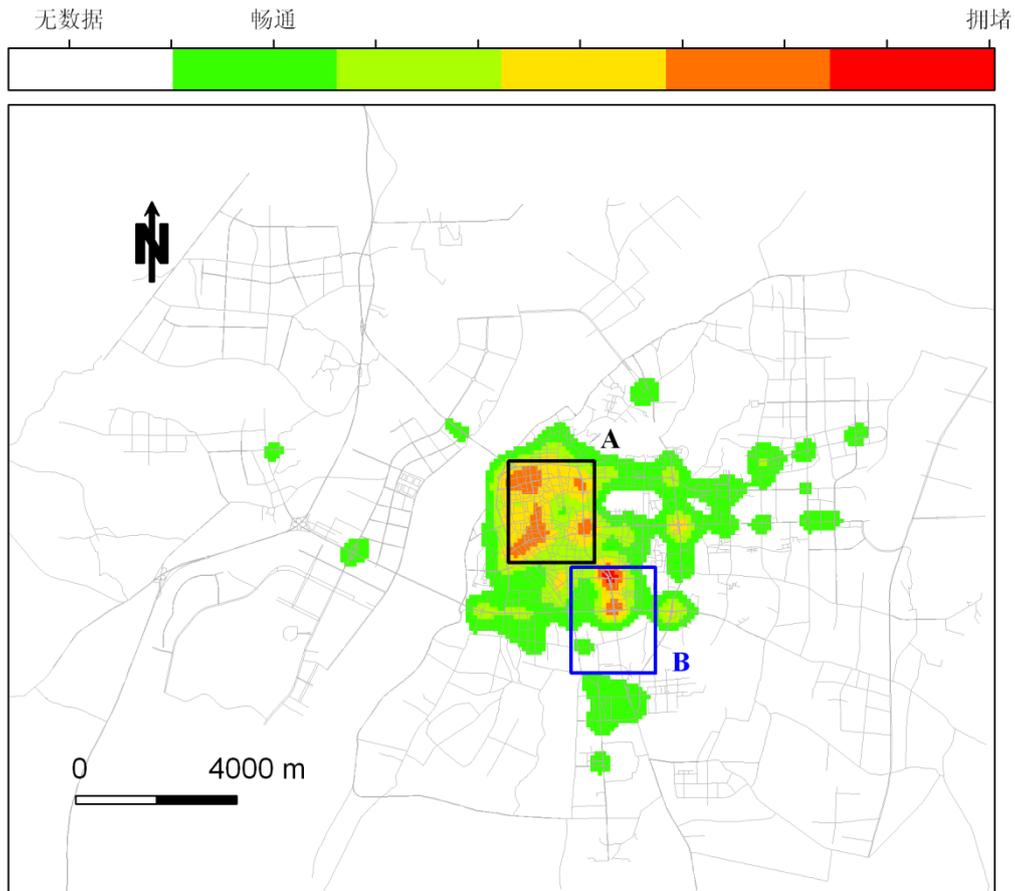


图8 南昌市工作日早高峰时段的行程速度密度图（区域A为商业区，区域B为居民区）

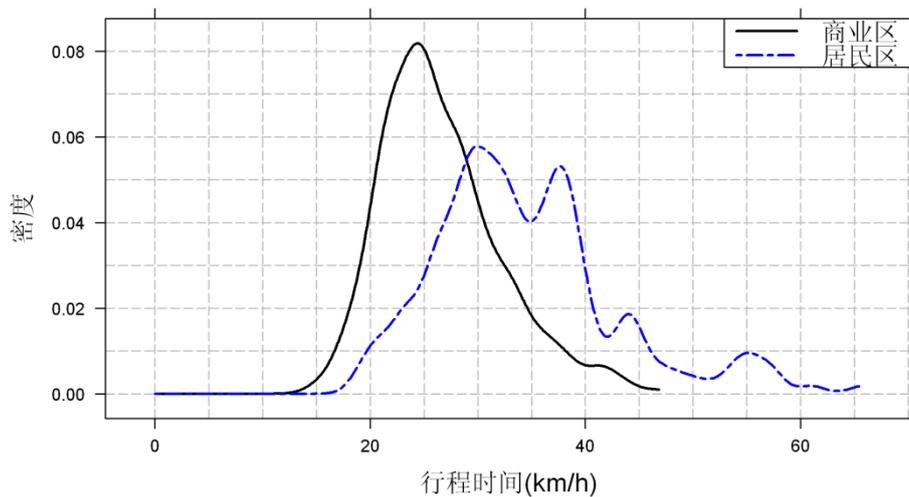


图9 商业区和居民区在工作日早高峰时段的行程速度分布

4.3 空间分布特征分析

在上一节中，笔者针对城市交通的空间异质性进行了分析，并且通过趋势面分析显示了路网行程速度在自由空间中大致分布情况，为了能够更好地发现交通运行中的空间模式，

在这一节中将基于路段位置和路网空间形态，详细地分析路网行程速度的空间分布特征，并给出一些量化的结果。

首先，笔者在实验中选取城市路网的几何中心作为空间基准：根据定义3获得所有路段位置的经纬度，并用其平均值作为城市路网的几何中心位置 (X_{center}, Y_{center}) ：

$$(X_{center}, Y_{center}) = \left(\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n} \right)$$

