

大数据驱动的人类移动模式和模型研究

刘 瑜¹ 康朝贵¹ 王法辉²

1 北京大学遥感与地理信息系统研究所,北京,100871

2 路易斯安那州立大学地理与人类学系,巴吞鲁日,路易斯安那州,70803

摘 要:大数据时代的到来使得基于个体粒度的海量时空轨迹获取人类移动模式成为可能。来自不同领域的学者基于手机通话数据、公交卡刷卡记录、社交网站签到数据、出租车轨迹、银行刷卡记录等进行了人类移动模式的研究,这些研究丰富了地理信息系统的时空分析方法,为从个体角度审视人与地理环境之间的交互关系提供了一个新视角,并可以应用于交通、公共卫生等领域。总结了基于大数据的人类移动模式研究流程,归纳了人类移动模式的基本度量方法,探讨了解释所观测移动模式的模型构建方法,指出了地理环境对于移动模型建立的影响。

关键词:大数据;人类移动模式;移动模型;地理环境影响

中图法分类号:P208

文献标志码:A

随着带有定位功能的移动计算设备等的广泛应用,大量具有个体粒度的时空标记的大数据(big data)的产生(包括手机通话数据、公交卡刷卡记录、社交网站签到数据、出租车轨迹、银行卡刷卡记录等)为长时间、高精度、高效地跟踪个体空间移动提供可能^[1]。同时,地理信息系统与科学的不断发展使得时间地理学从数据采集、存储、分析到可视化等整个过程都获得了极大的便利^[2-5],其强大的可视化和空间分析功能增强了个体时空轨迹的直观显示及其隐含模式的识别能力^[6-7]。实践表明,个体移动模式识别与分析能够为城市规划与管理^[8-9]、交通监控与预测^[10-11]、信息与疾病传播^[12-13]、旅游监测与分析^[14-15]等众多领域的研究提供工作基础与方法指导。基于大数据研究个体或群体行为,发现活动中蕴含的空间认知规律及空间行为和交互模式,建立以人为本的地理信息服务,进而支持个体或群体时空行为决策,已成为地理信息科学研究的前沿问题。此类研究也将为联系“计量”和“行为”两个地理学思想流派以及重新审视人上-地关系提供一个全新的视角,从而推动理论地理学的进一步发展。

1 人类移动模式研究流程

图 1 展示了基于地理空间大数据进行人类移

动模式研究的基本过程。基于原始数据,可以进行数据预处理并提取海量的轨迹,其中每条轨迹对应于一个个体。对于手机和签到数据而言,一个用户通常通过电话号码及用户 ID 对应于唯一的一条轨迹。对于出租车和公交卡数据,一条轨迹对应于个体的一次出行,在整个数据集中,可以有多条轨迹对应于同一个人。一条轨迹可以表示为 $\langle x_j, y_j, t_j \rangle, j=1, \dots, n$, 其中 n 为停留点的数目; (x_j, y_j, t_j) 表示轨迹上一个点的时空位置。在基于签到数据获取的轨迹中,通常包含每个停留点对应的活动信息,如就餐、购物等,从而可以更好理解人的行为规律。

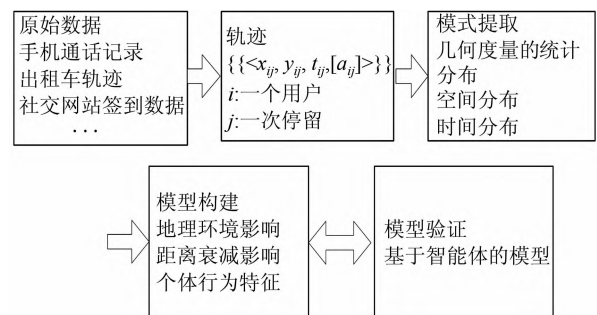


图 1 大数据驱动的人类移动模式研究流程

Fig. 1 Flowchart of Big Data-Driven Human Mobility Studies

在基于移动轨迹的模式发现中,现有研究主

要关注其几何度量的统计分布特征以及移动的时空分布特征。来自于复杂系统的学者多关注步长分布是否为幂律或具有重尾特征。目前针对分布的拟合多基于频率直方图进行, Clauset 等给出了一个更为严格的检验方法^[16], 然而由于人的移动行为的复杂性, 通过该检验并不容易。

