
第 10 章

地理大数据及其在城市研究中的应用

袁一泓 康朝贵 曹 鹏 刘 瑜

10.1 引 言

随着大数据时代的到来,学者们认识到了大数据对于地理学以及相关学科研究的机遇和挑战(Batty, 2013a; Goodchild, 2013; Kitchin, 2013; Miller and Goodchild, 2015)。目前,关于大数据并没有一个标准的定义(Batty, 2013a)。考虑到大数据这一概念的出现与传感网、移动互联网等信息技术的发展密切相关,通常把如下途径获取的数据归于大数据的范畴:直接的(如监控设备采集的视频影像)、自动的(如手机、公交卡获取的数据)、志愿的(如社交媒体数据)(Kitchin, 2013)。其中第二、三类数据近年来在地理研究中受到了广泛关注,其共同特点是每条记录都具有时空标记并且可以关联到个体,因而一个数据库中包含了大量个体的行为信息。目前,地理大数据一般特指这两类数据,主要包括手机数据、公交卡数据、出租车轨迹、社交媒体数据等。对于一个个体而言,其空间行为具有随机性而难以提取有价值的规律特征,然而当样本量大,一个群体的行为模式的规律性就较为明显,这种规律性与地理环境,尤其是地理环境中的社会经济特征有关。因此,地理大数据为进一步定量理解社会经济环境提供了一种新的观测手段。

10.2 背景

随着不同类型的地理大数据大量涌现,分析居民空间行为模式和城市内部空间结构逐渐成为热门的研究方向。个体和群体时空活动中隐含的规律和所能提供的知识可以为城市规划、交通、旅游、社会关系、疾病传播等诸多领域的研究提供理论基础和指导(Ahas et al., 2008; Cho et al., 2011; Wesolowski et al., 2012)。同时,也促进了地理信息科学与技术的不断发展,使得个体行为研究从数据采集、存储、分析到可视化等整个过程都获得了极大的便利。其中,空间数据库技术及其对时间特征的支持,提供了高效存储和管理个体活动轨迹信息的手段(Zheng and Zhou, 2011)。

与此同时,一系列相关概念被提出,其中比较有影响力的是志愿式地理信息(volunteered geographic information, VGI)、城市计算(urban computing)、社会感知(social sensing)。其中志愿式地理信息(Goodchild, 2007)是指个人自愿提供的地理数据,它的出现与Web 2.0技术以及博客(blog)、维基(Wiki)等应用密切相关,典型的VGI采集网站如OpenStreetMap、Wikimapia等。值得指出的是,Goodchild在提出VGI时,移动互联网和智能手机还没有被广泛应用,因此,在其最初的描述中,并没有涵盖后来基于位置的服务(location-based service, LBS)等产生的地理大数据,尽管这些数据很多也是以志愿方式产生和提交的。由于城市是各类地理大数据产生的最为密集的区域,这些数据为从多个维度刻画城市时空特征提供了充足的信息支持。集成分析这些数据,有助于支持城市管理和规划,服务于政府、企业和城市居民。在此基础上,Zheng等(2014)提出了城市计算的概念。它侧重分析方法,如机器学习、网络分析等,所采用的数据源既包括新型地理大数据,也包括传统地理空间数据与气象数据等。此外,Liu等(2015)提出了社会感知概念,指出地理大数据可从三个方面获取人的行为特征:①对地理环境的情感和认知(例如,基于社交媒体数据可以获取人们对于一个场所的感受);②在地理空间中的活动和移动(例如,基于出租车、签到等数据可以获取海量移动轨迹);③个体之间的社交关系(例如,基于手机数据可以获取用户之间的通话联系信息)。空间大数据包含了海量人群的时空行为信息,使得我们可以基于群体的行为特征揭示空间要素的分布格局、空间单元之间的交互以及场所情感与语义。简而言之,空间大数据提供的社会感知途径,第一次为地理学乃至相关人文社会科学研究打开了一条“由人及地”的研究范式。本章将以手机、出租车、社交媒体等常见地理大数据类型为例,介绍其特点以及在此基础上开展的研究。由于城市是产生地理大数据最为密集的区域,因此大部分研究都针对特定城市开展。

10.3 不同类型大数据的分析方法及应用

10.3.1 出租车轨迹数据分析与挖掘

为了获取城市交通的实时状况并进而支持城市交通管理和优化,往往通过内置全球定位系统(global positioning system,GPS)设备记录出租车的运行轨迹。这些出租车扮演了浮动车的角色,而海量的出租车轨迹数据也成为分析理解居民出行行为和城市功能结构的重要数据源。

10.3.1.1 出租车轨迹数据特征

出租车轨迹数据中的空间位置通过全球定位系统模块确定,这种方式具有空间定位精度高、采样间隔规则等特点。但是,全球定位系统容易在遮挡环境下丢失定位信号,从而在出租车轨迹采集过程中造成不规则的采样间隔,甚至产生数据缺失。考虑到车辆行驶受到道路网络的物理限制,一方面,出租车轨迹在空间上具有近似路网的形态(Liu et al. 2012);另一方面,出租车轨迹容易因较小的空间偏差或较大的采样间隔而偏离真实的行驶路径(Lou et al., 2009)。

结合行车状态信息(包括行驶方向、瞬时速度、载客状态等),出租车轨迹数据还可以提供乘客的出行起讫点(origin-destination,OD)信息。但是,由于无法对乘客进行个体识别,出租车轨迹数据无法反映乘客的连续出行活动。而且,在现实生活中,出租车轨迹数据中标识的出行起讫点通常会偏离乘客真实的活动地点(Gong et al., 2016)。

表 10-1 出租车轨迹记录样例

车辆编号	时间	经度	纬度	方向	速度	状态
1706	2014-05-31 00:00:03	114.312432	30.53968	10	37	重车
12040	2014-05-31 00:00:03	114.182297	30.604683	0	0	空车

出租车轨迹数据同时蕴含了与乘客、司机和道路等关联的出行活动信息,可以支撑从多个视角开展行为规律研究。同时,出租车轨迹数据中活动起讫地点和移动路径的潜在不确定性也给行为规律研究带来了诸多挑战。

10.3.1.2 主要研究方向及进展

出租车轨迹数据相关分析应用主要分为个体尺度与城市(群体)尺度两个方面,本节将分别论述各自领域的相关研究进展。

(1) 个体尺度

从个体尺度出发,关注移动轨迹的形态特征及隐含的居民出行活动,可以从出租车轨迹数据中挖掘三类主要规律(Kang and Qin, 2016)。

第一个方面是乘客出行活动模式分析。个体出行行为受到时空要素的外部制约。一方面,活动的发生位置依赖背景地理环境;另一方面,活动地点间的移动行为受到空间距离衰减效应的影响。基于此,大量研究通过出租车轨迹数据理解城市居民出行起讫点分布的空间分布,试图量化地理要素制约强度和距离衰减程度(Liu et al., 2012a)。目前,城市居民出租车出行距离的指数衰减现象(Liang et al., 2012)已被广泛接受和认可。值得指出的是,另有学者认为,基于路网的出租车出行距离符合幂率(power-law)衰减规律(Jiang et al., 2009)。此外,因受到时空的显著限制,乘客搭乘出租车出行行为往往体现出明显的时间节律和空间重现特征(Peng et al., 2012)。如何基于这些规律识别和分析城市居民的潜在活动地点及活动语义也受到了较多关注(Gong et al., 2016)。近年来,基于群体出行分布模式探索拼车服务(vehicle pooling)以提高交通出行的效率正逐渐发展成为热门的研究方向(Santi et al., 2014)。

第二个方面是对司机驾驶行为模式的分析。为优化出租车运营与服务效率,这类研究主要关注出租车司机的驾驶行为与策略(Liu et al., 2010; Kang and Qin, 2016)。其中,揽客、等客和歇业(包括休息、加油、交接班等)三类行为受到了较多的关注。基于需求热点探测(Chang et al., 2009)、载客历史经验(Li et al., 2011)以及需求量时间序列预测(Moreira-Matias et al., 2012)等的揽客策略被证明具有较好的效果,可以最小化平均能源和时间的消耗(Yuan et al., 2013)。等客行为研究主要考察等客位置和等客时长,常用的策略主要是估算下客点周围的历史上客量经验统计(Li et al., 2011)或转换为揽客策略的相对代价高低。歇业行为主要受经济学领域学者的关注,常作为基准收入依赖(income-reference dependency)和偿失效益(gain-loss utility)理论的验证案例(camerer et al., 1997; Farber, 2015)。城市和交通研究学者相对比较重视出租车的加油行为,为建立合理的加油站布置和排队策略提供辅助支持(Zhang et al., 2015)。

