

大数据时代的人类移动性研究

陆 锋, 刘 康, 陈 洁

(中国科学院地理科学与资源研究所 资源与环境信息系统国家重点实验室, 北京 100101)

摘要:人类个体/群体移动特征是多学科共同关注的研究主题。移动定位、无线通讯和移动互联网技术的快速发展使得获取大规模、长时间序列、精细时空粒度的个体移动轨迹和相互作用定量化成为可能。同时,地理信息科学、统计物理学、复杂网络科学和计算机科学等多学科交叉也为人类移动性研究的定量化提供了有力支撑。本文首先系统总结了大数据时代开展人类移动性研究的多源异构数据基础和多学科研究方法,然后将人类移动性研究归纳为面向人和面向地理空间两大方向。面向人的研究侧重探索人类移动特性的统计规律,并建立模型解释相应的动力学机制,或分析人类活动模式,并预测出行或活动;面向地理空间的研究侧重从地理视角分析人类群体在地理空间中的移动,探索宏观活动和地理空间的交互特征。围绕这两大方向,本文评述了人类移动性的研究进展和存在问题,认为人类移动性研究在数据稀疏性、数据偏斜影响与处理、多源异构数据挖掘、机器学习方法等方面依然面临挑战,对多学科研究方法的交叉与融合提出了更高要求。

关键词:人类移动性;大数据;数据挖掘;统计物理学;复杂网络

DOI:10.3724/SP.J.1047.2014.00665

1 引言

人类社会中个体/群体在地理空间的移动过程反映着纷繁复杂的区域人地关系。对这一移动过程特性的研究形成了地理学、社会学、物理学、流行病学、城市规划与管理等学科共同关注的主题——Human mobility。一直以来,Human mobility并无一个统一、公认的中文术语相对应,一般译为人群移动行为^[1]、人类移动性^[2]和人类流动行为^[3]等。为统一起见,本文将Human mobility译为“人类移动性”,表示人类个体/群体在地理空间中具有特定意义的“移动(Movement)”所隐含的社会系统要素时空分布与演化规律。从人类行为和迁移模式,到传染病的演化和疾病的传播,或是理解社会网络的形成机制,人类移动性在许多方面都具有重要的研究价值。例如,某些传染病正是由于人的旅行和相互接触才发生传播,研究群体的空间移动规律有助于深入理解这些疾病的传播机制,进而提出有效的预防和控制措施;在城市空间中,人在不同地点间的

移动直接导致交通网络上的各种复杂的流动现象,只有掌握了人类移动规律,才能合理规划交通设施,进而预防和控制交通拥堵。

人类个体/群体在地理空间的移动有多种表现形式。交通运输工具的位置变化、随身携带设备的位移过程、频率及规模等都是个体/群体空间移动特征的真实写照。过去人类移动性研究多基于观察、访问、调查问卷和出行日志等信息获取方式,信息获取成本高、样本量小、时间跨度短,且易受到问卷设计和主观判断的影响,难以大规模、长时间地观测和记录人的空间移动行为。进入大数据时代,随着传感器网络、移动定位、无线通讯和移动互联网技术的快速发展与普及,获取时空精细度更高的海量个体移动轨迹和相互作用成为可能;同时,移动轨迹处理技术、时空数据表达与挖掘技术的发展和物理学、计算机科学、地理学以及复杂性科学等多学科理论方法的交叉也为人类移动性研究提供了有力支撑,促进了移动行为特征分析的定量化。

图1描述了人类移动性的研究框架。源于现代

收稿日期:2014-07-08;修回日期:2014-08-18.

基金项目:国家自然科学基金资助项目(41271408、41101149);国家“863”计划资助项目(2013AA120305)。

作者简介:陆 锋(1970-),博士,研究员,博士生导师。研究方向为地理信息系统理论与方法、导航与位置服务、空间数据库技术等。E-mail: luf@lreis.ac.cn

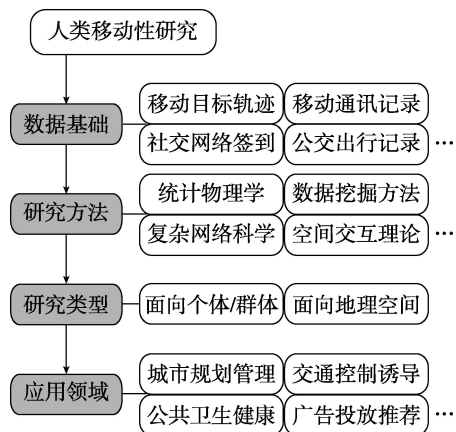


图1 人类移动性研究框架

Fig. 1 Research framework of human mobility

ICT技术的海量移动目标轨迹与多学科交叉的研究方法为人类移动性研究奠定了强有力的基础。在此之上,研究围绕面向人的研究和面向地理空间的研究两个方向展开。面向人的研究侧重人类移动特性的统计规律、移动特性建模、活动模式挖掘、出行或活动预测。面向地理空间的研究侧重人地交互过程和宏观动态活动特征。