在地理学研究中, 移动的时空分布特征可以揭示区域或城市的空间结构。其中, 移动的空间分布反映了城市土地利用的强度, 而时间分布则受到土地利用类型的影响^[17-20]。此外, 针对轨迹数据, 也有大量学者从数据挖掘^[21]、空间分析^[22]、可视化^[23]、时空数据库^[24]等方面开展了研究。

为了解释观测到的人类移动模式, 需要构造合适的模型加以解释。在模型中, 引入不同的因素, 包括地理环境的影响、距离衰减效应以及个体的行为特点等。为了验证模型, 多借鉴基于智能的模型, 从个体角度出发, 建立个体移动的概率模型, 通过蒙特卡洛方法引入随机性, 从而模拟生成大量虚拟的轨迹, 然后比较这些模拟得到的轨迹和观测到的轨迹的统计特征。如果比较的结果存在较大差异, 可以调整模型参数。而如果得到的两个分布相近或一致, 通常认为所建立的模型较好地解释了数据集所反映的人类移动模式。然而值得指出的是, 在比较时, 多采用聚集层面的统计特征, 而不同的个体层面的移动模型可以得到相同聚集层面的统计特征。因此, 一个更为保守的陈述是, 如果模拟得到的结果与实际一致, 那么不能“否定”所构造的移动模型^[25]。

在大数据驱动的人类移动模式研究中, 一个重要的相关联的问题是数据的代表性。目前, 研究人类移动模式所采用的各种数据都不是通过设计一个合理的采样原则对人群进行采样获得的。因此, 这些数据存在一定的偏差, 如手机数据只是反映了手机用户的空间移动特征, 而出租车轨迹数据只是反映了基于出租车的出行特征, 而不能发现乘坐公交车、自驾车等出行的模式。这就使得基于一种数

据源可能只是发现了人类移动模式的一个方面。如果能够得到相同时空范围的不同数据集, 就可以进行相互对比和验证。目前, 已有学者在该方向开展了初步研究, 如 Kang 等针对新加坡出租车和手机数据获取的移动模式进行了比较^[26]。

2 人类移动模式的度量

每个个体的移动模式可以表示为随机游走 (random walk) 模型, 根据随机游走模型, 个体每次移动的角度和步长 (displacement) 符合特定的概率分布。通常在建模过程中, 认为移动在不同方向上的分布是均匀的, 即各向同性, 而步长可以为固定值或者为正态或幂律分布。当步长分布为正态时, 为布朗运动; 而当步长为幂律分布, 并且幂指数在 1~3 时, 移动为列维 (Lévy) 飞行模型。图 2(a) 和 2(b) 展示了两个步长分布为幂律的列维飞行轨迹和一个步长分布为正态的布朗运动, 图 2(c) 和 2(d) 绘制了回转半径 (radius of gyration, ROG, ROG 大致反映了个体活动空间的大小) 以及离出发点的距离与移动步数的关系, 可以看出, 列维飞行模型对应较快的扩散速率。许多研究考察个体的轨迹发现, 在短距离移动之后, 总会伴随着一次大范围的转移, 这符合动物觅食的一般特点, 并且效率较高。因此, 目前针对移动模式的研究主要关注移动步长的统计分布特征。

由于在实践中难以获得一个个体长时间的多次移动轨迹, 大多数研究主要通过海量个体的轨迹数据的统计特征开展。表 1 总结了目前基于大数据的人类移动模式研究及发现的步长统计特征。

除了移动步长分布外, 还有一些度量可以用于分析人类移动模式, 如角度的分布。在一般的随机游走模型中, 角度的分布是均匀的, 即个体的每次移动没有方向的偏好性。在实际的移动模式中, 角度分布会出现各向异性的特点^[30,34]。如果