然后，将每个时段各路段距离中心位置的距离和行程速度视为一组样本 $S_p = (D, V)$ ，其中D是按照定义4计算出的路段距离中心位置的距离（单位是交叉口的个数），V表示路段的行程速度。

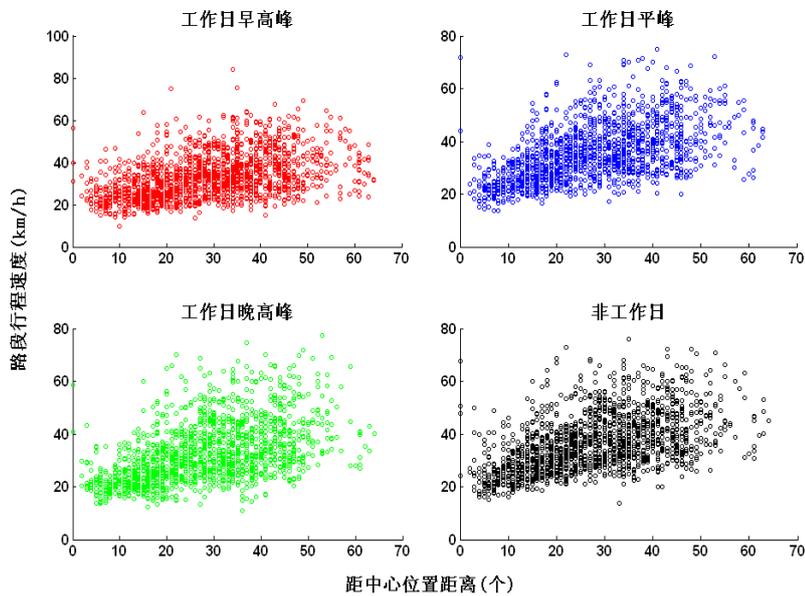


图10 典型时段路段离中距离和行程速度的关系

如图10所示的散点图结果可以直观地发现：（1）所有时段90%以上的样本都集中在一个的扇面区域内，这说明路段行程速度和位置之间不存在线性关系，呈现空间上的不稳定性；（2）扇面区域的张角在不同的时段具有不同的特征，这说明交通数据的空间分布随时间变化，呈现时态上的不稳定性，其中工作日早高峰时段的夹角最小，这反映工作日早高峰时段路网行程速度的空间分布较其他时段更为均匀；（3）城市交通数据受到路段空间位置的影响，本文仅以一个城市为例，在路网形态相异的其他城市可能会得到不同的分析结果。

上述的实验粗略地分析了交通数据和路段几何位置的关系，为了能够得到更加精细的结果，笔者将每个时段所有路段行程速度的样本按照从小到大的顺序排列，然后等分为三份，其中值最小的一份称为低速样本，值最大的一份称为高速样本，中间的称为中速样本。然后得出了每份样本的分布情况，结果如图11所示。