2 人类移动性研究数据基础

时空精细化程度更高的海量个体移动轨迹与时空分布数据是量化研究人类移动行为的基础。大数据时代,用于人类移动性研究的数据包括志愿者定位数据、装备卫星导航定位设备的浮动车行驶轨迹、手机终端定位与通讯记录、社交网络签到数据、公交IC卡刷卡记录和公共自行车租赁记录等。

2.1 志愿者定位数据

有些研究通过向志愿者发放GPS接收机来采集个体的移动轨迹信息^[4-5]。例如,微软亚洲研究院的GeoLife项目^[9]采集了128名志愿者近4年的GPS轨迹信息,用于经典旅游路线检测、交通方式判别、生活模式挖掘等研究^[6]。此外,当研究城市居民的通勤、购物、休闲和娱乐等各类活动行为特征时,志愿者GPS数据也可与调查问卷、活动日志等传统的个人时空数据采集方式相结合^[7]。志愿者GPS数据的优点是针对个人的采样率高,即个人轨迹较密集,能够真实反映个人完整的移动轨迹。缺点是样本量较小,不能进行大规模、大范围的人类移动行为分析。

2.2 浮动车定位数据

目前,我国在很多大中城市开展了基于浮动车(多为装备了卫星导航定位与无线通讯设备的出租车)采集方式的交通信息服务。浮动车轨迹包含着丰富的人群移动信息。值得注意的是,此类数据记录了部分乘客在城市路网限制下的部分时空轨迹,反映的多是作为公共交通工具的出租车的行驶轨迹特征^[8-9]和出租车司机的驾驶行为特征^[10]。因此,出租车轨迹数据多用于城市计算^[11]而非个体移动行为分析。

2.3 手机终端定位与通讯记录

相比于志愿者或浮动车定位数据,手机终端定位与通讯记录规模大、时间更连续,可广泛用于个体及群体移动行为规律探索^[12-13]。然而,由于手机终端定位一般来自于移动通信运营商,存在两个明显的局限性:(1)绝大多数运营商只有发生通信行为时才能感知手机用户的位置,因此从手机终端定位数据中构建的用户轨迹只是用户实际出行轨迹的一个粗略概化;(2)手机终端定位一般采取Cell-ID方式,所记录的用户位置是基于通信行为发生时所依附的路由基站位置的统计推断,并非用户精确位置。当然,如果通过在线网络地图服务或者导航系统获取基于WiFi或GPS的智能手机的位置,则可获取更高的用户定位精度。但此类定位数据由移动互联网服务商在用户知情的前提下主动采集,用户规模取决于在线地图类服务的普及程度。

2.4 社交网络签到数据

在基于位置的社交网络服务(如Foursquare、Twitter、新浪微博、盛大切客等)中,将用户通过移动终端上传当前位置的行为称为签到(Check-in),由此产生的带有位置标签(Geo-tag)的数据称为签到数据。签到数据广泛用于人类移动行为分析中^[14-15]。通常情况下,签到数据不但具有地理位置信息,还包含该位置上所承载的兴趣点(Point of Interest, POI)语义信息。因此在行为分析中可省略位置或轨迹语义化操作。然而,签到数据有一个明显的局限性,即用户只有在主动签到时才有记录,导致用户个人轨迹通常比较稀疏。

2.5 公交IC卡消费记录

公交IC卡刷卡数据是一种大规模的具有地理标识和时间标签的数据。公交IC卡目前已被广泛

用于公共交通费用支付,可反映海量城市居民的出行情况,能较全面地覆盖城市人群。通过分析公交IC卡刷卡数据,可有效地反映出持卡用户的出行行为特征、通勤行为与职住关系,以及城市土地利用特征和城市结构^[16-17]等。但此类数据只能粗粒度描述对象在不同刷卡位置间的移动,数据连续性欠佳。