第三个方面是地理学者较常从聚合和网络的视角考察出租车出行行为。通过建立描述不同空间区域空间交互的出行网络,可以进行空间交互网络建模,解释出行行为的发生机制。常见的模型包括引力模型、机会干预模型、辐射吸收模型及其变种等(Yan et al., 2014; Kang et al., 2015a)。同时,在网络聚合层面,通过观察出行网络的短期或长期变化,可以更加直观和有效地发现出租车乘客或司机在城市内部的频繁活动地点、路径和子区域(Guo et al., 2010; Yan et al., 2013),进而估计乘客出行活动的语义信息(Kang et al., 2015b)。这类研究体现了个体尺度出行行为的宏观汇总规律,具有较强的普适性。

(2) 城市尺度

城市研究学者侧重从出租车轨迹数据透视城市功能和运行状态。出行活动作为主体“人”的重要属性,一方面可以直观地反映出“地”(即地理环境,如城市空间结构等)对人类活动的影响;另一方面能够刻画出人类个体对地理空间的利用和改造(Axhausen and Gärling, 1992)。通过考察城市内部不同空间位置(或区域)的出租车出行活动强度及其时态变化,可以揭示城市土地利用分布、交通系统状态、大气环境污染等表征信息,有助于提高人们的生活品质,保护环境和促进城市运转效率(Zheng et al., 2014)。这方面的研究主要包括以下几类:

第一,城市规划识别与评估。准确掌握城市内部不同功能区域的分布对制定合理的城市规划有着极其重要的意义。城市系统的形态、要素之间的相互作用和空间结构(Bourne, 1982; Batty, 2013b)同出租车出行活动具有强关联关系。通过分析城市单个地块以及地块间的出行活动量的时序变化特征,可以揭示城市地块的用地属性(Liu Y et al., 2012b; Yuan et al., 2012; Liu X et al., 2016)。考察局部区域内部不同地块之间的出行活动集散程度,可以识别城市内部的功能子区域(Kang et al., 2013; Liu X et al., 2015)。在理解城市基本功能分布的基础之上,可以进而发现功能(或设施)规划与实际利用程度的对应关系。例如,Chawla等(2012)通过分析大规模车流轨迹数据在不同区域之间的行驶特征,识别连通性较差的区域对,以发现现有城市道路网的不足之处。

第二,交通状态估计与优化。出租车轨迹数据是探测城市道路交通状态参数的最常用的数据源(Castro et al., 2012)。参考出租车轨迹的形态和状态信息,基于轨迹聚类、模糊逻辑、支持向量机、神经网络等挖掘方法,可以有效提取城市道路的交通流量、车流速度和拥堵程度等信息(Li et al., 2007; Yuan J et al., 2010; Kong et al., 2016)。依据道路状态参数,可以进而进行交通预测。例如,Wang等(2014)利用出租车轨迹数据实时估计城市内部任意路线的车辆通行时间,并结合天气以及个人驾车习惯、技能和道路熟悉程度等因素,提出了个性化最快线路设计方法。Wang等(2013)利用出租车轨迹数据探索了城市内部拥堵路段间的状态传导机制。此类研究被证明可以有效降低驾车路程的平均时间,同时指导城市居民选择不同的道路进行出行,以缓解潜在出现的交通拥堵。

第三,能耗与污染物排放估算。此类研究重点关注基于城市交通流估算能耗和污染排放的方法。准确理解整个城市的汽车油耗和尾气排放情况有助于节能减排,对城市持续发展和保护公众健康至关重要。依据环境学理论中的车辆空气污染物排放模型,可根据车速、流量和平均排量计算出能耗和尾气排放情况(Nyhan et al., 2016)。此类研究不但可以向用户建议最低油耗路线,还可以做到细粒度的空气污染预警。对于长期数据的分析,可以发现城市中高能耗路段,以及估算汽车尾气排放的PM2.5占空气中总量的比重,从而为改进城市规划,制

定相关决策提供支持,例如,评估限制交通流量是否真的能够减轻污染状况(Shang et al., 2014)。

10.3.2 基于位置的手机数据

手机定位数据为人类行为规律的研究及众多应用领域提供了实时有效的数据支持,相关研究同样也包括个体尺度(微观层面)和城市尺度(宏观层面)的工作。

10.3.2.1 手机定位数据特征

移动通信设备可通过多种方式获取使用者的位置信息,如通过内置的全球定位系统模块,通过塔台基站(cellular tower)的位置信息,或两者结合(service-provider assisted GPS,A-GPS)。定位方式不同,手机定位数据的完整性及精确度也有所差异,但绝大多数此类数据涵盖以下三类信息:①定位信息;②服务使用信息;③使用者的个人信息,但此类信息通常受隐私保护条例限制,或需要参考外部人口统计学数据库获得(Dashdorj et al., 2013; Liu F et al., 2013; Reades et al., 2007; Yuan and Raubal, 2010)。本节将阐述不同定位技术所带来的手机数据的质量问题。

(1) 基站定位

基站定位作为应用最广泛的手机定位技术之一,通常依靠移动设备所连接的基站塔台来确定设备的即时位置。此类定位数据的精度与基站的分布密度密切相关。在城市中心等繁华地区,定位精度可达200~500 m。而在基站分布稀少的郊区,往往降低至5~10 km(Yuan et al., 2012a)。工业中也常采用三角定位(cell tower triangulation)和差分法(timing advance, TA)来提高基站定位精度(Calabrese et al., 2015)。此外,通过基站定位的手机数据也可能存在采样率的问题。如果使用基于事件触发的数据如呼叫详细记录(call detail records, CDRs),系统只有在用户发出或接收到呼叫信号时才会记录相应塔台位置,所以采样精度与用户对手机的使用频率直接相关。

(2) GPS定位

内置A-GPS模块的智能手机可提供更加精确的定位信息。A-GPS与传统GPS的主要差别在于,A-GPS首先与基站内部的定位服务器进行连接,获取由服务器解析的GPS位置数据,可快速得到更加精确的定位信息(精度可达5~10 m),且不需要依赖客户端的计算性能。这些位置信息也为手机应用研发者提供了极大的便利。

(3) 其他定位技术

除基站定位和GPS定位外,其他相关移动通信技术(如无线局域网络定位、

蓝牙定位等)也在手机定位中起到了重要作用。对于无线局域网络定位,移动设备可以通过连接的无线网络的互联网协议地址进行地理位置反解码。此外,蓝牙定位或惯性导航系统等新技术也为获取手机定位数据提供了全新的思路,如可以通过蓝牙连接来推断设备之间的相对位置。

由于人类活动规律的内在随机性及数据本身的限制,对手机定位数据质量的研究尚处于初级阶段。一些研究者通过数据交叉对比来探索不同类型手机定位数据的应用合理性。如美国田纳西大学萧世伦教授的研究组将精度较低的CDR数据与采样精度较高的基站信令数据(cellular handover)进行对比,量化建模了CDR在人类行为建模中的取样偏差(Zhao et al., 2016)。此外,手机定位数据还具有数据样本量大、时间空间尺度大、不易获得用户背景信息(隐私保护)等特征,为相关行为规律的研究带来了新的挑战。第10.3.2.2节将从个体研究和城市研究两个角度分述本领域的研究方向。

