

基于公共交通智能卡数据的城市研究综述*

龙 瀛 孙立君 陶 遂

提 要 基于智能卡的公共交通自动计费系统的广泛应用产生了大量的基于个体的微观时空数据。这种数据不仅记录了持卡者的出行行为, 同时也在个体维度揭示了城市空间的使用模式。由于其时空信息丰富、成本低、样本量大的优势, 公交卡数据受到了越来越多的城市/交通规划研究者的青睐, 开始在城市规划研究中发挥着日益重要的作用。本文主要总结了公共交通智能卡数据在定量城市研究中的应用, 综述并比较了国际国内基于公共交通智能卡数据研究的最近进展。综述涵盖了(1)数据处理与起讫推算, (2)公共交通系统的运行与管理, (3)城市空间结构分析, 以及(4)出行行为与社会网络等四个方面。最后, 总结了已有研究, 简要介绍了隐私保护和信息获取等问题, 并指出未来可能的研究方向。

关键词 公交系统; 城市空间结构; 城市研究; 出行行为

中图分类号 TU984
文献标识码 A
DOI 10.16361/j.upf.201503009
文章编号 1000-3363(2015)03-0070-08

作者简介

龙 瀛, 博士, 北京市城市规划设计研究院, 高级工程师,
longying1980@gmail.com
孙立君, 新加坡-ETH 中心未来城市实验室(Future Cities Laboratory), 博士后, lijun.sun@ivt.baug.ethz.ch
陶 遂, 澳大利亚昆士兰大学地理、规划与环境管理学院, 博士生,
s.tao@uq.edu.au

A Review of Urban Studies Based on Transit Smart Card Data

LONG Ying, SUN Lijun, TAO Sui

Abstract: The wide application of smart card for automated fare collection in public transit systems has produced large quantities of spatial-temporal data at an individual level. Such data not only records mobility behavior of cardholders, but also reveals the spatial functions of cities. Owing to its high spatial-temporal resolution, low cost and large quantity, transit smart card data attract a lot of attention from urban/transport planners and urban researchers. This article presents a comprehensive review of the latest quantitative urban studies enabled by smart card data from both China and overseas. The review covers the following four aspects: (1) data processing and origin-destination inference, (2) transit operation and management, (3) spatial structure of cities, and (4) mobility behavior and social networks. Finally, the review summarizes existing studies and gives a brief introduction to issues of privacy protection and information extraction. Areas of future research are also pointed out.

Keywords: transit systems; urban spatial structure; urban studies; mobility behavior

1 研究概况

随着信息通讯技术的大力发展以及政务公开运动的稳步推进, 手机、公共交通智能卡和信用卡等产生的大数据 (big data) 和来自商业网站和政府网站的开放数据 (open data) 共同促进了当前“新数据环境 (new data environment)”的形成。新数据环境下所产生的从多个维度描绘微观尺度下的人类活动和移动、以及环境要素特征的微观个体 (社会层面和物理空间层面) 数据, 与“以人为本”的新型城镇化和智慧城市建设不谋而合, 对城市空间结构和交通出行行为等研究起到了重要支撑作用。在这样的背景下, 笔者对当前国际和国内基于公共交通智能卡数据 (smart card data 或 SCD) 的城市研究进行综述, 并分析已有研究的不足, 提出未来可能的研究方向, 以期后续有更多学者基于 SCD 的相关城市研究工作开展。

* 国家自然科学基金项目“公共交通智能卡数据研究网络”(编号 51408039)

1.1 公共交通智能卡数据

城市公共交通（仅限于地面公交和城市轨道交通）作为一种空间资源占用量少、低碳节能、社会总出行成本低的交通方式，日益成为解决城市当前面临的各种交通问题的主要出路，这点已经成为很多国外城市发展的基本共识。目前，中国的城市公交优先发展已上升到国家战略层面，很多城市都在各个层面的规划或政策中提出了提高公共交通出行比例的目标。公共交通系统的自动计费(automatic fare collection或AFC)系统也较为普及(Blythe, 2004)。截至2007年，中国100个城市推行了AFC，需要乘客使用非接触式射频智能卡(RFID)或磁卡完成交易。除了计费的主要功能外，AFC也产生了丰富的公共交通智能卡持卡人出行信息的公共交通智能卡数据。

SCD的数据结构比较简单，一般记录了持卡人的ID、类型(如普通卡、学生卡和员工卡等)、上车/下车或上下车的详细时间和线路、车站编号，部分还记录了司机和车辆的ID。考虑到不同城市的公交计费方式不同，SCD的时空信息丰富程度也不同。如有的城市要求上下车都刷卡，费用根据乘车距离计算(distance fare)，如2015年1月之后的北京和目前的伦敦，以及多数城市的轨道交通。有的城市要求仅仅上车刷卡或者下车刷卡(如上海)，即一票制(flat fare)。有的城市则两种模式并用(如2005年之后的北京)。考虑到有的SCD仅仅记录线路和站点ID，对SCD进行额外的空间化(geocoding)的处理则成为其研究应用必要的一环，如将SCD记录的站点ID与经纬度坐标的地理位置进行匹配。此外，大量城市非一票制公交收费系统产生的SCD，其出行的起讫点信息均不完整，因此也需要大量工作对SCD进行信息补全。有大量研究致力于此。

考虑到公共交通智能卡多为匿名卡，申请时不需要提供个人信息，因此SCD大多不包含持卡人的社会经济属性信息。此外，部分城市的公共交通智能卡还可以用于日常消费(如在便利店和超市使用等)，所以有些SCD还包括消

费记录。笔者主要关注交通记录，对消费记录不做探讨。另外需要强调的是，部分城市针对个别群体有免票政策，如老人、残疾人等，这些人不需要刷卡上车，SCD则不包括这些人群的出行信息。