2.6 公共自行车租赁记录

近年来国内外部分城市为满足市民短途出行需求,提供公共自行车租赁公益服务。自行车租赁系统可自动记录所有自行车的出借/归还时间和位置,其租赁记录可以粗略反映一部分城市人群的出行信息,但信息量较少且覆盖人群较为有限,因此研究无法针对个体出行,一般只能从群体层面上对各租赁点的自行车使用情况作时序分析,以反映城市出行行为在空间上的动态变化特征^[18]。

3 人类移动性研究方法

统计物理学、数据挖掘、复杂网络等学科的交叉为人类移动性研究提供了理论方法支撑。不同学科由于关注的研究内容不同,所采用的研究方法也不同,使得人类移动性的研究方法呈现多样化特点。

3.1 统计物理学方法

统计物理学是利用概率统计方法对由大量粒子组成的宏观物体的物理性质及宏观规律作出微观解释的理论物理学分支。统计物理学方法已被广泛用于人类移动性研究中,通过定量统计大量人类行为事件,研究其中所隐藏的统计规律,并根据所研究的问题提出基本假设,建立理论模型,探索这些规律的产生机制和可能的动力学影响^[19-20]。基于这样的研究方法,近年来人们发现了人类移动行为中存在的统计规律,并建立了多种模型尝试对其深层机制进行解释^[21-23]。

3.2 数据挖掘方法

与统计物理学方法不同,数据挖掘方法侧重于从数据驱动的角度出发获取模式或规律。因此,针对海量人类移动数据,数据挖掘方法可以发挥其挖掘模式或规律的长处^[24]。然而,单纯的数据导向性也决定了数据挖掘方法难以直接揭示人类移动性的深层机理,只能根据挖掘的结果去寻求可能的、合理的解释。

3.3 复杂网络分析方法

复杂网络分析方法被广泛应用于交通网络、社交网络等网络的结构分析中^[25,26]。复杂网络的基本度量指标包括节点度、节点中介中心性、聚集系数、平均路径长度等,用以定量分析网络的拓扑结构特征。此外,社区探测是复杂网络研究中的重要主题,它根据网络中各节点间连接的紧密关系将其划分成若干社区,对理解复杂网络的内部构成和相互作用具有重要作用。在人类移动性研究中,复杂网络分析方法常被用来分析地理区域间的人类移动行为^[15,27],以揭示区域空间结构和交互作用。

3.4 空间交互理论方法

空间交互现象广泛存在,如地理区域之间的通勤、人口迁移、物品交换等。基于这些空间交互现象,业界提出空间交互理论,将区域间的流(flows)描述为依赖于起点和终点的某些特征,并和两点间距离呈负相关^[28]。

在人类移动性研究中,空间交互理论方法致力于揭示地理区域之间的人类移动行为规律,并预测地理区域之间的交互强度^[29]。然而,基于各种假设建立的模型在实际使用过程中或过于简化,或所需参数较多。因此,地理区域间群体移动行为特征规律的研究仍有待深入,以建立更加合理、实用的空间交互模型,服务于出行行为预测等应用。

4 人类移动性研究进展

人类个体/群体的移动行为产生于地理空间,又作用于地理空间,因此,本文从人类个体/群体和地理空间两个角度出发,分析人类移动性研究的进展。

4.1 面向人的移动性研究

4.1.1 人类移动特性统计规律

长久以来,由于缺乏精确数据和统计工具,生物迁移被认为是无序的随机游走过程。然而,随着Viswanathan等人发现信天翁的飞行时间近似服从幂律分布^[30]以后,这一观点遭到质疑。自2006年起,Nature、Science等期刊发表了多篇人类移动性研究成果,从统计物理学的角度揭示大规模人群移动步长的统计规律。针对混合交通方式出行,通过分析钞票流通记录^[31]、手机用户轨迹^[12]、志愿者GPS轨迹^[4]以及社交网站签到轨迹^[14]等,发现大规模群体的移动步长服从幂律或截尾幂律分布。针对单

一交通方式出行,通过分析出租车和民航乘客空间的移动特征^[32-33]、地铁和公交乘客的空间移动特征^[34-35]等,发现单一交通方式出行者移动步长一般服从指数分布而非幂律分布。

值得注意的是,群体水平上人类移动步长的幂律分布可能是移动模式各不相同的若干个体混合所致的,并不能据此推断每个个体的移动步长也服从相同的分布规律。闫小勇等对出行日志数据集中每个个体的出行距离分布进行统计,发现绝大多数个体的出行距离并不符合幂律分布,特别是发现学生、职员、退休者等不同个体有着完全不同的日常出行模式^[36]。