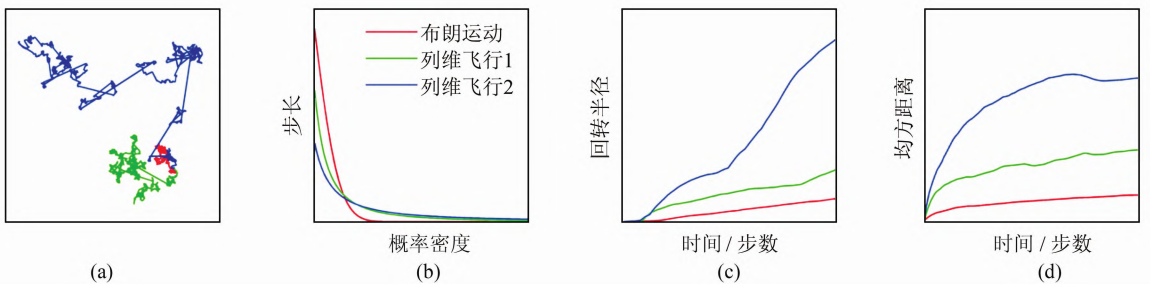


图 2 不同的移动步长分布比较

Fig. 2 Comparison Between Movement Models with Different Displacement Distributions

根据轨迹的形状确定一次移动,则角度分布还会受到道路形状的影响^[31]。对于一个个体移动轨迹,还可以计算更多的指标,如回转半径(ROG)和熵等^[35]。其中熵广泛应用于时间地理学研究中^[36-37],在文献[28]中,用 ROG 来度量人群活动的异质性,在文献[38]中,熵用于表征个体在移动中对不同地点的重访概率分布。关于更多的移动轨迹度量可参见文献[39]。

表 1 基于大数据的人类移动模式研究及发现的步长分布特征

Tab. 1 Various Displacement Distributions Obtained from Different Data Sets

研究	轨迹数据类型	步长(d)分布
Brockmann, et al. ^[27]	钱币数据	$d^{-\beta}, \beta=1.59 \pm 0.02$
González, et al. ^[28]	手机数据	$(d+d_0)^{-\beta} \exp(-ad),$ $\beta=1.75 \pm 0.15$
Kang, et al. ^[29]	手机数据	$\exp(-ad)$
Liu, et al. ^[30]	出租车数据	$(d+d_0)^{-\beta} \exp(-ad),$ $\beta=1.2 \pm 0.15$
Jiang, et al. ^[31]	出租车数据	$d^{-\beta}, \beta=2.5$
Liang, et al. ^[32]	出租车数据	$\exp(-ad)$
Noulas, et al. ^[33]	签到数据	$(d+d_0)^{-\beta}, \beta=1.5$
Liu, et al. ^[25]	签到数据	$\exp(-ad)$

3 移动模型

许多学者试图建立模型以解释观察到的人类移动模式。通常解释移动模式需要考虑的因素包括三个方面,即地理环境、距离衰减以及个体的空间行为特征。地理环境因素决定了个体移动潜在到访点的空间分布,通常该分布与人口密度分布正相关。距离衰减表示由于长距离移动的成本高,使得产生长距离移动的概率更低,目前基于海量个体轨迹所发现的移动步长分布都展示了这个特征。而个体的空间行为特征则反映了人们移动中一些个性化的规律,目前得到较多关注的是个体轨迹中的重访点,这是人类移动和动物移动存在较大差异的方面,即存在家和工地等频繁重访的地点,从而展示出较高的可预测性^[38]。基于此, Song 等建立了一个考虑重访概率的移动模型,即个体在每次移动时,以概率 ρ 访问一个新地点,而以概率 $1-\rho$ 重访已经到达的地点,从而解释所观察到的移动模式^[40]。