总体来讲，相比传统的交通出行调查数据，SCD的特点一般包括连续性好、覆盖面广、信息全面且易于动态更新，具有地理标识(geo-tagged)和时间标签，同时获取成本较低。因此SCD可以作为大数据的一种来支持城市研究工作。考虑到国家大力倡导公交发展，我国的公共交通覆盖率将继续上升。同时随着个别城市，如北京(2015年1月)，将公交改为分段计价(上下车刷卡)，SCD的价值也将继续提高。这些包含了大规模人群的详细出行记录的数据反映了乘客对公交系统以及城市空间的使用，在支持规划设计和评价上将具有巨大的应用潜力。此外，需要强调的是，SCD一般不涉及持卡人的社会经济属性信息，这点是出行调查数据的优势，两类数据的结合在一些研究中已经体现，详见“6.1 已有研究总结”。因此，充分结合传统数据，对SCD进行深度的挖掘分析，有望给城市空间结构、城市交通系统和出行行为等领域的研究带来突破和创新。

1.2 文献选择和分类

所综述的英文文章主要来自Web of Science，中文文章来自中国期刊网，还包括笔者和合作者的部分工作论文(arXiv及北京城市实验室 Beijing City Lab, www.beijingcitylab.com)：① Web of Science的搜索的关键词是，(smart card(s)/smartcard(s)/fare transaction) and (transit/transport/metro/subway/bus) in TITLE, /表示or, (s)表示可有可无，最终的意思是标题中含有前者中至少一个词和后者中至少一个词的论文；②期刊网：标题中含有“公交IC卡”；③工作论文：详见附录的论文列表。

此外，文献选择还侧重于学术期刊论文，不包括学位论文，也剔除了部分会议论文，也没有考虑铁路系统和区域间交通的SCD应用。

笔者共获得了92篇文献。这些基于SCD的研究大致可分为四类文献：第一类是数据处理，包括数据的补齐(比如一票制等)、OD推算和出行目的识别等；第二类是支持公交系统运行与管理；第三类是城市空间结构；第四类是针对社会网络和特定人群的出行行为。

1.3 文献一览

针对所搜集的92篇文献，笔者从多个维度进行了分析：

如：①使用的数据的对应中国40篇(北京、广州、深圳、昆明、南宁、江阴、常州、济南和苏州)，智利2篇(圣地亚哥)，英国7篇(伦敦、诺丁汉、Bradford and Southport)，新加坡9篇，西班牙2篇(Basque Autonomous Community, Gipuzkoa)，斯洛伐克1篇(Žilina)，日本2篇(高知、大阪)，美国5篇(纽约、芝加哥、Minneapolis-St. Paul metropolitan area)，加拿大11篇(Gatineau、Quebec、蒙特利尔)，韩国4篇(首尔)，法国1篇(The metropolitan area of Rennes)，巴西1篇(圣保罗)，澳大利亚2篇(布里斯班)；②28篇归为类别1，22篇为类别2，14篇为类别3，28篇属于类别4，4篇不属于任何分类(如综述、可行性探讨等)；③27篇涵盖公交和轨道数据，50篇纯公交，11篇纯轨道，其他则为没有明确指出；④47篇仅使用了SCD，而没有用到辅助性的小数据如居民出行调查、社会经济等数；3篇结合了调度数据，4篇结合了GTFIS数据，另外16篇结合了车载GPS数据；⑤数据时间跨度上，平均为209d，最小1.5h，最长9y (Devilleine, 等, 2012)，9篇超过或等于2年，43篇数据少于或者等于1周；⑥平均持卡人数量是371万人(中位数为200万人)，最少的1人(李海波和陈学武, 2013)，最大的1940万人(Long, 等, 2014a)；⑦50篇有至少一名中国大陆作者，33篇的第一作者来自中国大陆的研究机构。

从作者分布上看，海外机构的学者主要有，美国的Catherine Seaborn、Flavio Devillaine、Nigel Wilson、周江评，英国的M. Bagchi、Michael Batty、

钟晨, 新加坡的孙立君, 加拿大的 C. Morency、M.P. Pelletier、Ka Lee Alfred Chu, 智利的 M. Munizaga, 澳大利亚的 Mark Hickman、陶遂, 韩国的 Sang Gu lee, 爱尔兰的 Markus Hofmann, 以及香港的王冬根(香港浸会大学)。中国大陆机构的学者主要有中国科学院深圳先进技术研究院尹凌、张帆, 北京航空航天大学马晓磊和冷彪, 北京工业大学的陈艳艳, 北京联合大学的孟斌, 华南理工大学的徐建闽, 重庆市城市交通规划设计研究所的周涛, 东南大学的陈学武, 首都师范大学的罗来平, 以及北京市城市规划设计研究院的龙瀛和张宇。

下面将围绕这 92 篇文献从四个方面对相关研究进展进行介绍(考虑到个别典型研究的价值, 部分英文文献超越了所找到的 92 篇文献)。

2 数据处理与起讫矩阵推算

推算起讫矩阵(OD Matrix)是交通规划和管理的一个重要的基础部分。作为传统获取相关数据的方法, 交通行为调研往往费时费力, 同时也很难完全覆盖目标人群。这些缺点可能导致不准确的 OD 推算结果以及延缓 OD 的及时更新。鉴于 SCD 能够提供大量的时空出行数据, 研究者们意识到利用 SCD 可以克服上述的问题, 从而构建更准确、详细的公交乘客出行 OD 矩阵。

通常作为公交收费系统的副产物, SCD 所包含的信息及格式并没有规范化的标准。如前所述, 某些 SCD 只记录部分公交出行记录(如上车点, 时间)而忽略其他重要的出行信息(如下车点, 时间, 出行目的等)。因此, 用 SCD 直接推算 OD 就变的很困难。有鉴于此, 很多国外研究者提出了一系列对 SCD 进行数据补充的方法。Barry 等(2002)率先提出了两条重要的出行模式假设说并推算出纽约公交乘客的出行 OD, 即: ①大部分乘客会回到上一段出行(trip)的终点作为下一段出行的起点; ②大部分乘客当天的最后一段出行的终点和当天的第一段出行的起点一样。图 1 直观描述了基于该