4.1.2 人类移动特性建模

前述研究发现人类空间移动模式存在相似特征,暗示着可能存在支配人类空间运动行为的普适性机制。通过建立模型描述人类空间运动行为,有助于揭示各种现象和统计规律形成的底层机制,深化人们对人类移动行为的理解。Brockmann等研究人类的移动规律时发现采用传统的Lévy飞行模型不能合理地再现钞票的空间移动特征,而采用同时具有幂律的步长分布和幂律的停留时间分布的连续时间随机游走模型来模拟钞票的空间移动,能够很好地再现钞票低速扩散的特征^[31]。目前,人们已经发现传统的随机游走模型和单纯的Lévy飞行模型不能合理解释人类的移动特性,并尝试通过唯象机制来解释各种标度特性的可能起源。Song等认为人类同时具有探索未知地点和返回之前熟悉地点的倾向,因此引入探索新地点和偏好返回这两种机制,建立了偏好返回模型^[20];Han等考虑到人们进行日常长距离旅行时由于城市越大交通越便利而导致途经较大城市中转几率也较大这一特征,建立了基于层次性交通系统的人类运动模型^[21];Hu等构建了一个带有返家机制的Lévy飞行模型,并基于个体在游走的过程中希望访问更多不同的地点,以获取尽可能多样的信息的假设,将其转为信息熵优化模型^[22]。

总体来说,现有模型从不同角度对人类空间运动行为中多种统计特性的产生机制进行了有意义的探讨,深化了人们对于自身空间运动行为的理解。但是,人类行为本身是高度复杂的,基于各种假设建立的模型往往过于简化。提出的假设是否合理、是否具有普适性,在人类独特的空间运动模式背后是否还有更为本质的机制尚未被触及,仍有

待于深入探索。

4.1.3 人类活动模式挖掘

个体的活动序列在一定程度上反映了个体的意图、喜好和空间行为模式,从个人移动轨迹中提取出蕴含的活动序列模式,对个体行为理解和个性化推荐至关重要。此类研究通常首先将个体原始轨迹语义化,并建立个体的活动序列,然后利用一定的方法分析和提取个体的活动模式。针对带有活动类别的人类移动数据如活动日志、签到数据,活动类别容易获取,个体的活动时间序列容易建立。针对不带活动类别的人类移动数据如手机定位数据、GPS定位数据,现有研究主要利用外部POIs和土地利用类型^[24,37],或从轨迹中识别出的锚点信息^[38]来获取活动信息,并利用基于空间距离和基于统计等方法语义化轨迹^[39]。然而,目前并没有一种通用的方法适合所有类型的轨迹,各种方法都具有自身的局限性。如何获取和融合更多外部活动信息,提高轨迹语义化的精度,还需进一步研究。

在个人活动时间序列建立的基础上,现有分析和提取个人活动模式的方法主要有聚类、主成分分析、概率模型和序列模式挖掘^[24,38,40]等。各类方法都具有自身的局限性,例如,在利用聚类方法时,如何设定合适的类别个数依然是有待解决的问题;序列模式挖掘方法的评价缺乏统一的标准和框架,其可信度、支持度与感兴趣度等阈值缺乏较好的设定方法,针对海量数据的序列模式挖掘效率还有待提高。

4.1.4 人类出行或活动预测

人类的空间移动具有较强的规则性和可预测性^[41]。预测人类出行或活动对于个性化推荐和交通管理具有十分重要的作用。传统的预测方法大多基于马尔科夫模型^[42]。然而该模型利用历史位置序列信息的能力有限,并且预测的位置只能是历史位置中出现过的,具有较大的局限性。当前人类出行或活动的预测方法主要分为概率模型方法^[43]和机器学习方法^[44-45]两类。前者需要建立合适的概率模型,并将影响个体去向的多种因素考虑进概率模型,后者则需定义合适的特征以提高预测精度。尽管各种预测方法层出不穷,然而不同方法在预测精度、预测稳定性、计算成本等方面各有千秋,不存在普适的方法。从已有研究来看,人类出行或活动预测将会融入更多个人偏好和社交关系等个性化信息,不断改进模型以提高预测性能。

4.2 面向地理空间的移动性研究

人类在地理空间中的移动行为,一方面导致了不同地理区域之间交互强度的差异,另一方面导致了不同区域活动强度的动态变化。从地理空间的视角出发,现有人类移动性研究主要围绕空间交互过程和宏观动态活动两方面展开。

