

基于手机基站数据的城市交通流量模拟

吴健生¹, 黄力¹, 刘瑜², 彭建³, 李卫锋⁴, 高松², 康朝贵²

(1. 北京大学深圳研究生院城市人居环境科学与技术重点实验室, 深圳 518055;
2. 北京大学遥感与地理信息系统研究所, 北京 100871; 3. 北京大学城市与环境学院, 北京 100871;
4. 香港大学城市规划与设计系, 香港)

摘要: 基于移动定位数据的城市内社会经济活动特征分析是人类移动性的重要研究内容, 而交通流量更是这些特征的基本反映。为还原城市道路网络的使用情况并分析其分布特征, 本文从产生交通流量的个体出发, 对包含基站位置的手机话单数据进行系统抽样, 利用蒙特卡洛方法产生个体的出行起止点, 并结合当地道路交通网络求得最短路径, 最后估算出一天内道路交通网络上的流量分布。通过分析发现: 城市内大部分道路的流量小, 使用率低, 大部分交通流量集中在小部分主干道路; 进一步统计分析可知, 当地道路交通流量符合20/80规律, 即大约20%的道路承担着80%的交通流量; 而对不同类型的道路, 流量分布也反映出其在城市道路网络中的地位和作用。此研究对于历史交通流量分布的重现、城市道路交通模式的研究以及基于此的道路网络规划情景模拟都有着重要意义。

关键词: 人类移动性; 手机基站数据; 城市交通流量模拟

1 引言

信息与通讯技术的广泛使用产生了海量时空数据, 基于此的人类移动性 (Human Mobility) 研究也逐渐成为热点^[1-7]。由于具有多维度、高精度、支持行为轨迹分析和适用于大规模数据分析等特点^[4], 手机移动定位技术在研究人类移动性方面应用广泛, 其中包括: 移动轨迹的地理可视化^[8]、旅游等出行方式的研究^[9-10]、通讯工具的使用对出行的影响研究^[11-12]、个体的移动性规律与模式挖掘^[13-19]、个体移动轨迹的预测^[20-21]、城市内部宏观活动特征与结构^[22-23]等。在人类移动性中, 随时间而变化的空间位置是其最基本反映, 以往的研究大都着眼于数据记录时刻个体所在空间位置及其反映的活动模式, 而忽略了相邻两个时刻之间的移动过程及其对区域空间结构的影响。而在实际的移动过程中, 数据记录的个体会在当地的道路交通网络上产生交通流量。

在基于自动感应设备采集交通信息数据的研究中, 常用的数据来源包括路面传感器、车载GPS和手机基站网络, 其中路面传感器只能采集网络中少数固定点的交通信息, 车载GPS由于信号多路径 (Signal Multipath) 和城市峡谷阻塞 (Urban Canyon Obstruction) 问题在城区效用并不高。手机基站网络虽然精度较低, 但是移动设备的广泛使用和基站的大量铺设提供了海量的个体时空数据, 通过聚合与处理, 结合个体属性数据, 与传统交通信息数据互补, 为模拟交通流量、理解出行行为和规划交通网络提供了有力的手段^[24]。因此, 本研究将利用手机基站数据模拟城区交通流量, 在此基础上对区域内不同等级道路的流量分布进行分析, 探讨其在城市道路网络中的功能和地位, 并通过对比前人基于车载GPS的研究结果, 验证本研究方法的可行性和准确性。

为了还原一定时段内城市道路交通网络的使用情况, 探讨不同等级道路的功能地位和

收稿日期: 2012-05-11; 修订日期: 2012-10-09

基金项目: 国家自然科学基金项目(41271101) [Foundation: National Natural Science Foundation of China, No.41271101]

作者简介: 吴健生(1965-), 男, 副教授, 主要研究方向为城市景观生态和GIS。E-mail: wujs@pkusz.edu.cn

通讯作者: 黄力, E-mail: huangli.ghost@gmail.com

1657-1665 页

道路网络总体流量分布特征,本研究通过编写基于 ArcEngine10.0 的 C# 程序,利用其中的 Network Analyst 模块库,从对移动过程进行模拟还原这一动态角度出发,将手机移动定位数据与道路网络结合起来,通过对大规模样本的模拟,将数据反映的个体空间位置的移动过程具体化,在网络中模拟根据最短路径原则所生成的交通流量。因此,本研究基于交通行为作为一种大量个体时空行为总和的本质特征,利用手机基站数据可方便快速获取大规模实时移动数据的优点,提供了一种新的模拟历史时期一定时段内交通流量空间分布的方法,并为交通网络流量的分布模式分析和未来交通规划奠定了数据基础。

2 研究区与数据来源

本文研究区为某省会城市的中心城区,2010年总面积近 200 km²,建成区面积约 50 km²,总人口 100 多万, GDP500 亿元以上。区域内道路网络由国道、省道、高速公路和城市内主次干道组成,总长约 1500 km,其中包括 3 条国道和 6 条高速公路。