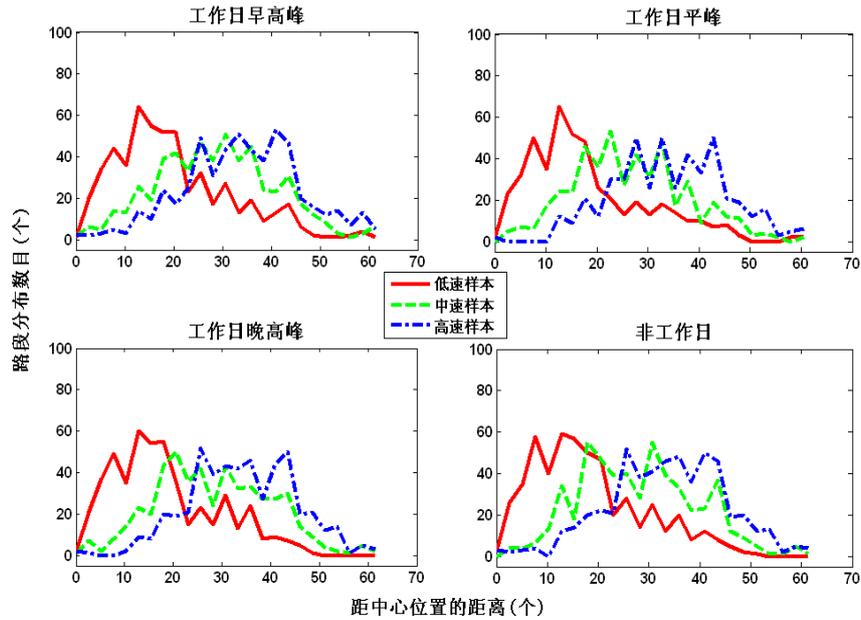


图11 典型时段不同速度等级路段的空间分布

从实验结果中可见不同时段的空间分布特征虽然局部上存在着差异性,但是从整体上看是相似的: (1) 低速样本主要分布在离中心位置较近的区域; (2) 中速样本比高速样本在距中心位置较近的区域分布的更多,但是在距中心位置较远的区域两者的分布是相近的; (3) 所有样本的分布曲线都具有多个波峰和波谷, 利用这些波峰和波谷可以发现城市交通中的“热点”和“冷点”区域。例如, 工作日早高峰时段低速样本分布频数最高值大约是15, 这说明距离城市中心15个交叉口附近的路段交通状态最差; 同样的, 从高速样本曲线可以发现工作日早高峰时段距离市中心25, 35和40个交叉口附近的路段交通运行比较理想。

5 结论

本文将城市交通系统看作一种复杂的空间过程, 将空间分析的方法引入交通数据分析, 并且针对城市路网几何结构和交通状态的关系提出了适合城市交通数据分析的空间模型, 这部分探索性的工作可以为交通领域的研究提供新的研究思路, 同时也为空间分析理论提供了新的应用领域。通过对从大量的浮动车数据中提取的真实交通数据的实例分析, 本文论证了城市路网交通数据中存在着空间依赖性和空间异质性, 这反映了城市交通状态存在着局部的相似性, 并且同时又在时空上呈现出不稳定性。

根据文中的分析结论, 一方面可以从宏观上发现城市交通实时运行的空间特征以及长时期的演变规律, 有效的发现城市交通系统的整体运行特点, 为城市管理者提供决策支持, 给解决城市交通问题带来新的途径; 另一方面还可以分析路段交通状态的影响力和范围, 为交通疏导、交通控制等微观层面的应用提供有效的信息支持。通过本文的研究可以证实空间过

程模型在交通中应用的可行性,因此,将这些具有空间特征的模型具体地用于解决各种实际的交通问题将是今后的工作重点。

【参考文献】

- [1] W. Willinger, M. Taqqu, and A. Erramilli. Stochastic Networks: Theory and Applications [M]. Oxford Science Publications, 1996:339-340.
- [2] 官晓燕,陈伟雄. 基于关联规则挖掘的路段流量关系的分析[J].公路交通科技, 2003,20(4):63-66.
- [3] Kumagai M., Fushiki T., Kimita K. and Yokota T. Spatial interpolation of real-time floating car data based on multiple link correlation I feature space [C]. Proc. of 13th World Congress of ITS, London, CD-ROM, 2006.
- [4] 韩卫国,王劲峰,高一鸽,胡建军. 区域交通流的时空预测与分析.公路交通科技, 2007,24(6):92-96.
- [5] Salvini PA, Miller EJ. ILUTE: an operational prototype of a comprehensive microsimulation model of urban systems [J]. Networks and Spatial Economics. 2005,5:217-234
- [6] H. J. Miller. Potential Contributions of Spatial Analysis to Geographic Information Systems for Transportation (GIS-T)[J]. Geographical Analysis, 1999. 31(4): 373-399.
- [7] H.J. Miller, Shih-Lung Shaw. Geographic Information Systems for Transportation: Principles and Applications[M]. Oxford: Oxford University Press, 2001.
- [8] 方志祥,李清泉,萧世伦.利用时间地理进行位置相关的时空可达性表达[J]. 武汉大学学报·信息科学版. 2010.35(9): 1091-1095.
- [9] W. R. Tobler. A Computer Movie Simulating Urban Growth in the Detroit Region[J]. Economic Geography, 1970. 46: 234-240.
- [10] 王劲峰等,空间分析[M]. 科学出版社:北京,2006.
- [11] 李清泉,尹建忠,贺奋琴.面向道路网的GPS浮动车覆盖率模型研究[J]. 武汉大学学报·信息科学版. 2009.34(6): 715-718.
- [12] 姜桂艳. 道路交通状态判别技术与应用[M]. 人民交通出版社,2004:49-52.
- [13] Yue Y., Zou H. X. and Li Q. Q. Urban road travel speed estimation based on low sampling floating car data [C]. Proceedings of the 9th International Conference of Chinese Transportation Professionals (ASCE), Harbin, China, August 5-9, 2009: 1719-1725.
- [14] Cliff A. D. and Ord J. K. Spatial process: models and applications [M]. London, Pion, 1981.
- [15] Anselin, Luc and Oleg Smirnov. Efficient algorithms for constructing proper higher order spatial lag operators [J]. Journal of Regional Science, 1996, 36:67-89.
- [16] Xie Z. and Yan J.. Kernel Density Estimation of traffic accidents in a network space [J]. Computers, Environment and Urban Systems, 2008, 32(5):396-406.

【作者简介】

邹海翔,男,博士,深圳市规划国土发展研究中心,规划师,工程师。电子信箱:
zou.mono@gmail.com