4.2.1 空间交互过程研究

群体在不同地理区域之间的移动行为实际上是一种空间交互过程。空间交互过程既能反映区域之间的社会经济联系强度,又能反映城市的空间结构。

1942年Zipf首次将万有引力定律引入城市间人类移动的研究中^[46]。现今引力模型及其变种已经成为研究地理区域间人类移动行为的最为常见的方法。然而引力模型所需参数较多,且模型的物理意义不够明确,受到很多学者的质疑^[27,43-44]。近年来,研究者提出了一些新模型。Noulas等提出了排序-距离模型,与引力模型相比,能够更为精确地捕捉城市之间的人类移动^[32]。Simini等根据人类空间移动行为特征构造了一个辐射模型,提高了区域间人口流动性的预测精度,而且该模型不需要任何参数^[47]。Lenormand随后提出了一种单一参数模型,并证明其预测精度优于辐射模型^[48]。尽管这些新模型在某些方面的表现优于引力模型,然而,部分原因在于有些研究应用引力模型的方法存疑,即直接用人口数量来近似地理区域的“质量”^[15]。Csáji等利用手机数据研究城市中的通勤时也指出,当用职、住地的数目而不是人口数量近似地理区域的“质量”时,引力模型拟合效果更好^[13]。

除了构建空间交互模型来解释和预测地理区域间的人类个体/群体移动,近年来,越来越多的研究从网络视角来分析地理区域间的人类个体/群体移动行为^[15,27,49]。群体在多个地理区域间的移动可以构建以地理区域为节点、地理区域间的移动强度为边权重的网络结构。除了分析网络节点度、边权重、聚集系数、平均路径长度等属性,社区探测方法也常用来识别移动交互强度紧密的社区结构,以分析区域空间结构、评价行政区划的合理性等。一些研究发现利用社区探测方法划分出的社区边界与现存行政区划较为一致^[27,49],说明人类的移动行为在一定程度上受到固有行政区划的内在限制。

进一步的研究可以将地理空间距离、城市社会经济要素、城市道路网络结构以及自然地理环境等

相结合,探讨多空间尺度下地理区域间的人类移动模式及其影响因素。

4.2.2 宏观动态活动研究

从宏观层面上看,人类群体的移动导致了不同地理区域活动强度的变化。业界通常采用手机基站的话务量^[50-53]或者不同区域出租车乘客的上下车次数^[54]等度量区域的活动强度。现有研究主要包括区域人类群体活动强度时空分析和区域功能区识别2个方面。

区域人类群体活动强度是区域动态性的真实反映。例如,Ratti等以意大利米兰232个手机基站的话务量为依据,开展了“移动景观(Mobile Landscapes)”研究,不仅刻画了城市的活动强度,而且剖析了城市的职住分离现象^[50]。Andrienko等使用自组织映射网络对米兰手机通话量的时空分布进行可视化聚类,研究了城市手机通话量时空分布模式^[51]。

特定空间单元的活动强度时间序列模式可以反映该空间单元的功能类型。现有研究大多通过构建各空间单元的活动强度时间序列,利用时间序列聚类方法^[52]或分类方法^[53-54]识别空间单元的功能类型。然而其识别精度普遍较低。究其原因,一方面是由于同一空间单元中可能存在多种功能类型,另一方面是由于用城市规划图来评价识别精度并不合理。进一步的研究可以考虑改进识别方法,并且融入遥感、POIs等更多信息以提高功能区识别精度。

5 总结与展望

大数据时代唤醒了社会各界对可用于分析的数据的认识,这些不同来源、结构纷杂、随时变化的时空信息产生于生产、生活的各个方面,并且大多具有空间隐喻^[55]。借助大样本、乃至全样本统计分析,有助于发现隐含的、稳健的规律,最大程度上减少小样本随机性带来的不确定性。

同时,地理学、统计物理学、计算机科学、复杂网络科学等多学科理论方法的交叉也为人类移动性研究提供了有力支撑。人类移动性研究的数据基础、研究方法和应用领域都呈现多样化特点,既有人类自身移动特性建模这种微观角度、分解(Disaggregated)层面的研究,又有城市间人群流动性研究这种宏观角度、聚集(Aggregated)层面的研究;出现了人类动力学等新兴研究方向,也使得重新审视