10.3.2.2 主要研究方向及进展

如前文所述,手机定位数据的广泛应用为人类行为规律的建模和分析提供了全新的思路。本节将从个体尺度与城市(群体)尺度分别论述本领域的相关研究进展。

(1) 个体尺度

个体尺度的研究侧重于对个体移动轨迹和出行规律的分析与建模。Larsen等(2006)将移动(mobility)分成5类:①人类在物理空间的移动;②物体在物理空间的移动;③想象空间的移动(imagination travel;例如,通过阅读书籍了解未知领域);④视觉旅行(visual travel;例如,通过互联网在谷歌地图上了解地形地貌);⑤交流沟通所带来的互动(communication travel;例如,和朋友之间的短信电话联系)。这5类移动行为并非相互独立,而基于手机定位的移动行为研究主要侧重于第一类和第五类。其中,计算物理和计算机科学领域的研究者关注对人类基本移动模型的拓展和改进(Rhee et al., 2011)。这一类研究将个体作为物理模型中的节点进行抽象化,而不过多考虑真实地理环境对模型参数和个体用户的影响(Gonzalez et al., 2008; Song et al., 2010)。而地理学界对人类行为规律的研究不局限于物理规律,而是更加侧重人类活动本身与环境的交互(activity-based study),以及地理空间异质性对于个体日常生活的意义(“9点回家还是9点去上班”)。Activity-based study在智能交通和智慧城市领域已经得到广泛应用(Shin et al., 2015; Liu F et al., 2014; Louail et al., 2014)。此类研究针对不同人群的日常活动,生活习惯和行为模式进行建模和分析(Jones et al., 1990),其中最为典型的是对用户活动空间(activity space)的建模。研究者将活动空间定义为“人类日常生活所覆盖的空间范围”(Mazey, 1981)。而手机定位数据为此类

研究提供了可靠的数据源 (Yuan and Raubal, 2016a; Calabrese et al., 2015; Ahas et al., 2015)。Yuan 和 Raubal (2016b) 基于行为空间的大小、形态、内部结构等参考量, 分析了不同城市手机用户行为空间的分布规律。爱沙尼亚塔图大学的研究者针对不同人种和不同生活模式的手机使用者行为空间的不同, 为城市隔离 (urban segregation) 现象提供了新的量化依据 (Silm and Ahas, 2014)。Xu 等 (2015) 则定义了一个新的指标“修正标准距离”(modified standard distance), 来度量以住宅为中心的用户活动空间的延伸范围, 并对同一住宅区的用户活动空间进行分布建模。

另一类基于手机定位数据的研究侧重于物理出行和即时通讯技术的交互影响。研究者将两者的交互作用分为 3 类: 取代、增强、综合 (Mokhtarian, 2009; Abler, 1975)。“取代论”的研究者认为, 通信技术(如智能手机)的发展将显著减少使用者物理出行的需求, 例如, 电话会议的出现降低了当面会谈的必要性 (Goulias, 2008; Pendyala et al., 1991; Saxena and Mokhtarian, 1997)。而“增强论”的研究者持相反意见, 认为信息通信技术反而使得人群的联系更加紧密和活跃, 从而增大物理出行的可能性。(Mokhtarian, 2009)。“综合派”则认为两者兼而有之 (Fiore et al., 2014), 或认为不能用单一的交互作用来描述通信技术对出行的影响 (Nathan et al., 2008; Kwan, 2007; Couclelis, 2004)。随着手机定位数据的广泛应用, 越来越多的研究者试图探寻这一交互作用在不同人口群体中的不同体现 (Nobis et al., 2005; Yuan et al., 2012b)。

(2) 城市尺度

另一大类手机定位研究侧重于城市层面的移动分析。此类研究侧重于利用大时间空间尺度的移动数据, 对城市不同区域的出行规律进行分类与归纳 (Caragliu et al., 2009; Miller, 2009)。研究结果可用于分析城市居民的移动模式与城市土地利用和交通网络等地理要素之间的关系, 突出空间异质性、距离衰减、锚点分布、道路网络等地理规律和环境变量的制约, 尝试建立城市尺度上的人类移动解释性模型, 为城市活动系统的“软性”优化及各种城市问题的有效解决提供科学支撑 (Ratti et al., 2007)。这方面的研究主要包括以下几类 (Yuan and Raubal, 2016a) :

- **城市规划与形态学研究:** 城市的行为特征可以看作城市居民个体行为的综合化和归一化 (Batty, 2005)。Kang 等 (2012) 基于中国东北部 8 个城市的手机数据, 探讨了不同城市形态与城市居民出行规律的关系。Phithakkitnukoon 等 (2010) 进行了更小尺度的研究, 探寻不同城市功能分区 (如商业区, 居民区等) 的行为规律的差异, 为城市规划与决策提供直接的依据。Xu 等 (2016) 定义了 3 个基于手机数据的用户行为指标: ①活动范围尺度; ②主要活动地点; ③移动频率, 并基于这三个指标对深圳和上

海手机用户的活动规律进行了横向对比,进一步分析了城市规划和区域异质性对居民出行的影响。

- **城市热点区域分布:**手机定位数据在城市热点区域的识别中起到了至关重要的作用。在犯罪学研究中,手机数据曾被用于筛选与识别高犯罪率城区(Chainey et al., 2008; Traunmueller et al., 2014; Lu, 2000)。同样的数据也可应用于对交通堵塞及出行规划的研究(Herrera et al., 2010),例如,Xu 等(2016)利用手机数据对不同城市区域的自行车出行频率进行了建模和预测。
- **城市时空动态建模:**此类研究可以作为对城市热点区域研究的拓展延伸,即对不同区域在不同时间段内的特征规律进行归纳建模,以期反映城市的实时动态特征(Schönenfelder, 2006; Yuan and Raubal, 2012; Hasan et al., 2012; Yang et al., 2016)。在麻省理工学院 SENSEable City 实验室的“Real-time Rome”课题中,研究者通过手机数据对意大利罗马的时空特征进行动态检测,为城市规划和决策者提供数据支持,例如,在特殊事件(节假日、体育比赛等)中的人群疏导问题。

综上所述,利用手机定位数据从个体角度出发研究宏观地理问题,有助于建立以人为本的地理模型并进行科学的空间决策(Liu Y et al., 2015)。科学技术的发展既改变了人们的活动模式,也提供了采集海量人类移动数据的可能性。本节概述了基于手机定位数据的人类行为模式分析,其研究成果不仅在挖掘时空模式等领域提供了第一手资料,也可在宏观领域服务于交通规划、犯罪模式分析、传染病控制等应用领域(Agar, 2004; Ahas et al., 2007, Ahas et al., 2010)。而个体层面的分析则有利于更深入地理解用户行为特征,为移动通信领域提供有效的反馈,也为未来的商业拓展提供可能性。

10.3.3 社交媒体数据

最近 10 年间,社交媒体诸如 Facebook、Twitter、Flickr、YouTube 及微博、豆瓣等的出现和发展使得每个人成为互联网的信息生产者。移动互联网的普及进一步提供了定位、异步交互、多种传感器等功能(Hu et al., 2015)。利用带位置的社交媒体数据可以同时分析人的移动和活动模式、场所情感以及社交关系等。

10.3.3.1 情绪分析

用户在社交媒体上发布的状态反映其情绪和情感,因此对这些文本的情绪分析可以了解社会层面的情绪状态。Dodds 等(2011)利用长达 3 年的 Twitter 文本数据,分析了人们情绪的时态变化特征,展示了宏观情绪在从日到周到年的不同时间尺度的周期性波动。Mitchell 等(2013)通过结合对 Twitter 的情绪分析和

传统的人口、健康、经济等调查数据,研究了与不同地区人们表现出的情绪相关的因素,例如,与结婚率呈正相关,与肥胖率呈负相关;并指出社交媒体数据可用于实时反映这些社会指标。Frank 等(2013)利用 Twitter 数据探讨了人们在 Twitter 上的用语变化与移动模式的关系,发现人们表达的愉快程度随着离开常在地点距离的增加而呈现出对数级的增加。Yang 和 Mu(2015)及 Yang 等(2015)利用 Twitter 数据来发现可能的情绪低落人群,并研究了季节和气候对低落情绪的影响,指出在不同的地理区域它们具有不一样的影响。Larsen 等(2015)开发了实时获取 Twitter 数据,对文本进行情绪分析和结果可视化的系统;对获取的数据分析表明,Twitter 上表达出的情绪与世界卫生组织(World Health Organization, WHO)给出的焦虑程度和自杀率呈现相关性。此外,他们用主成分分析对情绪特征进行分解,在排除了周期性的情绪变化之后,额外的情绪波动可用于事件检测。Abbasi 等(2014)对 20 种情绪分析工具在 5 个不同的 Twitter 语料数据集上进行了测试比较。Giachanou 和 Crestani(2016)则提纲挈领地描述了 Twitter 情绪分析的总体流程和不同类型的算法,介绍了相关研究领域,把对这类研究在工具选择方面提供了指导。