选取该研究区的原因包括以下 3 点:① 人口数量大,可提供海量的手机用户数据进行模拟;② 经济发达,基础设施如手机基站网络的建设较好,采集的数据质量精确可靠;③ 道路网络健全,个体的位置变化大部分倾向于沿道路网络进行。这些条件对于个体移动模拟的准确性至关重要。

本文使用的数据为城市道路网络数据和手机基站数据。其中,城市道路网络数据包括国道、省道、高速公路、城市道路、市区道路、市区杂路 6 个图层,来自 2007 年国家基础地理信息中心基本数据集中与道路相关的 6 个同名子目录,图层间道路无重复无遗漏。手机基站数据包括区域内的手机基站位置空间分布图以及 2007 年 7 月某天的匿名化处理后的各基站话单数据。其空间分布见图 1。

从图 1 可看出:手机基站的分布密度随区域内人类社会经济活动的增强而升高,表现为从城区外向城区内逐渐密集;高速公路与国道建立了城区与外部的连接,而省道除了连接城区内外,也构成城区内的道路网络骨架;城市道路是东部城区的主干道路,市区道路在上述道路基础上对网络进行了加密,市区杂路则在市区道路的基础上进一步加密并提供了小区尺度上的可达性;两大类型的道路在功能和接驳上分配较为合理,构成了当地健全的道路网络。

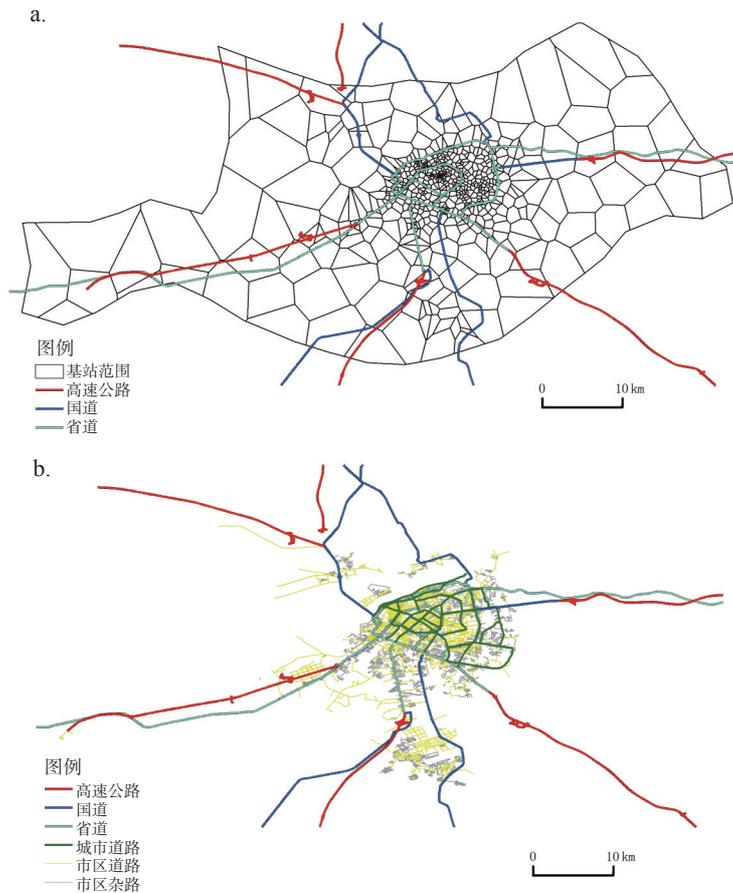


图 1 道路网络与基站范围

Fig. 1 Distribution of traffic network and cellular base stations

3 研究方法

为了利用手机基站数据对一天内的交通流量进行估算,就需要对数据记录的每个个体的移动进行模拟。当个体使用手机进行通话或发送短信时,其通讯业务会由距其最近的基站负责处理,由此产生一条话单记录。而如果在总的话单数据中该个体的手机号出现了两次或两次以上,则可以捕捉到个体移动的轨迹,数据中按时间先后形成的基站序号即为其移动轨迹中的重要节点。通过对所有个体的序列进行合并,得到每对基站间个体移动的总量。

但是,这只是说明移动的起止点在对应基站服务范围内,为了得到具体坐标,使用一个简单的蒙特卡洛 (Monte Carlo, MC) 方法^[25]:基站的服务范围由其 Voronoi 多边形确定,在其外包矩形内生成随机点,将点产生在多边形内的概率设为1,其余为0,由此产生起止点坐标。

利用 ArcEngine 10.0 的 Network Analyst 模块在构建好的城市道路交通网络数据集内求算起止点间的最短路径,得到移动产生的交通流量。通过对所有基站间的移动总量进行模拟,进而得出一天之内的交通总流量。