和精细化研究空间交互现象、交通出行预测等众多经典问题成为可能。

然而,人类移动性研究也存在诸多挑战。首先,数据的稀疏性需要引起高度重视。大数据与数据的稀疏性并不矛盾,例如,尽管我们可以获取签到位置总量巨大的社交网站签到数据,然而具体到某个个体,其签到轨迹十分稀疏,这对数据的精细化分析提出了挑战;其次,由于获取的数据只能反映部分人群的移动行为特征(例如社交网站签到数据一般只覆盖年轻人),因此利用各类数据获取的研究成果是否存在偏斜、是否具有代表性也常被质疑;最后,面对海量的异构数据,如何有针对性地发展与利用数据挖掘方法,如何进行深度学习、机器学习,从而能够从各种轨迹与出行活动数据中提取隐含的知识,是人类移动性研究的永恒挑战。

我们认为,未来人类移动性的研究将侧重以下4个方面:

(1) 多学科研究方法的交叉与融合

地理学、物理学、计算机科学和复杂网络科学等领域的学者从不同的角度出发探索人类移动模式及其内在机理。虽然研究的问题因学科性质不同而各有侧重,但研究方法更趋向于从大数据的应用和分析着手。因此,学科间研究方法的交叉和融合将成为未来发展的趋势,以推动人类移动性在理论方法和应用技术层面的不断创新。

(2) 多源异构数据的管理与计算

时空精细化程度越来越高的移动目标轨迹与其他实时数据具有多源异构性和动态性,对数据的存储、管理、处理提出了挑战。人类移动性研究一方面需要借助大规模数据集的并行处理技术和并行计算环境来提升数据处理效率、降低计算成本,另一方面还需尽可能融合多种类型数据,使得人类移动性分析更加全面。

(3) 多源异构数据的挖掘与分析

在传统分析与建模方法的基础上,面对前所未有的多源异构数据,还需要发展更多新方法。在大数据分析中,半监督学习、集成学习、迁移学习和概率图模型等技术尤为重要。此外,目前已到了需要重新思考“大数据+简单模型”的时候,而学术界和工业界十分推崇的深度学习(Deep learning)也许能帮助我们大数据中发掘出更多有价值的信息。

(4) 注重与实际问题的结合

人类移动性的相关研究应当注重与实际问题

相结合。例如,人类移动性与疾病传播学的结合研究将有助于深入理解这些疾病的传播机制,进而提出有效的预防和控制措施;人类移动性与交通工程学的结合研究则有助于合理规划交通设施,预防和控制交通拥堵。因此,人类移动性的研究一方面应当结合实际问题的领域背景知识,另一方面应当充分利用相关背景数据,如此,既有利于解决实际问题,又有利于深入理解人类的移动行为特征。

参考文献:

- [1] 徐赞新,王钺,司洪波,等.基于随机矩阵理论的城市人群移动行为分析[J].物理学报,2011,60(4):46-52.
- [2] 刘瑜,肖昱,高松,等.基于位置感知设备的人类移动研究综述[J].地理与地理信息科学,2011,27(4):8-13.
- [3] 丁益民,杨昌平.考虑人类流动行为的动态复杂网络研究[J].物理学报,2012,61(23):551-556.
- [4] Rhee I, Shin M, Hong S, *et al.* On the levy-walk nature of human mobility[J]. IEEE/ACM Transactions on Networking, 2011,19(3):630-643.
- [5] Zheng Y, Xie X, Ma W Y. GeoLife: A collaborative social networking service among user, location and trajectory [J]. IEEE Data(base) Engineering Bulletin, 2010,33(2):32-39.
- [6] 郑宇,谢幸.基于用户轨迹挖掘的智能位置服务[J].中国计算机学会通讯,2010,6(6):23-30.
- [7] 申悦,柴彦威.基于GPS数据的北京市郊区巨型社区居民日常活动空间[J].地理学报,2013,68(4):506-516.
- [8] Liu Y, Kang C, Gao S, *et al.* Understanding intra-urban trip patterns from taxi trajectory data[J]. Journal of Geographical Systems, 2012,14(4):463-483.
- [9] Veloso M, Phithakkitnukoon S, Bento C. Sensing urban mobility with taxi flow[C].//Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks. ACM, 2011:41-44.
- [10] Yuan J, Zheng Y, Zhang C, *et al.* T-drive: driving directions based on taxi trajectories[C].//Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, 2010:99-108.
- [11] 郑宇.城市计算与大数据[J].中国计算机学会通讯,2013,9(8):6-16
- [12] Gonzalez M C, Hidalgo C A, Barabasi A L. Understanding individual human mobility patterns[J]. Nature, 2008, 453(7196):779-782.
- [13] Csáji B C, Browet A, Traag V A, *et al.* Exploring the mobility of mobile phone users[J]. Physica A: Statistical Mechanics and its Applications, 2013,392(6):1459-1473.
- [14] Noulas A, Scellato S, Mascolo C, *et al.* An empirical study of geographic user activity patterns in Foursquare