10.3.3.2 事件分析与预测

由于社交媒体的实时性,利用其探测事件的发生可以快速发现异常事件并实施应对处理,因此这类研究最近几年也发展迅速。Chae 等(2012)利用周期趋势分解方法,将各主题 Twitter 数量变化分解为总趋势、周期成分和异常事件。Sakaki 等(2012)通过 Twitter 的时间序列来推断事件的发生,并用粒子滤波器找出事件的中心,其开发出的系统可以进行地震的快速反应。Walther 和 Kaisser(2013)首先找出空间中集聚的 Twitter 簇,再通过决策树判断其是否对应现实世界中的事件;Cheng 和 Wicks(2014)采用时空扫描统计(space-time scan statistics, STSS)的方法,利用时空圆柱对 Twitter 进行聚类,从而发现密集的主题所对应的事件。Liang 等(2015)用信号处理的方法,通过平滑滤波和带通滤波来提取 Twitter 中的事件。

犯罪行为作为一类特殊的事件,可以利用社交媒体数据进行分析和预测。Wang 等(2012)初步尝试利用 Twitter 数据,采用文本分析提取出相关主题建立线性回归模型,对驾车逃逸进行预测。Gerber(2014)则进一步将核密度估计方法所得的估计值和从 Twitter 文本中提取出的主题强度同时作为自变量,采用逻辑回归模型对 25 类犯罪进行预测,并在其中的 19 种得到了优于仅使用核密度估计法的结果。Wang 和 Gerber(2015)则利用 Twitter 文本辅助 Foursquare 的签到数据对用户的轨迹进行预测,并建立各类场所预测人口密度与犯罪率之间的相关性,训练支持向量机(support vector machine, SVM)对各类犯罪行为作出预

测。AlBon 和 Gerber(2016)则是利用用户的活动模式,利用空间单元内各类场所的活动强度和时间来训练逻辑回归(Logistic regression, LR)模型,对不同时空单元的各类犯罪率进行预测。Williams 等(2016)对用大数据进行犯罪的相关研究做了回顾,并结合“破窗理论”,探究了在犯罪率不同的地区,社交媒体上人们的发言对犯罪现象的不同指示作用。

10.3.3.3 地点认知

社交媒体数据由于其数据上传的自发性,对城市意象(image of the city)(Lynch, 1960)也提供了有效的支持。Dunkel(2015)通过 Flickr 上的照片及其标签的空间聚类,分析出大众对不同地点的感知。Zhou 等(2014)通过计算机视觉的方法从不同城市的 Flickr 照片中提取出 7 个不同维度的属性赋给城市,用于描述该城市的特征;接着构造机器学习模型识别给定照片所属的城市,识别率表示两城市的相似性,并发现了在空间上越接近的城市倾向于越相似。Oldoni 等(2015)则利用自组织映射(self-organizing map, SOM)收集的声音信号,刻画了不同地点的声音环境特征。Feick 和 Robertson(2014)用 Flickr 标签描述空间单元,探讨了在不同空间尺度下空间单元的标签相似性,显示了标签特性对空间尺度的依赖性。

10.3.3.4 移动性模式分析

尽管采样率较低,但是社交媒体数据仍然提供了观察人的活动和移动的途径。Cho 等(2011)通过社交网络 Brightkite、Gowalla 和手机的定位及社交关系数据,发现人们的短距离出行具有时间和空间上的周期性,而长距离的出行则更多受到社交网络结构的影响。Hawelka 等(2014)用带地理位置标记的 Twitter 数据来研究全球范围的人口移动,分析了出行率、出行距离、目的点的分布、流入流出差异等特征。Liu Y 等(2014)则利用中国的街旁网签到数据研究了人们出行活动的距离衰减规律。以上两篇文章分别在全球和中国的尺度上显示了人口移动的空间交互很大程度上仍然遵循国家或行政区域的界限。McKenzie 等(2015)研究了通过签到量的时间变化模式对不同场所的辨识能力,Chauhan 等(2016)研究了利用 Twitter 信息预测用户可能会前往的地点,体现了利用社交媒体数据对移动出行进行的研究从挖掘向预测过渡。

10.3.3.5 社交网络分析

社交媒体服务中用户的好友关系和互动都可以构成社交网络,对这些社交网络及其与地理空间的关系也是研究者关心的主题。关于社交网络中地理影响的较早研究是 Liben-Nowell 等(2005)对社交关系中距离影响的探讨。Crandall

等(2010)利用Flickr数据,通过时空共现来推断好友关系,而Sharma等(2014)则通过Twitter上的共同行为来推断好友关系。Stephens和Poorthuis(2014)探讨了Twitter网络结构与地理空间结构的关系,表明虽然社交网络使得远距离的交流变为可能,但实际上网络结构和信息传播在很大程度上仍然遵循距离衰减规律。Olteanu等(2016)通过量化隐私风险,研究了空间共现信息中潜在的位置泄露问题。Gao和Bian(2016)分析了尺度效应对社交关系网络分析的影响,对空间潜入社交关系网络研究中尺度的选择提供了参考。

10.3.4 其他数据

除较为常用的社交网络、手机定位数据和出租车数据外,其他类型的地理大数据也在人类行为模式的研究中有所应用,为大数据背景下的定量地理学研究提供了更多元化的解决途径。如:以下几种。

10.3.4.1 公共交通数据

与出租车数据类似,公共交通数据(如公交一卡通、地铁月票等)直接反应了城市居民的日常出行模式。Hasan等(2012)通过对地铁数据的分析,对不同城市区域之间的乘客流量进行建模,研究了城市功能分区的差异。从个体尺度,他们还对单独乘客的出行记录进行了模式挖掘和兴趣点提取(points of interest extraction)。而另一基于法国公交数据的研究则更侧重时间尺度的分析,对不同时间段内的公交乘客活动规律进行聚类(Mahrxi et al., 2014)。Yan等(2014)则结合了不同类型的数据,对城市居民的行为轨迹进行了有效预测。然而,与其他地理大数据类似,公交数据同样存在取样偏差问题,即只对公交服务的使用者而非整个人群进行采样。

10.3.4.2 商业数据

此类研究利用信用卡交易等商业数据对人类行为模式和地理学基本规律进行探索。Brockmann和Theis(2008)对纸钞的流动模式进行追踪,发现其流动轨迹依然可用逆重力模型建模,体现了较强的距离衰减性。Lenormand等(2015)利用信用卡数据分析了西班牙巴塞罗那和马德里两地不同年龄性别的居民的出行差异。

10.3.4.3 大众传媒数据

信息技术的进步不仅带动了社交网络和自媒体的蓬勃发展,也为传统大众传媒提供了新的信息分享平台。随着大量的新闻信息在网络上发布,互联网为传播学、社会学、政治学和地理学的研究者提供了大量实时更新的数据源。例

如,“Global Database of Events, Language, and Tone”(GDELT)是由雅虎、乔治城大学(Georgetown University)及其他合作者联合开发的在线新闻数据库,随时对全球在线新闻信息进行机器编码并存入结构化数据库,便于研究者使用^①(Leetaru and Schrot, 2013)。利用GDELT数据,大量研究者对国际前沿的新闻热点(如国家之间的双边冲突)进行了预测和建模(Hammond and Weidmann, 2014; Jiang and Mai, 2014; Yonamine, 2013)。Yuan和Liu(2015)基于时间序列分析,对中国过去30年内与邻国/友国的外交关系做了聚类分析,并为进一步探讨该数据在政治地理学中的意义提供了定量支持。类似的大众传媒数据库还包括World-Wide Integrated Crisis Early Warning System(ICEWS)。这些数据库具有数据量大、时间空间尺度大、事件影响力大等特点,因此非常适合作为大数据时代国际关系研究和政治地理学的源数据。