此方法虽然简单,但是在处理大规模数据时仍需改进。从对当天通讯数据的统计可以发现,在一天之内有个体移动的基站对数量为 13.04 万。以人次为单位,基站对间的移动量 (Flow) 最小为 1,最大为 9890,均值为 34.79,标准差为 185.20,叠加后的总量为 453.54 万。若处理 1 次移动需要 1 秒,则处理一天的数据需要 52 天,这对于得到一段时期内交通网络的流量分布模式和为未来交通规划提供依据的研究目的来说耗费时间过长,故需要从单个个体的模拟和整个模拟过程两个方面进行改进。

首先,对单个个体的模拟过程进行改进。对于求个体一次移动过程的最短路径的情形(图2),先对路段两两端点之间(图2中3~5、4~5、3~6、4~6共4种情景)求最短路径及其长度,并存储在中间文件中;然后将起止点到对应路段4个端点之间的距离考虑进来,得到从起点到终点的4条可能路径的长度,选取最小者,即是最短路径。而在处理移动数据的过程中,通过反复读取和扩充存储着道路交通网络节点间最短路径集的中间文件,可避免对最短路径解中有相同路段起止点的通行情景重复计算,从而提高单个个体的模拟效率(图3)。

其次,对整个模拟过程的改进。在选取的话单日记录中共有 453.54 万次基站间的移动,为了估算一天内的交通流量及其分布特征,估算过程中并不需要对整个记录进行模拟,而可以通过抽样的方式,选取记录的一部分进行模拟。

为了在缩减模拟规模的前提下保证模拟结果的有效性和准确性,在抽样方法上选择系统抽样,抽样间距为 100。对于 MC 方法生成的 400 多万次移动的起始终止点,其本质是一个

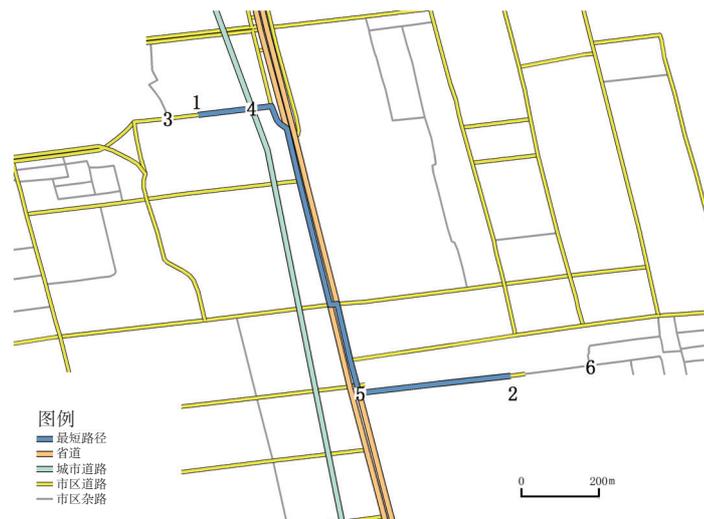


图2 最短路径计算示例

Fig. 2 Calculation of the shortest path

注:图中1、2分别为起始、终止点,3、4和5、6分别为离1、2的最近路段的两个端点

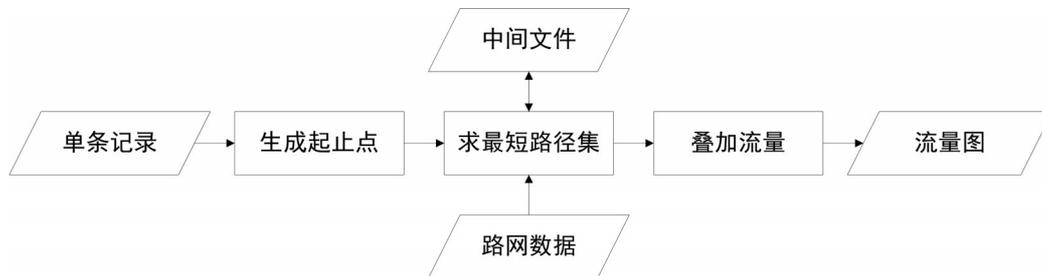


图3 单条移动数据处理流程

Fig. 3 Processing of a single record

离散的地理空间对象集合，它的分布疏密程度反映的是研究区内不同基站区域的人类活动程度的高低。而地理空间事物普遍存在空间自相关性^[26-28]，使用简单随机方法对其进行抽样会减低抽样的估计效率和抽样的总体代表性，在相同样本量下分层抽样和系统抽样的估计精度比简单随机抽样要高，而且当无先验知识指导分层时，系统抽样会比分层抽样得到更高的估计精度^[27]。

4 结果分析

4.1 城市交通流量模拟

对2007年7月某日的当地话单数据进行系统抽样，抽取共计4.54万人次基站范围之间的个体移动，分别计算最短路径并进行叠加，得到当日的交通流量估算结果。用自然间断点分级法(Jenks)将道路流量为分大中小3个等级，流量小的(0-748)道路设为青线，流量大的(2560-7107)道路设为蓝线，流量适中的(748-2559)道路设为紫线，得到城市交通流量模拟图(图4)。