地理学方面的研究主要关注地理环境和距离衰减对于人类移动模式的影响,其中重力模型集中反映了这两个方面的作用。重力模型可以表达为:

$$I_{ij} = \frac{P_i P_j}{f(d_{ij})}$$

其中, I_{ij} 表示区域 i, j 之间的交互; P_i, P_j 为两个区域的规模; d_{ij} 为两个区域之间的距离; $f(d_{ij})$ 为距离衰减函数,通常可以有幂律以及指数等形式。在聚集层面, I_{ij} 表示两个区域的交互量,如客流量、通讯量等;而在个体层面, I_{ij} 表示可以在两个区域间产生一次出行的概率。Liu 等采用该模型较好地解释了基于上海市出租车数据所发现的出行步长分布特征,并且指出空间行为步长分布所展示的距离衰减是由行为本身的距离衰减特性以及潜在轨迹停留点的分布共同作用、相互加强造成的,后者可以用人口分布来近似^[30]。该模型后来被研究者采用其他不同的数据所验证^[29,41]。

也有一些研究指出重力模型的不足,一些学者直接采用人口数作为区域规模来检查重力模型并得到了负面的结果^[42-43]。Simini 等指出了重力模型的不足,并且建立辐射模型(radiation model)以解释人类移动和迁移模式^[43]。但是在这些对重力模型的否定中,采用人口作为地理区域规模并不合适,因为人的出行驱动力并不严格与人口数正相关。一个应用重力模型的较好途径是根据地理区域间的交互量以及距离拟合重力模型,并估计产生实际出行流量的理论规模以及距离衰减函数^[44-45]。

在辐射模型中,关键是没有采用绝对距离而采用相对距离,这类似介入机会模型(intervening opportunities model),它使得在决定一次出行时候选目的点到当前点的距离排名而不是绝对距离起到了更重要的作用。文献[33,46]对距离采用了类似的处理方式,以解决地理空间分布异质性带来的影响。在建模中,尽管对于个体而言,采用不同的距离度量会造成不同的轨迹模式,但是在聚集层面上,这两者差异并不大。根据图 3(a)所示的随机分布的点,可以得到第 i 个候选目的点的平均距离与 i 正相关(图 3(b))。即对于所有个体而言,相对距离和绝对距离基本线性相关,这使得即使个体移动遵循基于距离位序的模型,但在聚集层面上依然可以较好地符合重力模型。

值得指出的是,目前所观测到的人类移动模式都是基于包含大量个体的移动轨迹数据得到的聚集层面的模式。不应忽视的是,由于个体差异的存在^[20,28],即使在宏观聚集层面得到步长的分布特征,也不能简单推测每个个体的移动简单受该函数的影响。如在城市尺度的移动中,每个人的通勤活动占据了较大比重。对于每个个体而

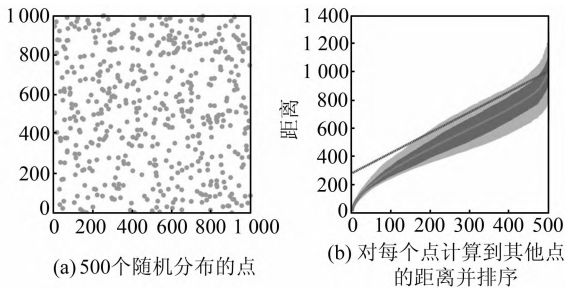


图 3 绝对距离和相对距离的关系

Fig. 3 Comparison Absolute Distance with Relative Distance

言,家和工作地点的距离分布频率较高,未必表现为步长的幂律分布。而在群体层面,由于长距离通勤的个体比例相对较低,这使得群体的移动步长会表现出距离衰减特征^[25]。González 等^[28]对此进行了探讨,指出所观察到的人类移动模式是群体异质性与个体移动模式卷积造成的结果。