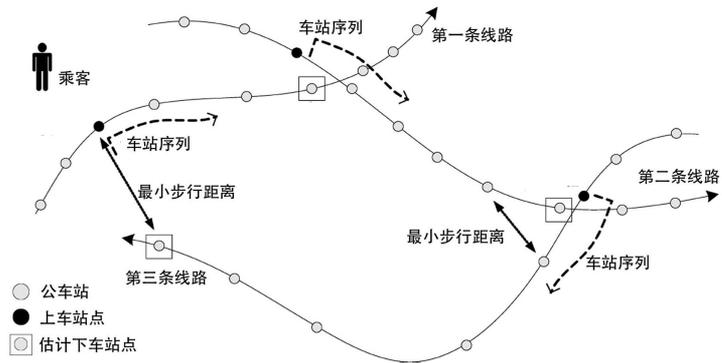


图1 对一般和最终出行下车点的估计模型
Fig.1 The alighting location estimation model for "normal" and "last" trips
来源: Trépanier等(2007).

假设推测乘客下车站点的过程。Barry 等的假设也为之后的 SCD 研究提供了重要的行为理论基础。如 Trépanier 等(2007) 基于类似假设用 Gatineau(加拿大)的 SCD 构建了 OD 矩阵。考虑到公交线路的复杂性, Munizaga 和 Palma(2012)提出了概括时间(generalized time)的概念从而加强了 Barry 等的出行假设的精细程度, 并以此构建 OD 矩阵。有些国外研究则将 SCD 同其他类型的数据, 如 GPS, general transit feed specification(GTFS)及 automatic passenger count(AFC)等结合以获得更可靠、精细的 OD 推算结果, 例如, Farzin(2008), Nassir 等(2011)。

除了 OD 推算, 不少国外研究者提出了进一步丰富补充 SCD 缺失信息(特别是出行目的)的方法。例如, Lee 和 Hickman(2014)的方法结合了用地和 GTFS 数据; Devillaine 等(2012) 的算法基于一系列对不同活动(如工作, 上学等)所需时限的假设; Kusakabe 和 Asakura(2014)的方法则将一个出行调研数据作为基准并利用了朴素贝叶斯分类器(naïve Bayes classifier)。这些方法都有效地对 SCD 进行了出行目的补充。另外, Chapleau 和 Chu (2008)通过核密度估计法(kernel density)找出学生乘客的活动密集点(anchor point)以推测出行目的地。值得注意的是, 类似于 Devillaine 等(2012)的基于时限假设的方法也经常用于识别换乘和构建完整的出行链, 如 Hofmann 和 O'Mahony 等(2005), Seaborn 等(2009)和 Jang(2010)。这对应用 SCD 进一步挖掘和

探索个人的出行行为有着重要的意义。

近年来, 国内很多研究者也开始关注新兴的公共交通智能卡数据。由于缺乏统一的规范标准, 国内的某些公共交通智能卡系统不是很完善, 导致较严重的出行信息的缺失(如, 同时缺失上下车刷卡信息)从而加大了用 SCD 推算 OD 的难度。很多研究提出了针对补充上下车站点等基本出行信息的方法, 包括戴霄和陈学武(2005)、戴霄等(2006)、章威和徐建闽(2006)、周涛等(2007)、徐建闽等(2008)、尹长勇等(2010)、陈君和杨东援(2013)、陈绍辉等(2012a, 2012b)、张宇等(2011)、周雪梅等(2012)、侯艳等(2012)、胡继华等(2014)、张萌萌等(2014)、张颂等(2014)、Luo 和 zhang(2013)、Chen 等(2011)、Chen 和 Yang(2013a)、Ma 等(2012), 以及 Yuan 等(2013)。常用的方法包括对乘客刷卡时间进行聚类分析并同其他数据(如公交调度和公交站地理坐标等)进行匹配, 如陈绍辉等(2012b), 尹长勇等(2010), 张颂等(2014); 以及建立吸引系数模型, 如周涛等(2007), 胡继华等(2014), 周雪梅等(2012), 或概率模型对刷卡站点进行推算, 如 Chen 等(2011), Ma 等(2012), 张萌萌等(2014)。显然, SCD 让大规模的 OD 推算变得更加可行。但在研究交通和城市问题上, SCD 的潜力不止于此。

3 公共交通系统运行与管理

由于用户群的庞大, 公共交通智能

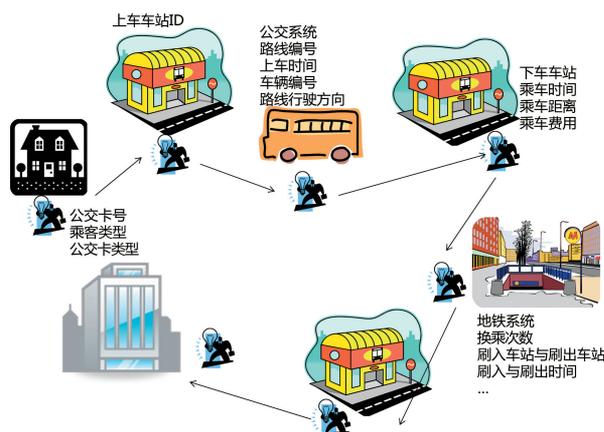


图2 公共交通智能卡记录所包含的个人及出行信息

Fig.2 Personal and travel information captured in smart card transactions

来源: 笔者自绘。

卡系统为交通运输领域的研究者提供了前所未有的海量数据。作为一种新兴的城市大数据,SCD在公共交通的各个层面都发挥着无可替代的作用,包括支持远期规划、服务中期应用以及分析和评价公共交通系统的即时运行状态。Pelletier等(2011)对SCD在城市交通中的应用作了很详细的综述。以这篇文献为基础同时延伸上一节的讨论,本小节着重介绍智能卡数据在交通系统运营与管理中一些新的具有代表性的应用。