- [J]. ICWSM, 2011,11:70-573.
- [15] Liu Y, Sui Z, Kang C, *et al.* Uncovering patterns of inter-urban trip and spatial interaction from social media check-In data[J]. PloS ONE, 2014,9(1):e86026.
- [16] 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.
- [17] 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.
- [18] Froehlich J, Neumann J, Oliver N. Sensing and predicting the pulse of the city through shared bicycling[M]. International Joint Conferences on Artificial Intelligence. 2009: 1420-1426.
- [19] 汪秉宏,周涛,周昌松.人类行为,复杂网络及信息挖掘的统计物理研究[J].上海理工大学学报,2012,34(2):103-117.
- [19] 周涛,韩筱璞,闫小勇,等.人类行为时空特性的统计力学[J].电子科技大学学报,2013,42(4):481-540.
- [20] Song C, Koren T, Wang P, *et al.* Modelling the scaling properties of human mobility[J]. Nature Physics, 2010,6(10):818-823.
- [21] Han X P, Hao Q, Wang B H, *et al.* Origin of the scaling law in human mobility: Hierarchy of traffic systems[J]. Physical Review E, 2011,83(3):036117.
- [22] Hu Y, Zhang J, Huan D, *et al.* Toward a general understanding of the scaling laws in human and animal mobility [J]. Europhysics Letters, 2011,96(3):38006.
- [23] Han X P, Wang B H. Impacts of distance and memory in the emergence of scaling mobility pattern of human[J]. Physics Procedia, 2010,3(5):1907-1911.
- [24] Di Lorenzo G, Calabrese F. Identifying human spatio-temporal activity patterns from mobile-phone traces[M]. Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on. IEEE, 2011:1069-1074.
- [25] Duan Y, Lu F. Structural robustness of city road networks based on community[J]. Computers, Environment and Urban Systems, 2013,41:75-87.
- [26] Kumar R, Novak J, Tomkins A. Structure and evolution of online social networks[M]. In: Yu Philip S., Han Jia-wei, Faloutsos Christos. Link mining: models, algorithms, and applications. New York: Springer, 2010:337-357.
- [27] Gao S, Liu Y, Wang Y, *et al.* Discovering spatial interaction communities from mobile phone data[J]. Transactions in GIS, 2013,17(3):463-481.
- [28] De Vries J J, Nijkamp P, Rietveld P. Alonso's theory of movements: Developments in spatial interaction modeling [J]. Journal of Geographical Systems, 2001,3(3):233-256.
- [29] Noulas A, Scellato S, Lambiotte R, *et al.* A tale of many cities: universal patterns in human urban mobility[J]. PloS ONE, 2012,7(5):e37027.
- [30] Viswanathan G M. Levy Flight search patterns of wandering albatrosses[J]. Nature, 1996,381:413-415.
- [31] Brockmann D, Hufnagel L, Geisel T. The scaling laws of human travel[J]. Nature, 2006, 439(7075):462-465.
- [32] Liang X, Zheng X, Lv W, *et al.* The scaling of human mobility by taxis is exponential[J]. Physica A: Statistical Mechanics and its Applications, 2012,391(5):2135-2144.
- [33] Jiang B, Jia T. Exploring human mobility patterns based on location information of US flights[J]. arXiv preprint arXiv:1104.4578, 2011.
- [34] 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.
- [35] 王明生,黄琳,闫小勇.探索城市公交客流移动模式[J].电子科技大学学报,2012,41(1):2-7.
- [36] 闫小勇.人类个体出行行为的统计实证[J].电子科技大学学报,2011,40(2):168-173.
- [37] Phithakkitnukoon S, Horanont T, Di Lorenzo G, *et al.* Activity-aware map: Identifying human daily activity pattern using mobile phone data[M]. In: Albert Ali Salah, Theo Gevers, Nicu Sebe, *et al.* Human Behavior Understanding. Berlin Heidelberg: Springer, 2010:14-25.
- [38] Farrahi K, Gatica-Perez D. Discovering human routines from cell phone data with topic models[C].//Wearable Computers, 2008. ISWC 2008. 12th IEEE International Symposium on. IEEE, 2008:29-32.
- [39] Huang L, Li Q, Yue Y. Activity identification from GPS trajectories using spatial temporal POIs' attractiveness [C].//Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks. ACM, 2010:27-30.
- [40] Eagle N, Pentland A S. Eigenbehaviors: Identifying structure in routine[J]. Behavioral Ecology and Sociobiology, 2009,63(7):1057-1066.
- [41] Song C, Qu Z, Blumm N, *et al.* Limits of predictability in human mobility[J]. Science, 2010,327(5968):1018-1021.
- [42] Ashbrook D, Starner T. Using GPS to learn significant locations and predict movement across multiple users[J]. Personal and Ubiquitous Computing, 2003,7(5):275-286.
- [43] Calabrese F, Di Lorenzo G, Ratti C. Human mobility prediction based on individual and collective geographical preferences[C].//Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on. IEEE, 2010:312-317.
- [44] Sadilek A, Kautz H, Bigham J P. Finding your friends and following them to where you are[C].//Proceedings of the fifth ACM international conference on Web search and data mining. ACM, 2012:723-732.
- [45] Baraglia R, Muntean C I, Nardini F M, *et al.* LearNext: learning to predict tourists movements[C].//Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. ACM, 2013: 751-756.