10.4 结语

地理大数据在当代信息社会中正得到越来越广泛的应用。本文简述了不同类型(手机数据、出租车数据、社交网络数据等)在人类行为规律建模以及在城市应用中的研究进展。这些基础研究对地理大数据的应用提供了新的思路,如智慧城市、智能交通、灾害应急预警和基于位置的个性化商业广告等应用领域。虽然本章仅集中介绍了手机数据等几种常见数据类型,但所涉及的方法论可广泛应用于其他丰富多样的地理大数据中。这些方法论不仅推进了地理学研究的前沿进展,还为计算机科学、社会科学、传媒学、政治学、城市规划等其他学科提供了全新的研究思路。除方法论外,大量的实证研究(empirical study)也为在信息社会背景下修订传统地理学理论框架提供了不同空间和时间尺度的数据支持。与此同时,地理大数据领域内也存在着许多悬而未决的研究问题,为研究者提供了新的机遇与挑战,例如:

- 统一的理论框架:目前,针对人类行为规律的地理大数据研究多侧重单一算法的实现或某个具体理论的拓展。这些研究往往缺乏统一的地理学理论框架作指导,也常常迷失于过于具体的实证研究中,难以提出高屋建瓴式的理论贡献。目前,地理大数据领域还缺乏类似传统区域科学中“时间地理学”这样灵活而全面的理论框架。
- 数据质量和不确定性问题:不确定性是地理大数据研究不可避免的问题。一方面,由于人的日常行为规律具有内在的不确定性,任何数据都不可能

^① www.gdeltproject.org/

支持百分之百的精确建模；另一方面，数据质量的参差不齐导致研究具有局限性。本章针对手机数据、出租车数据、社交网络数据的常见质量问题，如数据精确度问题、取样偏差、模型偏差等，进行了总结和阐述。然而大部分研究仅将数据质量问题作为研究局限性提出，而并未对其进行有效的量化和验证。如何建立一套行之有效的定量方法来度量地理大数据的不确定性，也是目前研究者们面临的挑战之一。

- 新方法和新理论的实证检验：如前所述，针对人类行为规律的地理大数据研究为城市规划等应用领域提供了大量的新方法和新理论。然而这些研究成果目前多停留在基础研究层面，较少有机会在实际应用中进行检验。未来本领域的研究应注重建立与政府相关部门和决策者的直接合作，使得地理大数据研究能够真正为民所用，为国所用。

参考文献

- Abbasi, A., A. Hassan, and M. Dhar. 2014. Benchmarking twitter sentiment analysis tools. In *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)*. Reykjavik, Iceland.
- Abler, R. 1975. Effects of space adjusting technologies on the human geography of the future. In: Abler, R., D.Janelle, A.Philbrick, and J.Sommer(eds). *Human Geography in a Shrinking World, North Scituate*.Duxburg, MA : Duxburg Press:35–56.
- Agar, J. 2004. *Constant Touch : A Global History of the Mobile Phone*.Cambridge, UK: Icon Books.
- Ahas, R., A.Aasa, U. Mark ,T.Pae, and A.Kull.2007. Seasonal tourism spaces in Estonia: Case study with mobile positioning data. *Tourism Management*, 28(3) : 898–910.
- Ahas, R., A.Aasa, A.Roose, Ü. Mark ,and S.Silm.2008. Evaluating passive mobile positioning data for tourism surveys: An Estonian case study. *Tourism Management*, 29(3) : 469–486.
- Ahas, R., S.Silm, O.Järv, E.Saluveer, and M.Tiru.2010. Using mobile positioning data to model locations meaningful to users of mobile phones. *Journal of Urban Technology*, 17(1) : 3–27.
- Ahas, R., A. Aasa, Y. Yuan, M. Raubal, Z. Smoreda, Y. Liu, C. Ziemlicki, M. Tiru, and M. Zook. 2015. Everyday space-time geographies: Using mobile phone-based sensor data to monitor urban activity in Harbin, Paris, and Tallinn. *International Journal of Geographical Information Science*, 29 (11) : 2017–2039.
- AlBoni, M. ,and M.S.Gerber.2016. Predicting crime with routine activity patterns inferred from social media. In *Proceedings of 2016 IEEE International Conference on Systems, Man, and Cybernetics*. Budapest, Hungary.
- Atefeh, F. ,and W.Khreich.2015. A survey of techniques for event detection in twitter. *Computational Intelligence*, 31(1) : 132–164.

- Axhausen, K. W., and T. Gärling. 1992. Activity-based approaches to travel analysis: Conceptual frameworks, models, and research problems. *Transport Reviews*, 12(4) : 323–341.
- Batty, M. 2005. *Cities and Complexity: Understanding Cities with Cellular Automata, Agent-based Models, and Fractals*. Cambridge, USA : The MIT Press.
- Batty, M. 2013a. Big data, smart cities and city planning. *Dialogues in Human Geography*, 3(3) : 274 –279.
- Batty, M. 2013b. *The New Science of Cities*. Cambridge, UK : The MIT Press.
- Bourne, L. S. 1982. *Internal Structure of the City: Readings on Urban Form, Growth, and Policy*. Oxford, UK : Oxford University Press.
- Brockmann, D., and F. Theis. 2008. Money circulation, trackable items, and the emergence of universal human mobility patterns. *IEEE Pervasive Computing*, 7(4) : 28–35.
- Calabrese, F., L. Ferrari, and V.D. Blondel. 2015. Urban sensing using mobile phone network data: A survey of research. *ACM Computing Surveys*, 47(2) : 1–20.
- Camerer, C., L. Babcock, G. Loewenstein, and R. Thaler. 1997. Labor supply of New York city cabdrivers: One day at a time. *The Quarterly Journal of Economics*, 112(2) : 407–441.
- Caragliu, A., C.D. Bo, and P. Nijkamp. 2009. Smart cities in Europe. *Journal of Urban Technology*, 18 (2) : 65–82.
- Castro, P.S., D. Zhang, and S. Li. 2012. Urban traffic modelling and prediction using large scale taxi GPS traces. In *Proceedings of the 10th International Conference on Pervasive Computing (Pervasive' 12)*. Newcastle, UK.
- Chae, J., D. Thom, H. Bosch, and J. Yun. 2012. Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In *Proceedings of the IEEE Conference on Visual Analytics Science and Technology (VAST)*. Brighton, UK.
- Chainey, S., L. Tompson, and S. Uhlig. 2008. The utility of hotspot mapping for predicting spatial patterns of crime. *Security Journal*, 21(1–2) : 4–28.
- Chang, H.W., Y.C. Tai, and J.Y.J. Hsu. 2009. Context-aware taxi demand hotspots prediction. *International Journal of Business Intelligence and Data Mining*, 5(1) : 3–18.
- Chauhan, A., D. Toshniwal, and R. Tejwani. 2016. Predicting future place of visit using user's personality profile. In *Proceedings of International Conference on Computational Techniques in Information and Communication Technologies*. New Delhi, India: 427–432.
- Chawla, S., Y. Zheng, and J. Hu. 2012. Inferring the root cause in road traffic anomalies, In *Proceedings of IEEE 12th International Conference on Data Mining*. Brussels, Belgium : 141–150.
- Cheng, T., and T. Wicks. 2014. Event detection using Twitter: A spatio-temporal approach. *PLoS ONE*, 9(6) : e97807
- Cho, E., S.A. Myers, and J. Leskovec. 2011. Friendship and mobility: User movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Diego, CA : 1082–1090.
- Couclelis, H. 2004. Pizza over the Internet: E-commerce, the fragmentation of activity and the tyranny of the region. *Entrepreneurship and Regional Development*, 16(1) : 41–54.