在公共交通的远期规划以及中期应用方面,智能卡数据的主要功能是帮助决策者深入理解公交系统的使用模式,从而系统地规划网络线路和建立综合公共交通模型(刘常平等,2008)。长期以来,关于出行者交通行为模式的研究主要是基于入户调查、出行日志和现场调查等方法。这些传统调查方法通常需要消耗大量的人力物力资源,但所得到的数据的质量和准确性一般很难保证。然而智能卡数据以其量大,时长,精准等优势弥补了调查数据的不足,迅速成为交通研究的先锋数据(Iseki,等,2007;Spurr,等,2014)。图2展示了公共交通智能卡数据所包含的用户个人属性以及其出行信息。为了研究智能卡数据的可靠性,Trepanier等(2009)比较了智能卡数据与入户调查数据的出行时空分布。他们发现,通过整合智能卡数据,入户调查数据的精确性和质量可以得到很大提高。Chu和Chapleau(2010)也强调了研究者应该更加充分地利用智

能卡数据的时空精细度,更好地研究用户微观行为。为了更细致地研究北京市的职住关系,Long和Thill(2013)建议将传统入户调查数据与智能卡数据进行结合。

SCD也在研究公交需求上发挥了重要作用。例如,Utsunomiya等(2006)通过研究乘客日常(day-to-day)的出行需求变化证实了提供面向需求的公交服务的重要性,并建议公交运营者根据每天的需求特征来制定具体的运营方案。Lim等(2008)重点研究了公交需求随时间的变化,并用此来预测未来的需求矩阵,为远期规划提供支持。Sun等(2014a)利用从智能卡数据中获取具有时间标记的OD矩阵进行了地铁时刻表的优化。通过优化运行时刻表,高峰时间的需求得到分散,乘客的平均等待时间也相应减少。王冬根(2015)用深圳市一周地铁刷卡数据研究了一系列地铁乘客出行特征,包括出行空间分布、出行距离、出行时间、站间交通流、站间网络关系,由此来刻画轨道交通系统的时空特征。Othman等(2014)利用地铁刷卡数据实现了一个轨道交通的智能体仿真系统,以此来研究轨道交通系统中的各种复杂动力学。

智能卡数据的短期应用价值主要体现在对公交系统的调度、运营和管理以及服务水平的评价上,即对SCD数据的微观应用及重现。在道路交通方面,Zhang等(2014)分别通过重现用户刷卡记录的方法来研究巴士的时空轨迹及占有率。Janosikova等(2014)利用公共交通智

能卡数据重现了乘客在公共交通网络中的路径选择行为,并用这种显示性偏好数据来估计路径选择模型。Sun等(2014b)通过上下车的刷卡记录来研究巴士乘客的上下车行为,并以此衡量了不同车型(如单层,双层和低底板巴士)对站台停滞时间的影响。对于闸口刷卡的轨道交通来说,由于智能卡数据的车辆信息标签缺失,很多研究也致力于解决路径选择未知时轨道交通客流分配的问题上,如刘建峰(2010),以及Sun等(2015)。此外,一些学者利用智能卡数据来衡量公交系统的运行指标和服务水平,如Trépanier等(2009),Ma等(2014)。SCD凭借着其量大质优的特点,已成为帮助解决公交管理问题的一种新兴的高精度,细颗粒的数据。与此同时,SCD的高时空精度也为更为微观的城市空间研究提供了可能。

4 城市空间结构分析

除了支持交通系统运行与管理,如何利用这样大规模的数据理解城市的空间结构成为了城市研究中具有深层意义的一个主要课题。这涉及城市计算(urban computing)的研究,侧重考虑从人的活动体现来城市空间结构,即人如何使用空间,以往更多依赖手机数据、签到数据和出租车轨迹数据(Zheng,等,2011;Yue,等,2014)。在这样的研究中,SCD也逐渐起到重要的数据支持作用,最早的相关文章发表在2011年。

城市的居住、就业和职住平衡是城市空间结构研究的重要问题之一。龙瀛等(2012),Long和Thill(2013)利用北京公交SCD,通过识别持卡人的居住地和就业地,研究北京的职住关系和通勤出行(其后续研究用则也纳入了地铁SCD,如Long的城市贫困研究),结果如图3所示。单纯利用地铁刷卡数据,孟斌(2015)做了类似探讨。作为龙瀛等(2012)的后续研究,Zhou和Long(2014)基于识别通勤出行,评价了北京的过度通勤(excess commute)。Zhou等(2014a,b)也评价了北京的通勤效率(commuting efficiency)情况。此外,Medina和Erath(2013)利用