- [46] Zipf G K. The P1 P2/D hypothesis: On the intercity movement of persons[J]. American sociological review, 1946: 677-686.
- [47] Simini F, González M C, Maritan A, *et al.* A universal model for mobility and migration patterns[J]. Nature, 2012,484(7392):96-100.
- [48] Lenormand M, Huet S, Gargiulo F, *et al.* A universal model of commuting networks[J]. PloS ONE, 2012,7(10): e45985.
- [49] Kang C, Sobolevsky S, Liu Y, *et al.* Exploring human movements in Singapore: A comparative analysis based on mobile phone and taxicab usages[C]//Proceedings of the 2nd ACM SIGKDD International Workshop on Urban Computing. ACM, 2013.
- [50] Ratti C, Williams S, Frenchman D, *et al.* Mobile landscapes: using location data from cell phones for urban analysis[J]. Environment and Planning B Planning and Design, 2006,33(5):727-748.
- [51] Andrienko G, Andrienko N, Bak P, *et al.* A framework for using self-organising maps to analyse spatio-temporal patterns, exemplified by analysis of mobile phone usage[J]. Journal of Location Based Services, 2010,4(3-4):200-221.
- [52] Pei T, Sobolevsky S, Ratti C, *et al.* A New insight into land use classification based on aggregated mobile phone data[J]. arXiv preprint arXiv:1310.6129, 2013.
- [53] Toole J L, Ulm M, González M C, *et al.* Inferring land use from mobile phone activity[C]//Proceedings of the ACM SIGKDD International Workshop on Urban Computing. ACM, 2012:1-8.
- [54] Pan G, Qi G, Wu Z, *et al.* Land-use classification using taxi GPS traces[J]. Intelligent Transportation Systems, IEEE Transactions on, 2013,14(1):113-123.
- [55] 陆锋,张恒才.大数据与广义GIS[J].武汉大学学报(信息科学版),2014,39(6):645-654.

Research on Human Mobility in Big Data Era

LU Feng*, LIU Kang and CHEN Jie

(State Key Lab of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China)

Abstract: Human mobility has received much attention in many research fields such as geography, sociology, physics, epidemiology, urban planning and management in recent years. On the one hand, trajectory datasets characterized by a large scale, long time series and fine spatial-temporal granularity become more and more available with rapid development of mobile positioning, wireless communication and mobile internet technologies. On the other hand, quantitative studies of human mobility are strongly supported by interdisciplinary research among geographic information science, statistical physics, complex networks and computer science. In this paper, firstly, data sources and methods currently used in human mobility studies are systematically summarized. Then, the research is comprehended and divided into two main streams, namely people oriented and geographical space oriented. The people oriented research focuses on exploring statistical laws of human mobility, establishing models to explain the appropriate kinetic mechanism, as well as analyzing human activity patterns and predicting human travel and activities. The geographical space oriented research focuses on exploring the process of human activities in geographical space and investigating the interactions between human movement and geographical space. Followed by a detailed review of recent progress around these two streams of research, some research challenges are proposed, especially on data sparsity, data skew processing and heterogeneous data mining, indicating that more integration of multidiscipline are required in human mobility studies in the future.

Key words: human mobility; big data; data mining; statistical physics; complex network

*Corresponding author: LU Feng, E-mail: luf@lreis.ac.cn