从图4可得出，在一天当中，区域内大部分道路流量较小(青线)，共计723条，其道路使用率低，道路交通网络的运行主要是靠小部分流量较大的道路支撑着(蓝线与紫线)，共计139条。分别统计大中小3个流量等级中6种类型道路的数量(表1)。从表1中可以发现，国道和省道在3个流量等级中分布较均一，大部分市区杂路属于小流量等级，而城市道路和市区道路在3个流量等级中呈阶梯式分布。这也反映出不同道路类型在当地道路网络中的地位：国道和省道作为干线道路，连接着城市内外的行政中心和经济中心，是当地道路网络的骨架，故相对而言其大中流量道路数量比例大；市区杂路为了方便居民的生活、工作和娱乐，需要提供对城市各地区广泛的可达性，是当地道路网络的枝叶，故其基本都是小流量道路；城市道路和市区道路作为当地道路网络的核心，需兼顾骨架和枝叶的功能，故其在流量等级中从大到小呈阶梯式递增分布。

4.2 城市交通流量20/80现象

作为道路网络的核心，市区道路也体现出图4模拟结果中整个城市道路网络体系的流量分布特征：即大部分道路流量小，小部分道路流量大，整体上既有对区域内广泛的交通可达性的贡献，又有着连接区域内重要的热点区域的作用。通过对模拟结果进行进一步处理，依据交通流量对所有道路进行降序排列，并对流量进行累加，得到大于等于某个排名的流量之和在总流量中的比例分布(图5)。

表1 不同道路类型的流量分布(单位:条)

道路类型	大流量	中流量	小流量	小计
国道	2	1	2	5
省道	8	15	11	34
高速公路	0	0	6	6
城市道路	2	4	30	36
市区道路	17	84	528	629
市区杂路	0	6	146	152
总计	29	110	723	862

对处理后的模拟结果按照道路排名的 1%、20% 和流量叠加和的 20%、80% 分别截取对应数据点 (表 2), 可以发现, 城市中道路网络流量分布在数值上表现为 20/80 现象, 即约 20% 的道路承担着 80% 的交通流量, 其中约 1% 的道路承担着 20% 的交通流量。

这一现象可以通过对比前人研究得到验证。Jiang 通过对瑞典耶夫勒市出租车 GPS 数据进行处理, 得到了当地道路网络流量的分布特征: 在以街道名称划分的道路网络中, 1% 的街道承担着约 25% 的流量, 20% 的街道承担着约 80% 的流量; 在以拓扑属性划分的道路网络中, 1% 的道路承担着约 40% 的流量, 20% 的道路承担着约 85% 的流量^[29]。但与表 2 相比, Jiang 的结果整体排名趋势略高, 在以拓扑属性划分的道路网络中偏差较大, 这是由于:

第一, 两者的道路划分标准不一。在 Jiang 以拓扑属性划分的道路网络中, 用来统计流量的道路是以道路连通性 (Street Connectivity) 作为依据由不同路段组合而成的, 有良好连通性和处于中心地位的路段易组成同一道路单位来统计出租车流量, 从而形成其中承担 40% 流量的前 1% 道路。本文以道路名称来区分不同道路单位, 依据的是其社会属性, 同样是连通性较好和处于中心地位的路段可能分属于不同名称的道路单位, 故虽然前 1% 的道路承担的 12.55% 流量比 40% 要小。但当统计范围扩展到前 20% 的道路时, 重要路段都包含其中, 两者的差距缩小, 基本反映道路流量的 20/80 规律。

第二, 两方法的数据来源不一。Jiang 的研究中采用的是出租车 GPS 数据, 反映的是此种出行方式下的人类移动性, 由于交通通达性和出行方式选择上的差异, 具体到