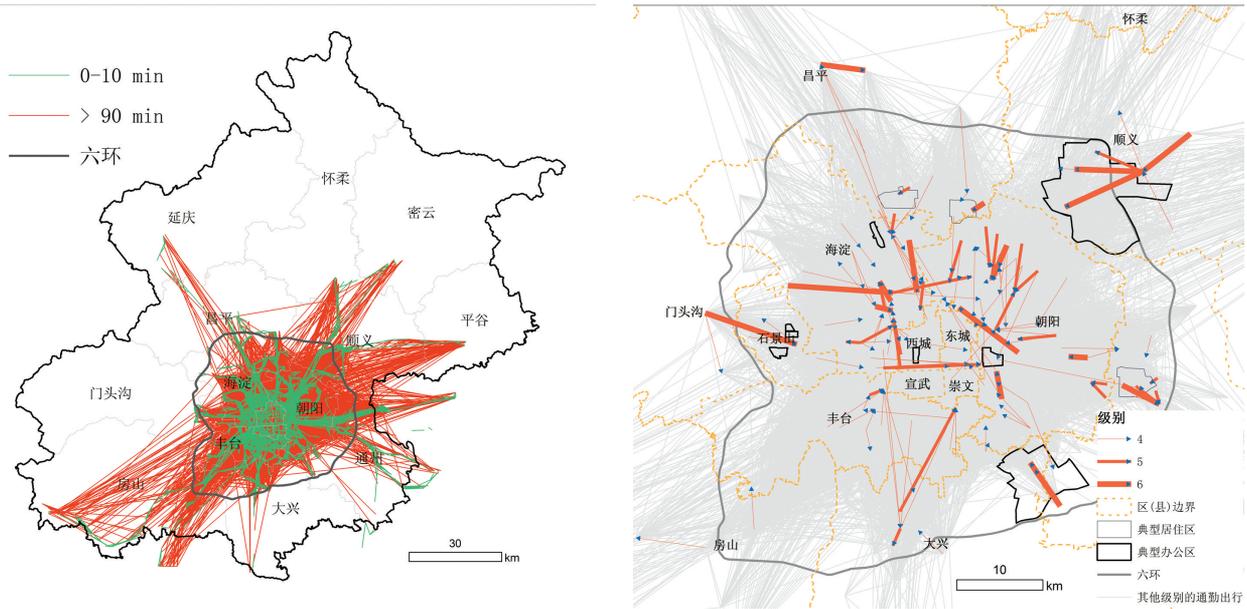


图3 基于公交刷卡数据识别的北京通勤出行 (左: 极端出行时间的通勤出行; 右: TAZ尺度的通勤链接)

Fig.3 Commuting trips identified from Beijing SCD (a) extreme commuting trips (b) commuting trips aggregated at the TAZ level
来源: 龙瀛等(2012).

SCD、出行调查和建筑信息, 提出了基于总步行时间的优化模型, 计算了新加坡的就业密度的时空分布。

通过借鉴复杂网络研究中的社团识别方法(community detection), 研究者们也开始利用个体流动数据来有效认识城市空间结构的变化。这方面 Gao 和 Long(2015)、Zhong 等 (2014)、Sun 等 (2014c)都做了探索, 其中后两者都利用跨越多年的SCD来对比了城市空间结构在较细粒度上的变化, 这点是传统数据难以实现的。

城市的多中心性是城市空间结构的另一个经典命题。对此, 有的学者用SCD来评价城市的多中心性, 如Zhou等 (2014c)在龙瀛等(2012)识别的居住地和就业地的基础上, 进而分别识别了北京的居住中心和就业中心。而 Roth 等 (2011)则利用轨道交通刷卡数据对伦敦的多中心性城市空间结构进行了评价。

最后, 在城市规划评价方面, Long 等(2014b)利用包括SCD在内的多种大数据及小数据评价城市增长边界 (urban growth boundary 或 UGB) 的实施效果。他们的研究结果显示, 虽然有大量的非正式开发分布在UGBs之外, 但UGBs所包含的区域内则容纳了95%以上的城市活动和移动。此外, 也有研究利用SCD

推算城市功能, 如Zhong等(2014)和Han等(2015)分别利用公交SCD识别城市用地功能和推导建筑功能。

5 出行行为与社会网络

SCD不仅在微观层面记录了每个持卡人的详细的出行行为, 在中观或宏观层面上也体现了用户的群体出行特征及相互关系。Bryan 和 Blythe(2007)早期对是否可以利用SCD研究持卡人行为进行了探讨, 目前这个研究方向正在蓬勃发展中。

在宏观层面上, SCD能够反映持卡人出行的时空分布特征。以往这样的研究只能借助小规模的数据, 而SCD为进一步丰富出行行为研究提供了可能。作为城市规划与交通研究中的热点课题, 用户出行行为的研究吸引了很多研究者。这类研究的首要问题是如何利用整理完善的SCD, 探索在不同空间规模下(如一个城市, 地区或一条线路)公共交通持卡人的时空动态特征——例如持卡人如何使用城市空间、是否有规律性, 工作日和周末、不同小时体现的高峰、上下车密度的空间分布等 (Nishiuchi, 等, 2012, Gong, 等, 2012, Kim 等, 2014, Ma, 等, 2014, Tao, 等,

2014ab, Zhao, 等, 2014)。除了基本的时空动态分析外, 部分研究深入探讨了SCD在特定领域的应用。例如, Liu 等 (2009)分析了出行行为与土地使用的关系; Chu 和 Chapleau(2010)利用一个月的SCD提出数据加强的算法, 对个人出行行为进行了深入的多维度分析, 如活动空间和位置、出发时间动态、月内活动安排以及行为规则等; Du 等(2013)利用北京SCD来分析持卡人的聚集特征; Zhang 等(2014)利用多源数据(包括手机、公交车、地铁和出租车数据)建立个人出行行为数据库来反映出行特征; Tao 等 (2014b)利用SCD探索了澳大利亚布里斯班的BRT与非BRT出行的时空特征。需要说明的是, 后两者提倡的是以人为本的城市与交通规划。作为一个交叉方向, 这类研究在城市研究中初露峥嵘, 吸引了社会科学和计算机科学等不同领域的学者。

为了更好地认识了解出行行为, 对所有持卡人进行系统分类则变得有必要。在这方面的研究, 如Agard等(2006)分析并比较了公交使用者的日常公交使用行为的模式, 将所有用户抽象为四个与年龄无关的行为组别; 除了分析持卡人出行行为的时间分布特征, Mohamed 等 (2014) 还将所有持卡人分为16类,

用精细化尺度的社会经济数据来对每类人的居住地进行评价。此外, Páez等(2011)利用SCD进行了持卡人的人口统计学的分析。

基于持卡人个体的特性, 研究者们还试图从SCD中挖掘出一般用户的行为特征, 如出行行为的规律性和变化性(Morency, 等, 2006)。Bagchi和White(2005)利用超过一个月的数据研究了个体行为一致性及随时间的变化。为研究公交用户的周期性行为, Morency等(2007)从长达十个月的智能卡数据中计算了用户的规律性指标。不同于以往的假设, 他们发现个体的出行在空间中并不是稳定的, 而是多样且随时间变化的。此项研究也从侧面说明了城市的非平衡特性。Ma等(2013)提出利用一种基于粗糙集的方法来分类和识别用户的出行模式以及规律性。