4 地理环境影响

地理因素对于人类移动模式的影响主要体现在两个方面,即距离衰减效应和地理环境的空间分布特征。距离衰减效应反映了距离对于出行的影响,而地理环境决定了潜在出行 OD(origin destination)点(可粗略地对应于轨迹中的停留点)的分布,进而影响了观测到的移动模式。前者可以通过距离衰减函数来表征,但是后者则难以用数学函数来表示,因为不同区域的地理分布格局存在较大的差异(尽管存在一定的规律性)。此外,在观测到的人类移动模式与地理分布之间建立解析函数关系也较为困难。

对应于三个尺度的移动,地理分布的表达也各不相同。在亚城市尺度,如公园,由于考察个体移动的时间跨度较短,通常将其假定为均质的。在城市尺度的移动中,人们的出行受到城市用地结构的影响。人口密度是理解城市结构的窗口,并决定了出行 OD 点的空间分布模式。研究表明,城市内部人口密度分布的规律性较为明显,随离市中心距离的增加而递减,并可以用一个负指数函数加以表达^[47]。除距离衰减效应外,这种单核心的城市人口分布形态进一步降低了长距离出行的概率^[30],使得最终观察到的步长分布的尾部不那么“重”,可以用指数分布或者指数截断的幂律分布来拟合^[29,41]。最后,在城市间尺度,城市体系的人口分布以及不同规模城市的空间分布会影响到人类移动模式。在城市规模的统计分布方面,

Zipf^[48]用美国的城市人口数据归纳出位序规模分布函数 $P_r = P_1 r^{-q}$, 式中, r 是城市从大到小排序的位序; P_r 是第 r 大城市的人口规模; P_1 和 q 是两个常数(P_1 可以解释为首位城市规模的理论值, q 表示城市人口规模随着位序上升而下降的坡度)。此后,许多学者也针对不同区域的城市体系展开了研究,并且发现了类似的统计规律^[49-50]。城市的位序规模分布在聚集层面决定了城市交互强度的分布,即少数大规模城市间的交互在交互总量中占据较大比例,而在个体层面,提高了大城市间距离对应的移动步长的概率。城市体系中不同规模的城市空间分布同样影响了观测到的移动模式。理想情况下,城市体系中不同规模的城市展示出层次结构,并存在有规律的空间分布模式。最近, Han 等探讨了城市体系中的层次结构对于人类移动模式的影响^[51]。该研究指出,人们在低层次城市之间的移动通常要路由高层次的城市,从而导致了步长的幂率分布特征,然而在其研究中并没有考虑不同层次城市人口数分布的影响。

5 结 语

人类移动模式是一个方兴未艾的领域,目前已经取得很多进展,但是如何更好地和地理空间相结合,并且将发现的时空模式应用于相关领域依然有许多问题需要研究和解决。

1) 人类移动性的尺度问题。尽管人的活动轨迹在时空间上是连续的,但个体移动模式却与分析的时空间尺度密切相关。考虑到人的活动的具体特点,可以将人的活动范围分为三个尺度,即城市间尺度、城市尺度和亚城市尺度。在这三个尺度上,人类的移动具有不同的特征,这些因素要在建模中加以区分考虑。

2) 人类移动模式数据的代表性问题。单一类型的大数据反映了特定人群的移动/活动特征,存在一定的采样偏差。因此,如何针对一个地理区域融合多源大数据,以更为全面地捕获移动规律,建立更为合理的解释模型,是目前的一个重要挑战。

3) 人类移动模型。不同的移动模型关注人类时空行为中的不同方面,以解释所观察到的移动模式。从地理学的角度,有两个问题需要关注: ①个体移动在空间上是不均匀分布的,这反映了地理环境的空间异质性。这种空间异质性难以用明确的数学函数加以表达,这给移动模型的解析形式的建立带来了困难。②活动是个体移动的根

本驱动,不同目的的出行往往具有不同的时空分布特征。然而,目前所获取的多类大数据都没有显式地表达活动信息,如何从轨迹中推断活动,从而区分不同目的的出行,有助于更好地理解人类移动模式。