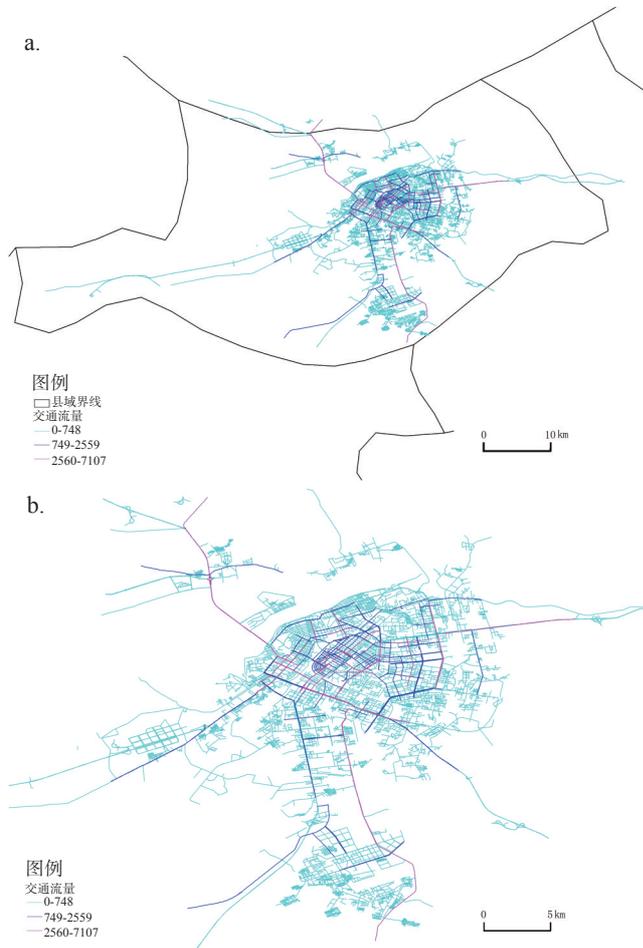


图 4 城市交通全日流量
Fig. 4 Traffic flow simulation result

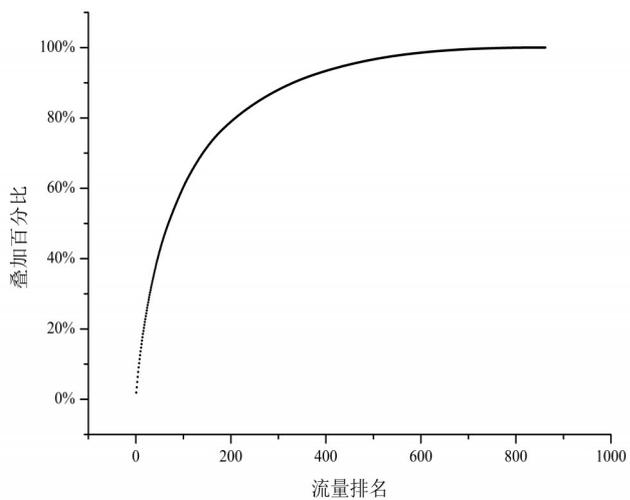


图 5 流量排名叠加
Fig. 5 Superimposed flow rank scale

道路网络上其计算出的流量分布本身具有向城市中心集聚的特征，故在同样是以街道名称为划分标准的模拟结果上，Jiang的排名趋势比本文结果有所上升，具体表现为虽然都是前1%的道路，由于其大都分布在城市的中心地区或交通热点地区，在Jiang的结果中承担了25%的流量，而本文中只承担了10%~20%的流量。但当统计范围扩展到20%的道路时，与以拓扑属性为划分标准的模拟结果一样，两者的差距缩小，基本反映道路流量的20/80规律。

因此，虽然与前人研究相比有所出入，但本文模拟结果与之都反映出道路流量分布的20/80规律，统计结果上在计入排名的道路范围扩大时符合较好，在空间分布上也体现出城市中心流量大边缘小的特征，可认为以此方法估算出的交通流量与历史时期真实的交通流量分布基本相符。

4.3 道路潜在易堵性识别

在流量分布图的基础上，可依据流量叠和百分比对当地的道路进行分类，作为其潜在易堵性，从而识别哪些道路可能由于承载的交通流量太大而发生堵车需要进行整治，哪些道路使用率较低需要撤并来改善通达性。本研究在处理数据的第一步将所有个体的基站点序列合并为在两两基站范围间移动的个体数量总和，这为识别出入量大的热点基站小区提供了可能。而更对应到模拟结果中的交通流量分布上是排名靠前，处于重点位置且可能发生堵车的热点道路。

按流量叠和百分比的前5%、5%~20%、20%~80%和80%~100%对道路网络进行划分，分为高、中、低以及无潜在易堵性(图6)。从图中可看出：当地潜在易堵性最大的道路集中在从南至北进入市中心的道路，其次的中潜在易堵性是从城市次中心进入城区的道路，潜在易堵性较低的是市中心及附近呈辐射延伸的道路，最后部分的城市道路和大部分市区杂路在只考虑个体通勤意愿的情况下几乎没有堵车的可能。在定位了高潜在易堵性和低使用率的道路之后，针对交通道路网络中不同的路段提出对应的整改意见，作为未来城市交通规划方案的基础：如对堵车风险高的热点道路改善其与邻接道路的换行条件，在附近铺设新道路、立交进行分流；对使用率低的道路的通行条件进行改善，撤除或关闭多余的交通信号灯和电子眼，降低路段中收费站的收费标准等。

5 结论与讨论

在数据来源方面，与GPS定位数据相比，手机基站数据的优点在于数据获得的便捷性

表2 道路排名及其流量和

Tab. 2 Flow rank and superimposed flow amount			
排名百分比(%)	排名	叠和百分比(%)	叠加流量和
1.04	9	12.55	46721
1.97	17	20.29	75540
20.07	173	75.41	280776
24.36	210	80.03	297956