需要强调的是上述研究多针对全部的持卡人, 而最新的研究开始将研究目标细化, 关注特定的持卡人群, 从而试图深入理解一些社会问题。如Wang等(2014)研究大学生如何出校园获得社会资源; Long等(2014a)用两年的SCD分析贫困人口并分析其居住地、就业地和通勤情况的变化; Long等(2015)分析四类极端出行人群(早出、晚归、长距离通勤、多次公共交通)。随着人口老龄化问题的普及, Eom和Sung(2011)将研究重心放在了老年持卡人身上, 用一周的数据调查了老年人的出行行为。他们发现老年人会明显地避开早晚出行高峰, 而选在早晚高峰之间的时间出行。

除了持卡人人群的细化, 一部分研究也将关注点放在一些特定的行为上。例如, Batty(2012)研究了伦敦公交罢工期间乘客的行为选择; Frumin和Zhao(2012)利用SCD分析了伦敦的事故发生行为(incidence behavior); Trépanier等(2012)利用3年的SCD评价了乘客对公交服务的忠诚度并分析了影响因素。Asakura等(2012)研究了轨道交通乘客对时刻表的感知情况, 发现乘客会慢慢改变自己的出行时间来适应新的时刻表; Jang(2010)从个体连续公交记录中研究了用户的换乘与换乘地点选择的问题。

与此同时, 还有很多研究者对公交换乘行为进行了细致的研究(Hofman和O'Mahony, 2005; Seaborn, 等, 2009; Wilson, 等, 2009; Lim, 等, 2013; Gordon, 等, 2013; 李海波和陈学武, 2013)。

值得注意的是新加坡国立大学的孙立君近年来在基于SCD研究复杂社会现象做了较早的探索, 图4分别展示了从公交车的一次运行中获得的车上所有乘客的时间相遇网络和某名乘客在一周内的“熟悉的陌生人”网络(Sun, 等, 2013)。这种复杂的身体接触网络为研究大规模的疾病传播提供了可能, 并被用于流行病爆发的识别(Sun, 等, 2014d)。总体上, 近年来这一领域的研究正在走向精细化, 如从所有人群到特定人群, 从所有行为到特定行为。

6 结论与展望

6.1 已有研究总结

笔者系统综述了基于公共交通智能卡数据的城市研究, 主要囊括了数据处理与OD推算、交通系统运行与管理、城市空间结构分析以及出行行为与社会网络等四个研究方向, 还有少量的研究是关于SCD的应用潜力、隐私等问题。总体上的文献特征包括四点: ①国外多于国内; ②近年来变热; ③多面向分析和优化交通系统运行而少针对城市规划的支持; ④规划相关(后两类)的研究国外多于国内, 国内的研究以笔者及合作者的

研究为主, 主要体现在利用SCD评价规划方案, 研究职住平衡, 关注乘客的出行特征及其对空间的使用。这些都可以支持规划编制和评价。

上述研究都紧密围绕着以提高居民生活质量为目标的规划倾向, 如: ①通过科学合理的数据处理, 对交通系统进行分析, 提高公共交通系统的运行效率和人性化水平; ②更为客观地刻画城市空间结构, 通过支持更为合理的决策达到提高居民生活质量的目的; ③通过关注个体出行行为, 提高特定人群的生活质量, 如学生、潜在的城市贫困者、极端出行乘客等; ④通过了解群体出行行为及乘客之间的同乘关系, 预防传染病的传播等。

虽然, 记录每个持卡人详细的上下车信息, 但SCD多为匿名卡, 所以不包含持卡人的社会经济属性信息, 这样就很难用SCD分析不同特征人群的交通行为和偏好。而传统的城市居民家庭交通出行调查数据, 两方面数据(特别是个体社会经济信息)都相对完善。笔者发现, 一半的SCD研究用到了传统的小数据如居民出行调查数据, 显示了SCD与传统数据结合的必要性。已有研究体现的数据结合思路主要有三种: ①利用小数据获得大数据挖掘的规则, 如在龙瀛等(2012)中, 从出行调查数据中识别人的居住地的特征和就业行为的特征, 以及步行距离的概率密度分布, 然后将这些规则用于识别大数据中每个持卡人的居住地和就业地; ②通过出行规律建立SCD

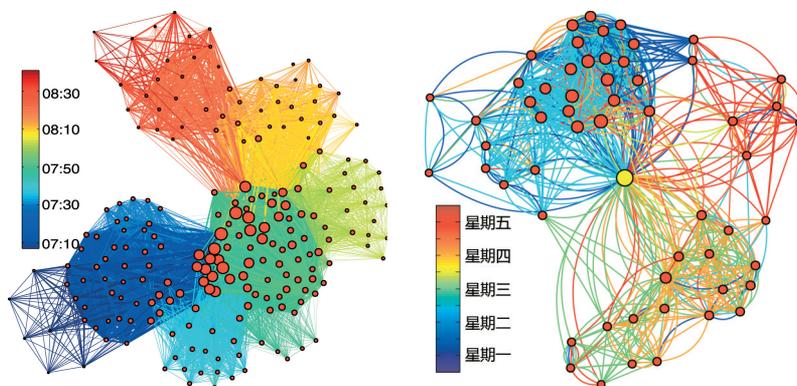


图4 在一辆公交车上的乘客相遇网络 (a) 及个体一周内“熟悉的陌生人”网络 (b)

Fig.4 (a) Temporal encounter network on a single vehicle and (b) Aggregated "familiar stranger" network for an individual
来源: Sun等(2013).

