

社会感知视角下的若干人文地理学基本问题再思考

刘瑜

(北京大学遥感与地理信息系统研究所,北京 100871)

摘要:近年来,不同类型大数据在地理研究中得到了越来越多的重视,许多学者基于手机、社交媒体、出租车等数据开展了大量实证研究。社会感知概念刻画了地理空间大数据基于大量人的行为时空模式获取地理环境特征的能力。社会感知手段有助于重新审视地理学研究中的—些基本问题,因而本文选择了空间分布和空间交互这两个基本地理概念以及定性方法和定量方法这两个人文地理基本研究方法展开讨论。大数据从微观个体和宏观群体两个层面同时感知空间分布和空间交互,可以定量分析其中的距离以及尺度效应。进而,由于小样本访谈人群和场所是定性研究的基础,而大数据可以通过定量方法识别特定人群和场所并进行刻画,因此,社会感知手段为集成定性和定量研究方法,构建混合地理学奠定了基础。

关键词:大数据;社会感知;空间分布;空间交互;定性方法;定量方法

DOI: 10.11821/dlxb201604003

1 大数据与社会感知

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(如城市规划、环境保护、交通等)研究的机遇和挑战^[1-4]。目前,关于大数据并没有一个标准的定义^[1]。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)^[3]。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具有时空标记并且可以关联到个体,而一个数据集中包含了大量个体的行为信息。在本文中,地理空间大数据特指这两类数据。目前研究较多的地理空间大数据有手机数据、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,当样本变大,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,地理空间大数据为人们进一步定量地理解社会经济环境提供了一种新的观测手段。最近,Liu提出了“社会感知(social sensing)”概念及研究框架,指出社会感知是指借助于各类海量时空数据研究人类时空行为特征,进而揭示社会经济现象的时空分布、联系及过程的理论和方法^[5]。社会感知数据可从三个方面提取人的时空行为特征:①对地理环境的情感和认知(如基于社交媒体数据中可以获取人们对于—个场所的感受);②在地理空间中的活动和移动(如基于出租

收稿日期:2015-08-24;修订日期:2015-12-24

基金项目:国家自然科学基金项目(41271386, 41428102) [Foundation: National Natural Science Foundation of China, No.41271386, No.41428102]

作者简介:刘瑜(1971-),男,山东诸城人,教授,博士生导师,中国地理学会会员(S110007302M),主要研究方向为地理信息科学。E-mail: liuyu@urban.pku.edu.cn

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 (如基于手机数据可以获取用户之间的通话联系信息)。社会感知的研究框架包括人、地、时三个基本要素。首先, 在“人”的方面, 社会感知数据可以获取人的活动与移动、社交关系、情感与认知等行为模式; 其次, 在“地”的方面, 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; 最后, 从“时”的视角, 可以发现地理过程 (尤其是人文地理过程, 如城市空间结构演化) 的规律和特征。与社会感知手段相比, 传统的遥感技术利用光谱特征对获取地物信息, 但无法有效地感知社会经济环境, 而大数据包含丰富的人群时空行为信息, 形成了对传统遥感数据的有力补充。

在传统的人文地理学研究中, 由于样本量通常较少, 难以直接根据人的行为特征去推断地理环境。而根据社会感知框架, 相关研究主要有如下几个方向: ① 基于活动时间变化特征分析用地功能^[6-8]; ② 基于空间交互发现城市或区域的空间结构^[9-10]; ③ 基于社交媒体语义数据提取特定地理事件与地理环境特征^[11-14]。正如这些研究所展示的, 时空大数据使基于人的时空行为特征感知并理解地理环境特征成为可能^[5]。回顾人文地理学发展的几次转向, 计量革命因为缺少对人的关注而受到批评, 而后期行为主义则受限于样本太小、缺乏定量模型。大数据的出现以及相关的社会感知手段研究, 在一定程度上弥补了上述两个方面的不足, 对于地理学尤其是人文地理学研究具有重要意义。对于地理信息系统与科学而言, 大数据的意义体现在两个方面。首先, 大数据管理和分析需要研究实现新的数据模型和空间分析方法; 其次, 大数据也为实现以人为本的地理信息系统提供了支持^[15]。