4) 人类移动模式的应用。目前,尽管已经有较多文献介绍了人类移动模式的应用,但是在实践中尚未见到成熟的应用系统。如何将量化的人类移动模式作为输入参数,在交通、公共卫生等系统中起到优化和决策支持作用,需要更为深入的研究。

参 考 文 献

- [1] Lu Y, Liu Y. Pervasive Location Acquisition Technologies: Opportunities and Challenges for Geospatial Studies[J]. *Computers, Environment and Urban Systems*, 2012, 36(2): 105-108
- [2] Kwan M P. Interactive Geovisualization of Activity-travel Patterns Using Three-dimensional Geographical Information Systems: A Methodological Exploration with a Large Data Set[J]. *Transportation Research Part C*, 2000, 8: 1-6
- [3] Kwan M P, Casas I. GABRIEL: GIS Activity-based Travel Simulator, Activity Scheduling in the Presence of Real-time Information[J]. *Geoinformatica*, 2006, 10(4): 469-493
- [4] Yu H, Shaw S L. Exploring Potential Human Activities in Physical and Virtual Spaces: A Spatio-temporal GIS Approach[J]. *International Journal of Geographical Information Science*, 2008, 22(4): 409-430
- [5] Shaw S L, Yu H. A GIS-based Time-geography Approach of Studying Individual Activities and Interactions in a Hybrid Physical-Virtual Space[J]. *Journal of Transport Geography*, 2008, 17(2): 141-149
- [6] Kraak M J, Koussoulakou A. A Visualization Environment for the Space-Time-Cube [C]. *Developments in Spatial Data Handling*. Berlin Heidelberg: Springer, 2005: 189-200
- [7] Ren F, Kwan M P. Geovisualization of Human Hybrid Activity-Travel Patterns[J]. *Transactions in GIS*, 2007, 11(5): 721-744
- [8] Ahas R, Mark Ü. Location Based Services—New Challenges for Planning and Public Administration? [J]. *Futures*, 2005, 37: 547-561
- [9] Ratti C, Frenchman D, Pulselli R M, et al. Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis[J]. *Environment and Planning B: Planning and Design*, 2006, 33: 727-748
- [10] Wang P, Hunter T, Bayen A M, et al. Understanding Road Usage Patterns in Urban Areas[J]. *Scientific Reports*, 2012(2): 1 001
- [11] Gao S, Wang Y L, Gao Y, et al. Understanding Urban Traffic-Flow Characteristics: A Rethinking of Betweenness Centrality [J]. *Environment and Planning B: Planning and Design*, 2013, 40(1): 135-153
- [12] Wang P, González M C, Hidalgo C A, et al. Understanding the Spreading Patterns of Mobile Phone Viruses[J]. *Science*, 2009, 324: 1 071-1 076
- [13] Bian L. Spatial Approaches to Modeling Dispersion of Communicable Diseases—A Review [J]. *Transactions in GIS*, 2013, 17(1): 1-17
- [14] Ahas R, Aasa A, Mark Ü, et al. Seasonal Tourism Spaces in Estonia: Case Study with Mobile Positioning Data[J]. *Tourism Management*, 2007, 28(3): 898-910
- [15] Tiru M, Kuusik A, Lamp M L, et al. LBS in Marketing and Tourism Management: Measuring Destination Loyalty with Mobile Positioning Data [J]. *Journal of Location Based Services*, 2010, 4(2): 120-140
- [16] Clauset A, Shalizi C R, Newman M E J. Power-law Distributions in Empirical Data[J]. *SIAM Review*, 2009, 51(4): 661-703
- [17] Chen C, Chen J, Barry J. Diurnal Pattern of Transit Ridership: A Case Study of the New York City Subway System[J]. *Journal of Transport Geography*, 2009, 17(3): 176-186
- [18] Gong Y, Liu Y, Lin Y, et al. Exploring Spatiotemporal Characteristics of Intra-urban Trips Using Metro Smartcard Records[C]. *The 20th International Conference on Geoinformatics*, Hong Kong, 2012
- [19] Liu Y, Wang F, Xiao Y, et al. Urban Land Uses and Traffic ‘Source-sink Areas’: Evidence from GPS-enabled Taxi Data in Shanghai[J]. *Landscape and Urban Planning*, 2012, 106: 73-87
- [20] Yuan Y, Raubal M. Extracting Dynamic Urban Mobility Patterns from Mobile Phone Data[J]. *Lecture Notes in Computer Science*, 2012, 7 478: 354-367
- [21] Lee J G, Han J, Li X, et al. TraClass: Trajectory Classification Using Hierarchical Region-Based and Trajectory-Based Clustering[C]. *Int Conf on Very Large Data Base (VLDB’08)*, Auckland, New Zealand, 2008
- [22] Long J A, Nelson T A. A Review of Quantitative Methods for Movement Data [J]. *International Journal of Geographical Information Science*,