与小数据的联系,进而充分利用两者特有的不同信息优势,如Long等(2015)提出了四种具有不同的极端行为的出行者,然后从SCD识别四类出行者的居住地、就业地、通勤和日常出行的分布,最后从出行调查数据获得了各类出行者的社会经济属性如收入、职业、受教育程度、户口情况、家庭私家车拥有情况等,进而从多个维度刻画这四类群体;③利用社会经济方面的GIS数据来对SCD中体现的空间信息(如居住地)进行背景化(contextualized),如Mohamed等(2014)。

SCD体现的是公共交通出行者的时空轨迹,与微博数据和手机数据的贡献者一样,不涵盖所有的人群。因此,虽然利用SCD支持公共交通系统的运行、管理和优化具有直接优势,在城市空间结构分析和出行行为与社会网络分析两个方面,如果仅仅用SCD来表征社会的整体特征则会失之偏颇。换句话说,特定的用户群导致了SCD的样本有偏性。针对这样的问题,不同的研究采取了不同的应对方法,主要包括:①充分利用SCD的有偏性,例如在中国城市中频繁乘坐公交的出行者中,低收入人群有更大的出现概率,Long等(2014a)利用SCD研究城市贫困对象在2008-2010年间的出行行为和居住地、就业地等变化情况;②关注局部群体的出行行为特征,如Zhou等(2014)利用学生SCD分析北京985高校学生的出行行为,以及Long等(2015)分析四类极端出行人群的出行行为及社会经济特征;③结合其他数据研究同一问题,来保证研究结果的稳定性,如Long等(2014b)利用SCD、传统居民出行调查数据、社交网站签到数据以及出租车轨迹数据,评价北京城市规划的实施效果,发现各类数据的评价结果都表明,超过95%的人类活动与规划边界一致;④关注公共交通使用者反映的城市空间结构,并期待与其他有偏数据反映的结果一起勾勒出完整的城市系统。

6.2 已有研究存在的不足

总体上,截止至此稿撰写,国际国内利用SCD的研究总量已不算稀少,但

还存在以下几点研究上的不足并值得更多学者的关注:①已有研究多关注计算机角度的数据处理、OD识别以及支持公共交通系统的运行与管理,与规划相关的后两类研究为数较少,而该方面的研究国外正在变热。发表在规划刊物上的SCD相关研究更是寥寥无几,在国内,笔者及合作者做了初步的探索性研究;②已有SCD研究多利用少于一周时长的刷卡记录,部分还处于可视化层次,鲜见利用覆盖多年的SCD进行相关研究,如评价公共交通基础设施的建设对城市空间结构的影响,评价特定事件对城市运行的长期影响,以及评价规划的实施情况等。Long等(2014a和2015)和Zhong等(2014)、Sun等(2014c)的研究是例外;③由于SCD的高时空精细度,即使数据经过了匿名处理,数据使用者也可以轻易地在数以百万的用户中辨识特定个体身份。在这种情况下,个人隐私的保护显得尤为重要。如何在隐私保护的前提下尽可能地保持SCD的应用价值日渐成为了新的研究课题(de Montjoye,等,2013);④城市大数据赋予了人们解决不同城市问题的机会,但大多集中于某些领域或某些具体问题。未来的趋势应该是应用城市数据来感知城市的各个方面,作为整个城市的传感器。

6.3 未来研究方向

考虑到已有研究的不足,未来的SCD研究可基于(但不限于)以下几个研究方向:①结合车载GPS数据,开发简单直观的方法,构建完整的持卡人OD网络。多数城市的SCD只包括一次刷卡信息,而不是完整的上下车信息,这个阶段在目前的研究中都比较复杂且耗时,构建完善的OD网络对于开展后续城市研究工作具有较大的支撑作用;②需要进一步加强SCD与各类小数据的结合,丰富信息维度;③使用多年的SCD分析城市动态,评价基础设施建设的影响,规划评估,评价公共交通系统设计等;④继续关注某些特定群体,如学生和城市贫困者,深入研究,比如犯罪预防,以提高居民生活质量。

致谢:本研究受到国家自然科学基金项目(编号51408039),“公共交通智能卡数据研究网络”成员提供了大量研究资料和信息,厦门大学王慧教授对此文提供了宝贵建议,人民大学吴凌燕女士协助整理了文献,在此一并表示感谢。参考文献的Endnote/LaTeX库在北京城市实验室的网站可以下载(<http://www.beijingscitylab.com/projects-1/3-bus-landscapes/>)。

参考文献 (References)

- [1] BAGCHI M, WHITE P R. The potential of public transport smart card data[J]. *Transport Policy*, 2005, 12(5): 464-474.
- [2] DEVILLAINE F, MUNIZAGA M, TRÉPANIÉ R M. Detection of activities of public transport users by analyzing smart card data[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2012, 2276: 48-55.
- [3] GORDON J B, KOUTSOPOULOS H N, WILSON N H M, ATTANUCCI J. P. Automated inference of linked transit journeys in London using fare-transaction and vehicle location data[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2013, 2343: 17-24.
- [4] HAN H, YU X, LONG Y. Discovering functional zones using bus smart card data and points of interest in Beijing[R]. arXiv preprint at arXiv: 1503.03131. 2015.
- [5] JANG W. Travel time and transfer analysis using transit smart card data[J]. *Transportation Research Record: Journal of the Transportation Research Board*, 2010, 2144: 142-149.
- [6] KIM K, OH K, LEE Y K, KIM S, JUNG J Y. An analysis on movement patterns between zones using smart card data in subway networks[J]. *International Journal of Geographical Information Science*, 2014, 28(9): 1781-1801.
- [7] LONG Y, THILL J C. Combining smart card data and household travel survey to analyze jobs-housing relationships in Beijing[R]. arXiv preprint at arXiv: 1309.5993. 2013.
- [8] LONG Y, LIU X, ZHOU J, et al. Profiling underprivileged residents with mid-term public transit smartcard data of Beijing[R]. arXiv preprint at arXiv: 1409.5839. 2014a.
- [9] LONG Y, HAN H, TU Y, ZHU X. Evaluating the effectiveness of urban growth boundaries using human mobility and activity records[R]. Beijing City Lab. Working paper 56. 2014b.
- [10] LONG Y, LIU X, ZHOU J, et al. Early birds, night owls, and tireless/recurring itinerants: an exploratory analysis of extreme transit behaviors in Beijing, China[R]. arXiv preprint at arXiv: 1502.02056.2015.
- [11] MA X, WANG Y. Development of a data-driven platform for transit performance measures using smart card and GPS data[J]. *Journal of Transportation Engineering*, 2014, 140(12): 04014063-01-04014063-12.