地理空间大数据及其提供的社会感知手段, 带来了全新的研究范式, 促进了人文地理学的研究^[16], 也有助于重新审视地理学研究的一些基本问题。由于空间分布和空间交互是地理学研究的两个重要传统^[17-18], 它们共同表征了空间异质性 (spatial heterogeneity), 而定性方法与定量方法是人文地理学研究的两条重要途径, 因此, 本文围绕空间分布和空间交互以及定性方法和定量方法这两组问题, 归纳最新研究进展, 并探讨社会感知支持下的空间分布和空间交互研究, 以及如何从“人”和“地”这两个视角去集成地理学中的定性与定量方法。

2 空间分布和空间交互

空间异质性是地理学的基础概念, 它表现为观测变量的一阶分布 (distribution) 以及场所间二阶交互 (interaction) 的时空变化特征。空间分布和空间交互在现代地理学研究中受到广泛重视, Ullman、Hartshorne、Castells 等著名地理学家均对此问题进行了探讨^[19-21]。Tobler 地理学第一定律正是阐述了分布的空间依赖 (spatial dependence) 及交互所受到的距离影响, 从而在理论上支持相关的空间建模与分析^[22-25]。

2.1 感知空间分布和空间交互

大数据同时提供了对于空间分布和空间交互的感知手段。首先, 人类活动密度的空间差异及时间变化表达了相应地理现象的分布特征, 如图 1a 所示, 利用超过 50 万用户一年内的社交媒体签到位置, 展现了中国城市化程度较高的区域以及主要交通干线的空间分布特征。

除了签到数据, 手机的通话记录以及出租车的上下车点都可以用于量化人群活动的时空分布^[26-27]。在城市尺度上, 由于相同功能地块具有相近的人群活动密度以及日变化特征, 例如, 城市商业中心区的人群活动量呈现出“昼高夜低”的模式, 而在居民区, 其

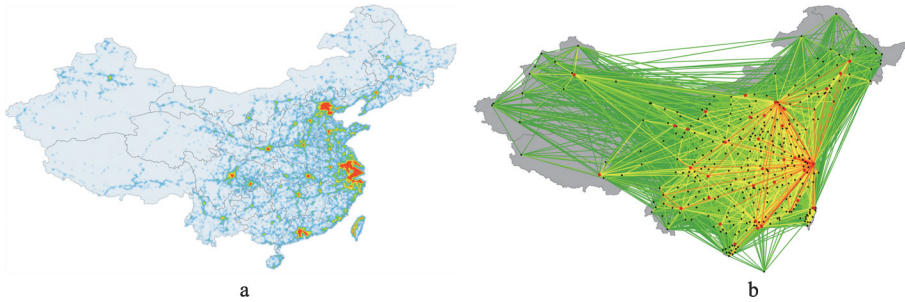


图1 利用社交媒体签到数据得到的签到点分布(a)以及城市间交互模式(b)
(其中暖色调表示高值,冷色调表示低值。出于简化原因,南海诸岛未显示)

Fig. 1 (a) Spatial distribution of check-in points; (b) Spatial interactions between 370 cities in China. Both maps are created using a social media check-in data set, where the warm and cool colors denote high values and low values, respectively. The South China Sea Islands are not shown for simplicity.

活动变化模式正好相反。因此,可以基于不同地块的活动时间变化曲线对研究区域进行土地利用分类^[6, 8]。还可以从带有时空标记的社交媒体数据获取个体的认知和情绪信息,从而在群体层面构建与不同地点相关联的语义与情感,并刻画地理空间异质性,如Dodds利用Twitter数据研究了区域的“幸福感(happiness)”分布特征^[28]。不论是活动随时间变化的特征,还是语义与情感特征,都可以用于表征地理单元间的空间差异性和依赖性,以及在此基础上展示的空间分布模式。