- 2013, 27(2): 292-318
- [23] Demsšr U, Verrantaus K. Space-time Density of Trajectories: Exploring Spatio-temporal Patterns in Movement Data[J]. *International Journal of Geographical Information Science*, 2010, 24 (10): 1 527-1 542
- [24] Gütting R H, Böhlen M H, Erwig M, et al. A Foundation for Representing and Querying Moving Objects[J]. *ACM Transactions on Database Systems*, 2000, 25(1): 1-42
- [25] 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
- [26] Kang C, Sobolevsky S, Liu Y, et al. Exploring Human Movements in Singapore: A Comparative Analysis Based on Mobile Phone and Taxicab Usages [C]. The 2nd International Workshop on Urban Computing (UrbComp'13), Chicago, USA, 2013
- [27] Brockmann D, Hufnagel L, Geisel T, The Scaling Laws of Human Travel[J]. *Nature*, 2006, 439: 462-465
- [28] González M C, Hidalgo C A, Barabási A L. Understanding Individual Human Mobility Patterns [J]. *Nature*, 2008, 453: 779-782
- [29] Kang C, Ma X, Tong D, et al. Intra-urban Human Mobility Patterns: An Urban Morphology Perspective[J]. *Physica A: Statistical Mechanics and Its Applications*, 2012, 391(4): 1 702-1 717
- [30] 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
- [31] Jiang B, Yin J, Zhao S. Characterizing the Human Mobility Pattern in a Large Street Network[J]. *Physical Review E*, 2009, 80: 021136
- [32] Liang X, Zheng X, Lü W, et al. The Scaling of Human Mobility by Taxi is Exponential[J]. *Physica A: Statistical Mechanics and Its Applications*, 2012, 391(5): 2 135-2 144
- [33] 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
- [34] Rhee I, Shin M, Hong S, et al. On the Lévy Walk Nature of Human Mobility[J]. *IEEE/ACM Transactions on Networking*, 2011, 19(3): 630-643
- [35] Yuan Y, Raubal M, Liu Y. Correlating Mobile Phone Usage and Travel Behavior - A Case Study of Harbin, China[J]. *Computers, Environment and Urban Systems*, 2012, 36(2): 118-130
- [36] Kwan M P. Gender Differences in Space-time Constraints[J]. *Area*, 2000, 32(2): 145-156
- [37] Wong D, Shaw S L. Measuring Segregation: An Activity Space Approach [J]. *Journal of Geographical Systems*, 2011, 13(2): 127-145
- [38] Song C, Qu Z, Blumm N, et al. Limits of Predictability in Human Mobility[J]. *Science*, 2010, 327: 1 018-1 021
- [39] Csájia 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): 1 459-1 473
- [40] Song C, Koren T, Wang P, et al. Modelling the Scaling Properties of Human Mobility[J]. *Nature Physics*, 2010, 6: 818-823
- [41] Liang X, Zhao J, Dong L, et al. Unraveling the Origin of Exponential Law in Intra-urban Human Mobility[J]. *Scientific Reports*, 2013, 3: 2 983
- [42] Thiemann C, Theis F, Grady D, et al. The Structure of Borders in a Small World[J]. *PLoS ONE*, 2010, 5: e15422
- [43] Simini F, González M C, Maritan A, et al. A Universal Model for Mobility and Migration Patterns [J]. *Nature*, 2012, 484: 96-100
- [44] O'Kelly M E, Song W, Shen G. New Estimates of Gravitational Attraction by Linear Programming [J]. *Geographical Analysis*, 1995, 27: 271-285
- [45] Xiao Y, Wang F, Liu Y, et al. Reconstructing Gravitational Attractions of Major Cities in China from Air Passenger Flow Data 2001-2008: A Particle Swarm Optimization Approach[J]. *The Professional Geographer*, 2013, 65: 265-282
- [46] Liben-Nowell D, Novak J, Kumar R, et al. Geographic Routing in Social Networks[J]. *Proceedings of the National Academy of Sciences of the USA*, 2005, 102(33): 11 623-11 628
- [47] Clark C. Urban Population Densities[J]. *Journal of the Royal Statistical Society*, 1951, 114: 490-494
- [48] Zipf G K. Human Behavior and the Principle of Least Effort[M]. Cambridge, MA: Addison-Wel-sey, 1949
- [49] Chen Y, Zhou Y. The Rank-size Rule and Fractal Hierarchies of Cities: Mathematical Models and Empirical Analyses[J]. *Environment and Planning B; Planning and Design*, 2003, 30: 799-818
- [50] Batty M. Rank Clocks [J]. *Nature*, 2006, 444: 592-596
- [51] Han X, 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: 036117