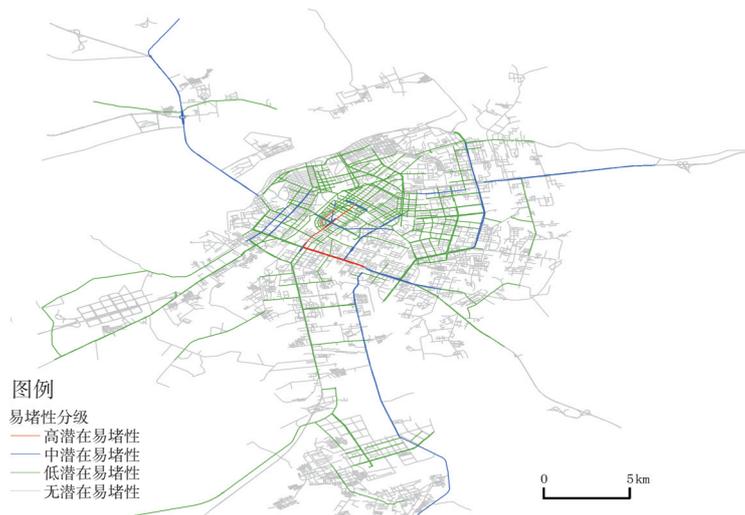


图6 道路潜在易堵性

Fig. 6 The potential jammed road in traffic network

和代表性。由于GPS数据的采集需携带设备,成本较高,限制了可研究对象的数量,依托于出租车平台的GPS数据反映的是此种交通模式下的个体移动,削弱了样本的代表性。手机基站数据依托于当地通讯基础设施,分布范围广,可研究对象数量多,可形成多时间序列数据。但由于其中个体属性信息丰富,保护个人隐私也成为此类研究中的关键问题,对此国外已开始通过立法来解决^[30]。在本文中,通过匿名化预处理将属性信息剔除,只留下基站编号反映个体位置信息,并对研究区和数据获取日期进行模糊化描述,在结果中不涉及当地路段的具体名称,避免个人隐私和研究区位置的泄露。但同时,本文只采用了当地规模最大的手机运营商数据,与所有手机数据相比,此运营商服务的个体在城市空间分布上可能存在一定程度的聚集,从而对当地所有个体的代表性存在偏差,这需要在进一步的研究中通过其他运营商的数据进行修正。

为达到快速实时的应用目的,本文在模拟的有效性和准确性前提下使用系统抽样对运算量进行了压缩,但相比实际交通流量的产生群体,本方法的样本存在偏性。以日均手机通讯频次为标准,产生交通流量的个体可分为每天两次及以上、每天一次或没有、无手机3类,本研究只涉及第一类个体,且由于被动式的位置获取方式,手机提取的起止点模拟轨迹与用户实际移动轨迹可能存在偏差^[3]。但与传统的调查统计数据相比,手机数据克服了其周期长、成本高、精度低、时效性和准确性不高的缺点,已成为未来个体时空行为分析的重要手段,学者们在此基础上进一步结合个体社会经济属性进行了许多相关研究^[4]。而针对偏性的改进可考虑基于空间抽样与统计推断的基本原理,使用优于块状克里金法(Block Kriging)和比率估计法(Ratio Estimator)的B-Shade法进行纠正^[31,32],以使本方法在满足城市尺度的交通流量估算和流量分布特征分析的目的基础上,得到更为精准的流量分布。

针对基站数据定位精度不高的问题,本文利用蒙特卡洛方法,对Voronoi多边形外包矩形内的子区域赋予不同概率,通过随机取点模拟个体的起止点。而在求最短路径时,对单次移动的模拟过程进行改进,提高了方法的处理效率。但精度问题也是基站数据相比GPS定位数据不足的地方,体现在随机点与真实点之间的误差和个体未离开基站范围内的移动模拟上。由于采用二值化的随机方法,个体在Voronoi多边形内的概率处处相等,这种假设与现实不符。但在基站尺度下,静态的人口密度概念不适用于人类移动性的研究目的。可在进一步的研究中结合基站内建筑密度和容积率分布,对高密度区域赋予高概率值,低密度赋予低值,反映个体出现的可能性。对于未离开一个基站范围内的移动过程,数据记录为同一编号的两次相邻通讯,但由于是否移动或者移动是否产生交通流量并不可知,故在模拟中不考虑。

在处理结果上,本文通过系统抽样选取适量样本,以最短路径方法模拟了当日交通流量分布,为识别当地交通网络的重要道路和热点地区提供了可能。对道路流量的统计分析结果揭示了流量分布的20/80现象:小部分路段承担了大部分的流量,支撑着城市道路网络的正常运行;而大部分路段的流量较小,但也为城市道路网络提供了更多可达性。值得注意的是,本文在高速公路的模拟结果上存在一定失真,其所有道路都属于小流量等级,与现实不符。原因是只采用了研究区内的手机基站数据,使用高速公路出入区域的个体起始点或终止点在区域外,数据不可得而未进行模拟,所以出现了失真现象。