地理学有很久的空间交互研究传统,其中空间交互的解释模型受到了学者的广泛关注^[29]。基于大数据所反映的个体移动和联系,可以在聚集层面量化地理单元间的空间交互。例如,基于同样社交媒体签到数据获取的个体城市间移动轨迹,可以进行汇总并表达城市间的空间交互模式(图1b)^[30]。近年来,随着大数据的广泛应用,许多学者基于空间交互构建嵌入空间的网络(spatially embedded network),即网络的每个节点对应一个地理单元,并引入网络科学分析方法,如中心性度量和社区分割(community detection),从而定量评价地理单元的重要性并发现研究区的结构特征。区域划分是处理地理空间异质性的一条重要途径。在大数据的支持下,目前有两类主要分区方法。第一种方法考虑地理单元所关联的活动时变特征相似性,或语义情感的相似性,利用聚类方法,如K均值聚类,将相似性高的区域进行合并;第二种则利用地理单元之间的联系强度,利用网络社区发现算法,将联系较为紧密的地理单元划分到同一区域^[9-10]。这两种方法分别基于空间分布的依赖性和空间交互的强度,其区划结果的地理含义存在差异。前者将会得到特定属性(如土地利用)较为均质的区域,而后者得到的分区则往往拥有更为丰富的内部结构,并且可以归因于分区内地理单元功能的差异性和互补性(图2)。

2.2 空间分布和空间交互中的距离和尺度问题

2.2.1 距离衰减效应

正如Tobler第一定律所陈述的,空间邻近的区域往往具有相似的属性以及更强的交互。在空间分析中,前者可以通过空间自相关指数(如Moran's I等)加以度量,而后者则可通过重力模型等途径定量表征距离衰减。大数据提供了基于人的空间行为特征量化现象分布和交互中的距离影响的支持。

地理分布中的距离衰减效应意味着空间距离近的区域具有相似的观测值,即表现为正的空间自相关。地理空间的这种分布特性,对于空间分析至关重要,因为它是空间插值的理论基础^[25]。例如在地统计学分析中,即利用方差随距离变化表示相似度的距离衰减,进而构建插值函数。对于自然地理现象,相似度的距离衰减容易被观测和理解,如

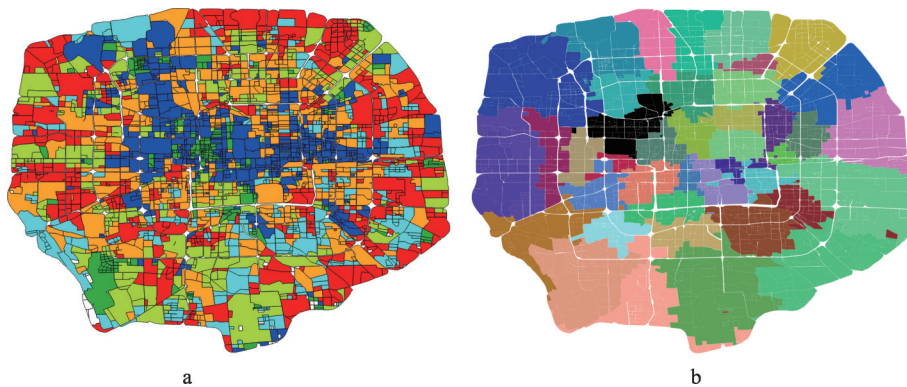


图2 北京五环以内的城区基于相似性(a)以及交互强度(b)的空间划分方法(其中不同颜色表示不同区块,采用数据为出租车轨迹数据。a. 基于出租车打车次数的日变化特征进行非监督分类得到,同一类型地块具有相似的土地利用特征;b. 基于地块之间的交通量利用社区分割算法得到,同一组团内部具有更为紧密的联系)

Fig. 2 Two regionalization schemata based on similarity (a) and spatial interaction (b). Different zones are depicted using different colors. The study area is the urban area inside the Fifth Ring Road of Beijing. Both are computed using a taxi trajectory data set. The subfigure a is created by classifying the diurnal variations of pick-ups and drop-offs in parcels, indicating that parcels in the same category have similar land use characteristics. The subfigure b is computed based on spatial interaction strengths between parcels, indicating that parcels in the same zone have relatively strong connections