- Crandall, D.J., L.Backstrom, D.Cosley, S.Suri, D.Huttenlocher, and J.Kleinberg.2010. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences*, 107 (52) : 22436–22441.
- Dashdorj, Z., L.Serafini, F.Antonelli, and R.Larcher.2013. Semantic enrichment of mobile phone data records. In *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia*. Luleå, Sweden : 35.
- Dodds, P.S., K.D.Harris, I.M.Kloumann, C.A.Bliss, and C.M.Danforth.2011. Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PLOS ONE*, 6 (12) : e26752.
- Dunkel, A.2015. Visualizing the perceived environment using crowdsourced photo geodata. *Landscape and Urban Planning*, 142 : 173–186.
- Farber, H.S.2015. Why you can't find a taxi in the rain and other labor supply lessons from cab drivers. *The Quarterly Journal of Economics*, 130 (4) : 1975–2026.
- Feick, R., and C.Robertson.2014. A multi-scale approach to exploring urban places in geotagged photographs. *Computers, Environment and Urban Systems*, 53 : 96–109.
- Fiore, F.D., P.L.Mokhtarian, I.Salomon, and M.E.Singer.2014. “Nomads at last”? A set of perspectives on how mobile technology may affect travel. *Journal of Transport Geography*, 41 : 97–106.
- Frank, M.R., L.Mitchell, P.S.Dodds, and C.M.Danforth.2013. Happiness and the patterns of life: A study of geolocated tweets. *Scientific Reports*, 3 (7466) : 274–274.
- Gao, P., and L.Bian.2016. Scale effects on spatially embedded contact networks. *Computers, Environment and Urban Systems*, 59 : 142–151.
- Gerber, M.S.2014. Predicting crime using Twitter and kernel density estimation. *Decision Support Systems*, 61 : 115–125.
- Giachanou, A., and F.Crestani.2016. Like it or not: A survey of twitter sentiment analysis methods. *ACM Computing Surveys*, 49 (2) : 1–41.
- Gong L., X.Liu, L.Wu, and Y.Liu.2016. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*, 43 (2) : 103–114.
- Gonzalez, M.C., C.A.Hidalgo, and A.L.Barabasi.2008. Understanding individual human mobility patterns. *Nature*, 453 (7196) : 779–782.
- Goodchild, M.F.2007. Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69 (4) : 211–221.
- Goodchild, M.F.2013. The quality of big(geo) data. *Dialogues in Human Geography*, 3 (3) : 280–284.
- Goulias, K.G.2008. Travel behaviour and demand analysis and prediction. In Meyers, R.A. (ed.). *Encyclopedia of Complexity and Systems Science*. New York, USA: Springer: 9536–9565.
- Guo, D., S.Liu, and H.Jin.2010. A graph-based approach to vehicle trajectory analysis. *Journal of Location Based Services*, 4 (3–4) : 183–199.
- Hammond, J., and N.B.Weidmann.2014. Using machine-coded event data for the micro-level study of political violence. *Research and Politics*, 1 (2) : 1–8.
- Hasan, S., C.Schneider, S.V.Ukkusur, and M.C.Gonzalez.2012. Spatiotemporal patterns of urban

- human mobility. *Journal of Statistical Physics*, 151(1) : 304–318.
- Hawelka, B., I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti. 2014. Geo-located twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41 (3) : 260–271.
- Herrera, J.C., D. B. Work, R. Herring, X. G. Ban, Q. Jacobson, and A. M. Bayen. 2010. Evaluation of traffic data obtained via GPS-enabled mobile phones: The mobile century field experiment. *Transportation Research Part C-Emerging Technologies*, 18(4) : 568–583.
- Hu, X., T.H.S.Chu, V.C.M.Leung, and C.H.Ngai. 2015. A survey on mobile social networks: Applications, platforms, system architectures, and future research directions. *IEEE Communications Surveys and Tutorials*, 17(3) : 1557–1581.
- Jiang, B., J. Yin, and S. Zhao. 2009. Characterizing the human mobility pattern in a large street network. *Physical Review E*, 80(2) : 021136.
- Jiang, L., and F. Mai. 2014. Discovering Bilateral and Multilateral Causal Events in GDELT. In *Proceedings of the 6th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction*. Washington, DC, USA.
- Jones, P., F. Koppelman, and J. P. Orfueil. 1990. Activity analysis: State-of-the-art and future. In: Jones, P.A. (ed.). *Developments in Dynamic and Activity-Based Approaches to Travel Analysis*. UK: Avebury: 34–55.
- Kang, C., X. Ma, D. Tong, and Y. Liu. 2012. Intra-urban human mobility patterns: An urban morphology perspective. *Physica A: Statistical Mechanics and its Applications*, 391 (4) : 1702 –1717.
- Kang, C., S. Sobolevsky, Y. Liu, and C. Ratti. 2013. Exploring human movements in Singapore: a comparative analysis based on mobile phone and taxicab usages. In *Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing*. Chicago, USA : 1–7.
- Kang, C., Y. Liu, D. Guo, and K. Qin. 2015a. A generalized radiation model for human mobility: Spatial scale, searching direction and trip constraint. *PLoS ONE*, 10(11) : e0143500.
- Kang, C., Y. Liu, and L. Wu. 2015b. Delineating intra-urban spatial connectivity patterns by travel-activities: A case study of Beijing, China. In *Proceedings of the 23th International Conference on Geoinformatics*. Wuhan, China : 1–6.
- Kang, C., and K. Qin. 2016. Understanding operation behaviors of taxicabs in cities by matrix factorization. *Computers, Environment and Urban Systems*, 60 : 79–88.
- Kitchin, R. 2013. Big data and human geography: opportunities, challenges and risks. *Dialogues in Human Geography*, 3(3) : 274–279.
- Kong, X., Z. Xu, G. Shen, J. Wang, Q. Yang, and B. Zhang. 2016. Urban traffic congestion estimation and prediction based on floating car trajectory data. *Future Generation Computer Systems*, 61 : 97–107.
- Kwan, M. P. 2007. Mobile communications, social networks, and urban travel: Hypertext as a new metaphor for conceptualizing spatial interaction. *Professional Geographer*, 59 : 434–446.
- Larsen, J., J. Urry, and K. W. Axhausen. 2006. *Mobilities, Networks, Geographies*. Hampshire, UK: Ashgate.

- Larsen, M., T.Boonstra, P.Batterham, B.O' Dea , C.Paris, and H.Christensen.2015. We feel: Mapping emotion on twitter.*IEEE Journal of Biomedical and Health Informatics*, 19(4) : 1246–1252.
- Leetaru, K., and P.Schrodt.2013.GDELT: Global data on events,language ,and tone ,1979–2012.In *Proceedings of the International Studies Association Annual Conference*.San Diego,USA.
- Lenormand, M., T.Louail, O.G.Cantú-Ros, M.Picornell, R.Herranz, J.M.Arias J ,M.Barthelemy, M.S. Miguel, and J.J.Ramasco.2015.Influence of sociodemographic characteristics on human mobility. *Scientific Reports*, 5 : 10075.
- Li, B., D.Zhang, L.Sun, C.Chen, S.Li, G.Qi, and Q.Yang.2011.Hunting or waiting? Discovering passenger-finding strategies from a large-scale real-world taxi dataset.In *Proceedings of the IEEE International Conference on Pervasive Computing and Communications(PERCOM)*.Seattle , USA : 63 –68.
- Li, X., J.Han, J.G.Lee, and H.Gonzalez.2007.Traffic density-based discovery of hot routes in road networks.In *Proceedings of the 10th International Symposium on Spatial and Temporal Databases*. Boston , USA : 441–459.
- Liang, X., X.Zheng, W.Lv, T.Zhu, and K.Xu.2012.The scaling of human mobility by taxis is exponential.*Physica A : Statistical Mechanics and its Applications*, 391 : 2135–2144.
- Liang, Y., J.Caverlee, and C.Cao.2015.A noise-filtering approach for spatio-temporal event detection in social media.*Advances in Information Retrieval*, 9022 : 233–244.
- Liben-Nowell, D., J.Novak, R.Kumar, P.Raghavan, and A.Tomkins.2005.Geographic routing in social networks.*Proceedings of the National Academy of Sciences of the United States of America*, 102 (33) : 11623–11628.
- Liu, F., D.Janssens, G.Wets, and M.Cools.2013.Annotating mobile phone location data with activity purposes using machine learning algorithms. *Expert Systems with Applications*, 40 (8) : 3299 –3311.
- Liu, F., D.Janssens, J.X.Cui, Y.P.Wang, G.Wets, and M.Cools.2014.Building a validation measure for activity-based transportation models based on mobile phone data. *Expert Systems with Applications*, 41 (14) : 6174–6189.
- Liu, L., C. Andris, and C. Ratti.2010.Uncovering cabdrivers' behavior patterns from their digital traces. *Computers ,Environment and Urban Systems*, 34 (6) : 541–548.
- Liu, X., J.Biagioni, J.Eriksson, Y.Wang, G.Forman, and Y.Zhu.2012.Mining large-scale ,sparse GPS traces for map inference: Comparison of approaches.In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.Beijing, China : 669–677.
- Liu, X., L.Gong, Y.Gong, and Y.Liu.2015.Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography*, 43 : 78–90.
- Liu, X., C.Kang, L.Gong, and Y.Liu.2016.Incorporating spatial interaction patterns in classifying and understanding urban land use. *International Journal of Geographical Information Science*, 30 (2) : 334–350.
- Liu, Y., C.Kang, S.Gao, Y.Xiao, and Y.Tian.2012a.Understanding intra-urban trip patterns from taxi trajectory data. *Journal of Geographical Systems*, 14 (4) : 463–483.