Towards Big Data-Driven Human Mobility Patterns and Models

LIU Yu¹ KANG Chaogui¹ WANG Fahui²

1 Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 100871, China

2 Department of Geography and Anthropology, Louisiana State University, Baton Rouge, LA 70803, USA

Abstract: In the big data era, massive volumes of individual-level movements, extracted from various geospatial data, including mobile phone data, public transportation card records, social media check-in data, taxi trajectories, and bank card records, are available for scholars in different fields to study human mobility patterns. These studies enrich spatio-temporal analysis methods in GIS and provide a new perspective to human-environment interactions. Observed human mobility patterns and models can be applied to many applications such as transportation and public health. This paper presents a generic workflow for big-data-driven human mobility analyses and summaries major movement measures. By comparing a number of models used to interpret and reproduce the observed pattern, this paper emphasizes the geographical impact on human mobility patterns.

Key words: big data; human mobility pattern; movement model; geographical impact

First author: LIU Yu, professor. His research interest is geographical information science. E-mail: liuyu@urban.pku.edu.cn

Foundation support: The National Natural Science Foundation of China, Nos. 41271386, 41271385, 41171296.

(上接第 644 页)

Big Data GIS

LI Qingquan^{1,2} LI Deren²

1 Shenzhen Key Laboratory of Spatial Smart Sensing and Services, Shenzhen University, Shenzhen 518060, China

2 State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

Abstract: Big data is changing the world, and also posing challenges for GIS. The volume, velocity, and variety of these data challenge the data management ability of GIS, while the veracity and value issues of big data challenge spatial analysis theory and methods. Thus, as a tool focusing on spatial data management, analysis and visualization, GIS has to make necessary adjustments and changes to meet the big data requirements. This paper discusses the challenges based on the 5V properties of big data, and then, analyzes three characteristics of future GIS in the big data era, which are: ① scalable data management, ② data-driven modeling and data mining, and ③ geo-computational visual analytic.

Key words: big data; GIS; spatial data management; spatial data analysis; visualization

First author: LI Qingquan, PhD, professor, specializes in spatial-temporal data mining, multi-sensor integration, and industry and engineering surveying. E-mail: liqq@szu.edu.cn

Foundation support: The National Natural Science Foundation of China, No. 41371377; Shenzhen Scientific Research and Development Funding Program, Nos. ZDSY20121019111146499, JSGG20121026111056204; Shenzhen Dedicated Funding of Strategic Emerging Industry Development Program, No. JCYJ20121019111128765.