- [12] MA X, WU Y J, WANG Y, et al. Mining smart card data for transit riders' travel patterns[J]. Transportation Research Part C: Emerging Technologies, 2013, 36: 1-12.
- [13] MORENCY C, TRÉPANIÉ M, AGARD B. Measuring transit use variability with smart-card data[J]. Transport Policy, 2007, 14(3): 193-203.
- [14] MUNIZAGA M A, PALMA C. Estimation of a disaggregate multimodal public transport origin-destination matrix from passive smartcard data from Santiago, Chile[J]. Transportation Research Part C: Emerging Technologies, 2012, 24: 9-18.
- [15] PELLETIER M P, TRÉPANIÉ M, MORENCY C. Smart card data use in public transit: a literature review[J]. Transportation Research Part C: Emerging Technologies, 2011, 19(4): 557-568.
- [16] ROTH C, KANG S M, BATTY M, et al. Structure of urban movements: polycentric activity and entangled hierarchical flows[J]. Plos One, 2011, 6(1): e15923.
- [17] GAO S, LONG Y. Finding public transportation community structure based on large-scale smart card records in Beijing[R]. Beijing City Lab. Working Paper 59.2015.
- [18] SUN L, AXHAUSEN K W, LEE D H, et al. Understanding metropolitan patterns of daily encounters[J]. Proceedings of the National Academy of Sciences, 2013, 110(34): 13774-13779.
- [19] SUN L, JIN J G, LEE D H, et al. Demand-driven timetable design for metro services[J]. Transportation Research Part C: Emerging Technologies, 2014a, 46: 284-299.
- [20] SUN L, TIRACHINI A, AXHAUSEN K W, et al. Models of bus boarding and alighting dynamics [J]. Transportation Research Part A: Policy and Practice, 2014b, 69: 447-460.
- [21] SUN L, JIN J G, AXHAUSEN K W, et al. Quantifying long-term evolution of intra-urban spatial interactions[J]. Journal of the Royal Society Interface, 2014c, 12(102): 20141089.
- [22] SUN L, AXHAUSEN K W, LEE D H, et al. Efficient detection of contagious outbreaks in massive metropolitan encounter networks[J]. Scientific Reports, 2014d, 4: 5099.
- [23] SUN L, LU Y, JIN J G, et al. An integrated Bayesian approach for passenger flow assignment in metro networks[J]. Transportation Research Part C: Emerging Technologies, 2015, 52: 116-131.
- [24] TAO S, CORCORAN J, MATEO-BABIANO I, et al. Exploring Bus Rapid Transit passenger travel behaviour using big data[J]. Applied Geography, 2014a, 53: 90-104.
- [25] TAO S, ROHDE D, CORCORAN J. Examining the spatial-temporal dynamics of bus passenger travel behaviour using smart card data and the flow-comap[J]. Journal of Transport Geography, 2014b, 41: 21-36.
- [26] TRÉPANIÉ M, MORENCY C, AGARD B. Calculation of transit performance measures using smartcard data[J]. Journal of Public Transportation, 2009, 12(1): 79-96.
- [27] YUAN N J, WANG Y, ZHANG F, et al. Reconstructing individual mobility from smart card transactions: a space alignment approach[C]. 2013 IEEE 13th International Conference on Data Mining (ICDM), IEEE. Dallas, TX, USA. 2013.
- [28] ZHONG C, ARISONA S M, HUANG X, et al. Detecting the dynamics of urban structure through spatial network analysis[J]. International Journal of Geographical Information Science, 2014, 28(11): 2178-2199.
- [29] ZHOU J, LONG Y. Jobs - housing balance of bus commuters in Beijing exploration with large-scale synthesized smart card data[C]. Transportation Research Board 92nd Annual Meeting. Washington, DC. 2014.
- [30] ZHOU J, MURPHY E, LONG Y. Commuting efficiency in the Beijing metropolitan area: an exploration combining smartcard and travel survey data[J]. Journal of Transport Geography, 2014, 41: 175-183.
- [31] ZHOU J, MURPHY E, LONG Y. Visualizing the minimum solution of the transportation problem of linear programming (TLP) for Beijing's bus commuters[J]. Environment and Planning A, 2014, 46(9): 2051-2054.
- [32] ZHOU J W M, LONG Y. Big data for intrametropolitan human movement studies: a case study of bus commuters based on smart card data[C]. Beijing City Lab. Working Paper 50. 2014.
- [33] 陈君, 杨东援. 基于智能调度数据的公交IC卡乘客上车站点判断方法[J]. 交通运输系统工程与信息, 2013(13): 76-80. (CHEN Jun, YANG Dongyuan. Identifying boarding stops of bus passengers with smart cards based on intelligent dispatching data[J]. Journal of Transportation Systems Engineering and Information Technology, 2013(13): 76-80.)
- [34] 李海波, 陈学武. 基于公交IC卡和AVL数据的换乘行为识别方法[J]. 交通运输系统工程与信息, 2013(13): 73-79. (LI Haibo, CHEN Xuewu. A method to identify public transit transfers based on IC and AVL data[J]. Journal of Transportation Systems Engineering and Information Technology, 2013(13): 73-79.)
- [35] 龙瀛, 张宇, 崔承印. 利用公交刷卡数据分析北京职住关系和通勤出行[J]. 地理学报, 2012, 67(10): 1339-1352. (LONG Ying, ZHANG Yu, CUI Chengyin. Identifying commuting pattern of Beijing using bus smart card data[J]. Acta Geographica Sinica, 2012, 67(10): 1339-1352.)

修回: 2015-05