- Liu, Y., F. Wang, Y. Xiao, and S. Gao. 2012b. Urban land uses and traffic ‘source-sink areas’: Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning*, 106(1): 73–87.
- Liu, Y., Z. Sui, C. Kang, and Y. Gao. 2014. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLOS ONE*, 9(1): e86026.
- Liu, Y., X. Liu, S. Gao, L. Gong, C. Kang, Y. Zhi, G. Chi, and L. Shi. 2015. Social sensing: A new approach to understanding our socioeconomic environments. *Annals of the Association of American Geographers*, 105(3): 512–530.
- Lou, Y., C. Zhang, Y. Zheng, X. Xie, W. Wang, and Y. Huang. 2009. Map-matching for low-sampling-rate GPS trajectories. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*. San Diego, CA: 352–361.
- Louail, T., M. Lenormand, O. G. Cantu-Ros, M. Picornell, R. Herranz, E. Frias-Martinez, J. J. Ramasco, and M. Barthelemy. 2014. From mobile phone data to the spatial structure of cities. *Science report*, 4: 5276.
- Lu, Y. 2000. Spatial cluster analysis of point data: Location quotients versus kernel density. In *2000 University Consortium of Geographic Information Science (UCGIS) Summer Assembly Graduate Papers*. Oregon, Portland.
- Lynch, K. 1960. *The Image of the City*. Cambridge, USA: The MIT press.
- Mahrsi, M.K.E., E. Côme, J. Baro, and L. Oukhellou. 2014. Understanding passenger patterns in public transit through smart card and socioeconomic data: A case study in Rennes, France. In *Proceedings of the 3rd International Workshop on Urban Computing (UrbComp 2014)*. New York, USA.
- Mazey, M.E. 1981. The effect of a physio-political barrier upon urban activity space. *Ohio Journal of Science*, 81: 212–217.
- McKenzie, G., K. Janowicz, S. Gao, and L. Gong. 2015. How where is when? On the regional variability and resolution of geosocial temporal signatures for points of interest. *Computers, Environment and Urban Systems*, 54: 336–346.
- Miller, H.J. 2009. Geographic data mining and knowledge discovery: An overview. In: Miller, H.J., and J. Han (eds.). *Geographic Data Mining and Knowledge Discovery (Second Edition)*. London, UK: CRC Press. pp.3–32.
- Miller, H.J., and M.F. Goodchild. Data-driven geography. 2015. *GeoJournal*, 80(4): 449–461.
- Mitchell, L., M. R. Frank, K. D. Harris, P. S. Dodds, and C. M. Danforth. 2013. The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place. *PLOS ONE*, 8(5): e64417.
- Mokhtarian, P.L. 2009. If telecommunication is such a good substitute for travel, why does congestion continue to get worse? *Transportation Letters*, 1: 1–17.
- Moreira-Matias, L., J. Gama, M. Ferreira, and L. Damas. 2012. A predictive model for the passenger demand on a taxi network. In *Proceedings of the 15th international IEEE Conference on Intelligent Transportation Systems (ITSC)*. Alaska, USA: 1014–1019.
- Nathan, R., E. M. Getz, E. Revilla, M. Holyoak, R. Kadmon, D. Saltz, and P. E. Smouse. 2008. A

- movement ecology paradigm for unifying organismal movement research. *Proceedings of the National Academy of Sciences of the United States of America*, 105(49) : 19052–19059.
- Nobis, C., B. Lenz, and C. Vance. 2005. Communication and travel behaviour: Two facets of human activity patterns. In Timmermans, H. (ed.) *Progress in Activity-based Analysis*. Oxford, UK: Elsevier. pp.471–488.
- Nyhan, M., S. Sobolevsky, C. Kang, P. Robinson, A. Corti, M. Szell, D. Streets, Z. Lu, R. Britter, S. R. H. Barrett, and C. Ratti. 2016. Predicting vehicular emissions in high spatial resolution using pervasively measured transportation data and microscopic emissions model. *Atmospheric Environment*, 140: 352–363.
- Oldoni, D., B. De Coensel, A. Bockstael, M. Boes, B. De Baets, and D. Botteldooren. 2015. The acoustic summary as a tool for representing urban sound environments. *Landscape and Urban Planning*, 144: 34–48.
- Olteanu, A. M., K. Huguenin, R. Shokri, M. Humbert, and J. P. Hubaux. 2016. Quantifying Interdependent Privacy Risks with Location Data. *IEEE Transactions on Mobile Computing*, 16 (3) : 829–842.
- Pendyala, R. M., K. G. Goulias, and R. Kitamura. 1991. Impact of telecommuting on spatial and temporal patterns of household travel. *Transportation*, 18: 383–409.
- Peng, C., X. Jin, K. C. Wong, M. Shi, and P. Lio. 2012. Collective human mobility pattern from taxi trips in urban area. *PLOS ONE*, 7(4) : e34487.
- Phithakkitnukoon, S., T. Horanont, G. Di Lorenzo, R. Shibasaki, and C. Ratti. 2010. Activity-aware map: Identifying human daily activity pattern using mobile phone data. In Salah, A. A., T. Gevers, N. Sebe, and A. Vinciarelli (eds.). *HBU 2010*. Heidelberg: LNCS, Springer. pp.14–25.
- Ratti, C., A. Sevtsuk, S. Huang, and R. Pailer. 2007. Mobile landscapes: Graz in real time. In: Gartner, G., W. Cartwright, and M. P. Peterson (eds.). *Location Based Services and TeleCartography*. Berlin, Heidelberg: Springer Berlin Heidelberg. pp.433–444.
- Reades, J., F. Calabrese, A. Sevtsuk, and C. Ratti. 2007. Cellular census: Explorations in urban data collection. *IEEE Pervasive Computing*, 6(3) : 30–38.
- Rhee, I., M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong. 2011. On the Levy-walk nature of human mobility. *IEEE/ACM Transactions on Networking*, 19: 630–643.
- Sakaki, T., M. Okazaki, and Y. Matsuo. 2012. Tweet analysis for real-time event detection and earthquake reporting system development. *IEEE Transactions on Knowledge and Data Engineering*, 99 (4) : 919–931.
- Santi, P., G. Resta, M. Szell, S. Sobolevsky, S. H. Strogatz, and C. Ratti. 2014. Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences*, 111 (37) : 13290–13294.
- Saxena, S., and P. L. Mokhtarian. 1997. The impact of telecommuting on the activity spaces of participants. *Geographical Analysis*, 29(2) : 124–144.
- Schönfelder, S. 2006. Urban rhythms: Modelling the rhythms of individual travel behaviour. Zurich: Swiss Federal Institute of Technology.