基于手机基站数据的交通流量估算方法最重要的意义在于其为重现真实交通流量分布、研究道路网络的交通模式和基于此的情景模拟和规划提供了一种新的思路和方法。借助于海量话单数据和详细的交通网络数据集,可重现数天或者一天中数个时段内的交通流量分布,在此基础上还可以分析不同类型道路在不同时期的流量分布反映出的城市动态景观^[22]。通过截取不同类型的时序数据,可模拟多种情景下的居民出行:对工作日的时序序列,提取个体在工作时间中的位置点集,能确定其大致的工作地点,同理提取个体的居住

地点。对所有个体进行处理,得到当地居民的职住分布,为模拟工作日的通勤出行提供依据;对节假日的非休息时间段进行提取,可得居民休闲出行的目的地分布,为模拟节假日的休闲出行提供依据。同时,根据时间间隔和道路上的距离对个体移动求速度,可判断出交通工具的可能选择,分上下班时间段和休闲时段等不同出行目的进行讨论,可得当地居民的交通模式偏好特征。结合两者,可评定不同道路交通规划思路或方案的好坏以及对交通状况和居民出行的改善程度。

参考文献 (References)

- [1] Lu Zi, Zhao Yahong, Wu Shifeng et al. The time distribution and guide analysis of visiting behavior of tourism website users. *Acta Geographica Sinica*, 2007, 62(6): 621-630. [路紫, 赵亚红, 吴士锋 等. 旅游网站访问者行为的时间分布及导引分析. *地理学报*, 2007, 62(6): 621-630.]
- [2] Mao Xia, Xu Rongrong, Li Xinshuo et al. Fine grid dynamic features of population distribution in Shenzhen. *Acta Geographica Sinica*, 2010, 65(4): 443-453. [毛夏, 徐蓉蓉, 李新硕 等. 深圳市人口分布的细网格动态特征. *地理学报*, 2010, 65(4): 443-453.]
- [3] Liu Yu, Xiao Yu, Gao Song et al. A review of human mobility research based on location aware devices. *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. *地理与地理信息科学*, 2011, 27(4): 8-13.]
- [4] Chai Yanwei, Zhao Ying, Ma Xiujun et al. Mobile positioning method for spatial-temporal behavioral data collection and its geographical applications. *Areal Research and Development*, 2010, 29(6): 1-7. [柴彦威, 赵莹, 马修军 等. 基于移动定位的行为数据采集与地理应用研究. *地域研究与开发*, 2010, 29(6): 1-7.]
- [5] Kwan M, Dijst M, Schwanen T. The interaction between ICT and human activity-travel behavior. *Transportation Research Part A: Policy and Practice*, 2007, 41(2): 121-124.
- [6] Brockmann D, Hufnagel L, Geisel T. The scaling laws of human travel. *Nature*, 2006, 439(7075): 462-465.
- [7] Brockmann D, Theis F. Money circulation, trackable items, and the emergence of universal human mobility patterns. *Pervasive Computing, IEEE*, 2008, 7(4): 28-35.
- [8] Kang C, Gao S, Lin X et al. Analyzing and geo-visualizing individual human mobility patterns using mobile call records//18th International Conference on Geoinformatics, 18-20 June 2010, Beijing.
- [9] Ahas R, Aasa A, Roose A et al. Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management*, 2008, 29(3): 469-486.
- [10] Ahas R, Aasa A, Mark Ü et al. Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Management*, 2007, 28(3): 898-910.
- [11] Nobis C, Lenz B. Communication and mobility behaviour: A trend and panel analysis of the correlation between mobile phone use and mobility. *Journal of Transport Geography*, 2009, 17(2): 93-103.
- [12] Schwanen T, Kwan M. The Internet, mobile phone and space-time constraints. *Geoforum*, 2008, 39(3): 1362-1377.
- [13] Phithakitnukoon S, Horanont T, Di Lorenzo G et al. Activity-aware map: Identifying human daily activity pattern using mobile phone data. *Human Behavior Understanding, Lecture Notes in Computer Science*, 2010, 6219: 14-25.
- [14] Ahas R, Aasa A, Silm S et al. Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C: Emerging Technologies*, 2010, 18(1): 45-54.
- [15] González M C, Hidalgo C A, Barabási A. Understanding individual human mobility patterns. *Nature*, 2008, 453(7196): 779-782.
- [16] Piórkowski M. Sampling urban mobility through on-line repositories of GPS tracks//Proceedings of the 1st ACM International Workshop on Hot Topics of Planet-Scale Mobility Measurements, 22-22 June 2009, Kraków, Poland.
- [17] Candia J, González M C, Wang P et al. Uncovering individual and collective human dynamics from mobile phone records. *Journal of Physics A: Mathematical and Theoretical*, 2008, 41(22): 224015.
- [18] Sevtsuk A, Ratti C. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. *Journal of Urban Technology*, 2010, 17(1): 41-60.
- [19] Yuan Y, Raubal M, Liu Y. Correlating mobile phone usage and travel behavior: A case study of Harbin, China. *Computers, Environment and Urban Systems*, 2012, 36(2): 118-130.
- [20] Song C, Qu Z, Blumm N et al. Limits of predictability in human mobility. *Science*, 2010, 327(5968): 1018-1021.
- [21] Sabbata S D, Mizzaro S, Vassena L. Where do you roll today? Trajectory prediction by space rank and physics models//Gartner G, Rehl K. *Location Based Services and Telecartography II*. Heidelberg: Springer, 2009: 63-78.
- [22] Ratti C, Pulselli R M, Williams S et al. Mobile Landscapes: using location data from cell phones for urban analysis. *Environment and Planning B: Planning and Design*, 2006, 33(5): 727-748.
- [23] Huang W, Dong Z, Zhao N et al. Anchor points seeking of large urban crowd based on the mobile billing data//Cao L,