- Shang, J., Y. Zheng, W. Tong, E. Chang, and Y. Yu. 2014. Inferring gas consumption and pollution emission of vehicles throughout a city. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA : 1027–1036.
- Sharma, U., A. Suman, and S. Shannigrahi. 2014. Inferring social ties from common activities in twitter. In *Proceedings of the 25th ACM conference on Hypertext and social media*. Santiago, Chile.
- Shin, D., D. Aliaga, B. Tunçer, S. M. Arisona, S. Kim, D. Zünd, and G. Schmitt. 2015. Urban sensing: Using smartphones for transportation mode classification. *Computers, Environment and Urban Systems*, 53 : 76–86.
- Silm, S., and R. Ahas. 2014. Ethnic differences in activity spaces: A study of out-of-home non-employment activities with mobile phone data. *Annals of Association of American Geographers*, 104 (3) : 542–559.
- Song, C.M., Z.H. Qu, N. Blumm, and A.L. Barabasi. 2010. Limits of predictability in human mobility. *Science*, 327 : 1018–1021.
- Stephens, M., and A. Poorthuis. 2014. Follow thy neighbor: Connecting the social and the spatial networks on Twitter. *Computers, Environment and Urban Systems*, 53(3) : 331–346.
- Traunmueller, M., G. Quattrone, and L. Capra. 2014. Mining mobile phone data to investigate urban crime theories at scale. In *Proceedings of 6th International Conference on Social Informatics (SocInfo)*. Columbus, USA : Springer; 396–411.
- Walther, M., and M. Kaisser. 2013. Geo-spatial event detection in the twitter stream. *European Conference on Information Retrieval*, 7814 : 356–367.
- Wang, M., and M.S. Gerber. 2015. Using Twitter for next-place prediction, with an application to crime prediction. In *Proceedings of 2015 IEEE Symposium Series on Computational Intelligence*. Cape Town, South Africa.
- Wang, X., M.S. Gerber, and D.E. Brown. 2012. Automatic crime prediction using events extracted from twitter posts. International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction, 7227 : 231–238.
- Wang, Y., Y. Zheng, and Y. Xue. 2014. Travel time estimation of a path using sparse trajectories. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, USA : 25–34.
- Wang, Z., M. Lu, X. Yuan, J. Zhang, and H. V. D. Wetering. 2013. Visual traffic jam analysis based on trajectory data. *IEEE Transactions on Visualization and Computer Graphics*, 19(12) : 2159–2168.
- Wesolowski, A., N. Eagle, A.J. Tatem, D.J. Smith, A.M. Noor, R.W. Snow, and C.O. Buckee. 2012. Quantifying the impact of human mobility on malaria. *Science*, 338(6104) : 267–270.
- Williams, M.L., P. Burnap, and L. Sloan. 2016. Crime sensing with big data: the affordances and limitations of using open source communications to estimate crime patterns. *The British Journal of Criminology*, doi: 10.1093/bjc/azw031.
- Xu, Y., S.-L. Shaw, Z. Zhao, L. Yin, Z. Fang, and Q. Li. 2015. Understanding aggregate human mobility patterns using passive mobile phone location data: A home-based approach. *Transportation*, 42 (4) : 625–646.

- Xu, Y., S.-L. Shaw, Z. Fang, and L. Yin. 2016. Estimating potential demand of bicycle trips from mobile phone data-An anchor-point based approach. *ISPRS International Journal of Geo-Information*, 5 (8) : 131.
- Xu, Y., S.-L. Shaw, Z. Zhao, L. Yin, F. Lu, J. Chen, Z. Fang, and Q. Li. 2016. Another tale of two cities: Understanding human activity space using actively tracked cellphone location data. *Annals of the American Association of Geographers*, 106 (2) : 489–502.
- Yan, X. Y., X. P. Han, B. H. Wang, and T. Zhou. 2013. Diversity of individual mobility patterns and emergence of aggregated scaling laws. *Scientific Reports*, 3: 2678.
- Yan, X. Y., C. Zhao, Y. Fan, Z. Di, and W. X. Wang. 2014. Universal predictability of mobility patterns in cities. *Journal of The Royal Society Interface*, 11 (100) : 20140834.
- Yang, W., L. Mu, and Y. Shen. 2015. Effect of climate and seasonality on depressed mood among twitter users. *Applied Geography*, 63: 184–191.
- Yang, W., and L. Mu. 2015. GIS analysis of depression among Twitter users. *Applied Geography*, 60: 217–223.
- Yang, X., Z. Fang, Y. Xu, S.-L. Shaw, Z. Zhao, L. Yin, T. Zhang, and Y. Lin. 2016. Understanding spatio-temporal patterns of human convergence and divergence using mobile phone location data. *ISPRS International Journal of Geo-Information*, 5 (10) : 177.
- Yonamine, J.E. 2013. Predicting future levels of violence in Afghanistan district using GDELT. Dallas, USA : The University of Texas at Dallas.
- Yuan, J., Y. Zheng, C. Zhang, W. Xie, X. Xie, G. Sun, and Y. Huang. 2010. T-drive : Driving directions based on taxi trajectories. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*. San Jose, USA : 99–108.
- Yuan, J., Y. Zheng, and X. Xie. 2012. Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Beijing, China : 186–194.
- Yuan, N.J., Y. Zheng, L. Zhang, and X. Xie. 2013. T-finder: A recommender system for finding passengers and vacant taxis. *IEEE Transactions on Knowledge and Data Engineering*, 25 (10) : 2390 –2403.
- Yuan, Y., and M. Raubal. 2010. Spatio-temporal knowledge discovery from georeferenced mobile phone data. In *Proceedings of the 2010 Movement Pattern Analysis*. Zurich, Switzerland.
- Yuan, Y., M. Raubal, and Y. Liu. 2012. Correlating mobile phone usage and travel behavior—A case study of Harbin, China. *Computers, Environment and Urban Systems*, 36: 118–130.
- Yuan, Y., and M. Raubal. 2012. Extracting dynamic urban mobility patterns from mobile phone data. In *Geographic Information Science—7th International Conference*. Columbus, USA : Springer; 354 –367.
- Yuan, Y., M. L. Giger, H. Li, N. Bhooshan, and C. A. Sennett. 2012. Correlative analysis of FFDM and DCE-MRI for improved breast CADx. *Journal of Medical and Biological Engineering*, 32 (1) , 42 –50.
- Yuan, Y., and Y. Liu. 2015. Exploring inter-country connection in mass media; A case study of China.

- In *Proceedings of the International Conference on Location-based Social Media*. Athens, Georgia.
- Yuan, Y., and M. Raubal. 2016a. Exploring georeferenced mobile phone datasets—A survey and reference framework. *Geography Compass*, 10: 239–252.
- Yuan, Y., and M. Raubal. 2016b. Analyzing the distribution of human activity space from mobile phone usage: An individual and urban-oriented study. *International Journal of Geographical Information Science*, 30(8), 1594–1621.
- Zhang, F., N.J. Yuan, D. Wilkie, Y. Zheng, and X. Xie. 2015. Sensing the pulse of urban refueling behavior: A perspective from taxi mobility. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3): 37.
- Zhao, Z., S.-L. Shaw, Y. Xu, F. Lu, J. Chen, and L. Yin. 2016. Understanding the bias of call detail records in human mobility research. *International Journal of Geographical Information Science*, 30(9): 1738–1762.
- Zheng, Y., and X. Zhou. 2011. *Computing with Spatial Trajectories*. New York, USA: Springer Science and Business Media.
- Zheng, Y., L. Capra, O. Wolfson, and H. Yang. 2014. Urban computing: concepts, methodologies, and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3): 38.
- Zhou, B., L. Liu, A. Oliva, and A. Torralba. 2014. Recognizing city identity via attribute analysis of geo-tagged images. In *Proceedings of European Conference on Computer Vision*. Zurich, Switzerland.

作者简介

袁一泓,德克萨斯州立大学地理系助理教授,主要从事地理大数据的分析与建模。研究方向包括个体行为模式分析、城市计算、时空数据不确定性等。E-mail:yuan@txstate.edu

康朝贵,武汉大学遥感信息工程学院地理信息工程系教师,主要从事时空活动数据挖掘和分析相关研究和教学,尤其关注个体轨迹数据在城市信息学领域的应用。E-mail:cgkang@whu.edu.cn

曹鹏,北京大学遥感与地理信息系统研究所硕士研究生,主要从事社交媒体数据的挖掘研究,E-mail:caopeng@pku.edu.cn

刘瑜,北京大学遥感与地理信息系统研究所教授,主要从事地理信息科学领域的理论及方法研究,目前关注地理大数据的挖掘及应用。E-mail:liuyu@urban.pku.edu.cn