- Feng Y, Zhong J. *Advanced Data Mining and Applications*. Heidelberg: Springer, 2010: 346-257.
- [24] Calabrese F, Colonna M, Lovisolo P et al. Real-time urban monitoring using cell phones: A case study in Rome. *IEEE Transactions on Intelligent Transportation Systems*, 2011, 12(1): 141-151.
- [25] Strachan S, Williamson J, Murray-Smith R. Show me the way to Monte Carlo: Density-based trajectory navigation// *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2007: 1245-1248.
- [26] Jiang Chengsheng, Wang Jinfeng, Cao Zhidong. A review of geo-spatial sampling theory. *Acta Geographica Sinica*. 2009, 64(3): 368-380. [姜成晟, 王劲峰, 曹志冬. 地理空间抽样理论研究综述. *地理学报*, 2009, 64(3): 368-380.]
- [27] Cao Zhidong, Wang Jinfeng, Li Lianfa et al. Strata efficiency and optimization strategy of stratified sampling on spatial population. *Progress in Geography*, 2008, 27(3): 152-160. [曹志冬, 王劲峰, 李连发 等. 地理空间中不同分层抽样方式的分层效率与优化策略. *地理科学进展*, 2008, 27(3): 152-160.]
- [28] Tobler W R. A computer movie simulating urban growth in the Detroit Region. *Economic Geography*, 1970, 46: 234-240.
- [29] Jiang B. Street hierarchies: A minority of streets account for a majority of traffic flow. *Int. J. Geogr. Inf. Sci.*, 2009, 23 (8): 1033-1048.
- [30] Markkula J. Dynamic geographic personal data: New opportunity and challenge introduced by the location-aware mobile networks. *Cluster Computing*, 2001(4): 369-377.
- [31] Wang Jinfeng, Jiang Chengsheng, Li Lianfa et al. *Space Sampling and Statistical Inference*. Beijing: Science Press, 2009. [王劲峰, 姜成晟, 李连发 等. *空间抽样与统计推断*. 北京: 科学出版社, 2009.]
- [32] Wang Jinfeng, Reis B Y, Hu Maogui et al. Area disease estimation based on sentinel hospital records. *PLoS ONE*, 2011, 6(8): e23428.

Traffic Flow Simulation Based on Call Detail Records

WU Jiansheng¹, HUANG Li¹, LIU Yu², PENG Jian³,
LI Weifeng⁴, GAO Song², KANG Chaogui²

(1. *Shenzhen Graduate School of Peking University, The Key Laboratory for Environmental and Urban Sciences, Shenzhen 518055, Guangdong, China;*

2. *Institute of Remote Sensing and Geographical Information System, Peking University, Beijing 100871, China;*

3. *College of Urban and Environmental Sciences, Peking University, Beijing 100871, China;*

4. *Department of Urban Planning and Design, The University of Hong Kong, Hong Kong, China)*

Abstract: Urban social and economic activity analysis based on mobile location data is a magnificent context for human mobility research and the traffic flow is one of the most basic activities. In order to restore the use of urban transportation network and examine its distribution, we apply a novel approach to draw a traffic flow distribution map of local road network based on a large number of individual cellphone detailed records. We reconstruct details of individual user's mobility and generate its traffic flow step by step: 1. Sampling cellphone records from local operator; 2. simulating the random start point and end point for each individual by Monte Carlo; 3. working out its route through the shortest path. After sampling and simulating thousands of records in one day, we finally draw a traffic flow distribution map of local road network, in which we uncover that a large portion of roads contains a small portion of flows and vice versa. In further statistical analysis, we reach the 20/80 principle of traffic flow: 20% of the top roads accommodate 80% of traffic flow. And flow distribution of different road types reflects the function of urban transportation networks. This research make a contribution to the reconstructed historical traffic flow distribution, studies on urban road network pattern and scenario simulation of transportation planning.

Key words: human mobility; mobile phone data; traffic flow simulation