Zhao 对于中国气候以及植被相似度的空间衰减进行了定量研究^[31]。而对于人文地理现象,尤其是与人的行为有关的现象,其空间自相关程度尚需进一步研究。由于地理障碍的影响,导致人文地理现象空间自相关不显著甚至负的空间自相关,典型的如城市的居住隔离(residential segregation)现象。在图3所示的北京市昌平区遥感影像中,仅仅一路(京藏高速公路)之隔,分别是别墅区和年轻中低收入者租住的城中村,无疑,其居民收入水准以及空间行为模式会有明显的差异。Liu通过扩展Moran's I 系数,利用城市出租车数据,初步印证了以上论断^[32]。人文现象的这一分布特征,使得研究中基于观测样本的插值操作^[33]需要谨慎。

Couclelis认为,所有经典的人文地理模型在表征空间时都将活动视为距离的函数^[34]。随着信息通讯技术的发展,距离的空间阻隔作用被大大削弱^[35],因此许多学者提出了“距离的消亡(death of distance)”^[36]。大数据提供了检视这一论断的支持,除了基于个



图3 人文地理现象更为强烈的变化梯度
(a. 北京市昌平区北四村以及碧水庄园别墅区的遥感影像;b. 北四村为著名的“蚁族村”,有约九万年轻人租住于图示范围内;c. 环境优美的碧水庄园别墅区,与北四村最近距离不到1 km)

Fig. 3 Rapid spatial change of human geography phenomena. (a) The remotely sensed image of Beisicun and Bishunzhuangyuan villa district. Subfigure b shows the street view of Beisicun, which is a well-known ant tribe village in Beijing. As contrast, Bishunzhuangyuan, only one kilometer away, has a beautiful view (c).

体的空间移动度量场所之间的联系强度，还可以通过用户间利用信息通讯手段（如手机通话、微博互粉）等建立的联系感知空间交互。对于前者，人或物在空间中的移动由于成本原因会出现距离衰减；而对于后者，即基于信息通讯手段建立的联系中距离影响的程度，目前的研究表明该影响依然存在。例如Kang利用手机数据，通过重力模型拟合，得到幂律（power-law）形式的距离衰减函数，其指数为0.5^[37]。这说明在基于信息通讯技术建立的联系中，距离衰减效应尽管较弱，但并非已经“消亡”。这可以归因于人们在网络空间的联系可以认为是真实世界中联系的映射，即两个区域间的社会经济关联越强，其间居民的联系也越多，即两者存在正相关关系。由于前者存在距离衰减效应，使得在群体层面的联系依然受到空间约束，而不是与距离无关。

2.2.2 尺度效应 空间分布和空间交互在传统地理学研究中，多在区域聚集层面进行分析。由于地理学缺乏天然的分析单元^[38]，研究结果依赖于空间单元形状，即产生了可变量状单元问题（modifiable areal unit problem, MAUP）^[39]。Armhein指出该问题的产生包括两种情形：从小面积单元到大面积单元的不同尺度的聚合方式，以及研究区的不同划分方案^[40]。由于大数据的基本粒度是个体，使得研究者可以同时从个体和群体两个层面观察空间分布和交互模式（图4）。很明显，在个体层面的分布和交互模式中，并不存在分析尺度的影响。只有当试图依据不同空间分析单元概括群体层面的模式时，才需要处理MAUP问题。因此可以基于大数据感知到的个体模式在不同聚合过程中的变化考察不同地理现象的尺度效应。

与空间分析中的尺度效应相类似，基于大数据研究人的空间行为模式需要注意生态学谬误（ecological fallacy）问题。由于大数据具有海量的个体样本，使得研究者可以很容易观察整个样本的空间行为模式并建立解释模型。然而，由于人群异质性（population heterogeneity）的存在，基于整个人群得到模式和模型未必适用于每个个体^[30]。例如，对于一个人群中所有具有联系的个体之间的距离分布，尽管基于手机数据已经观察到距离衰减效应^[37]，然而具体到每个人，其联系对象的空间分布未必随距离增加而变得稀疏。不论是地理空间的MAUP问题还是人群的生态学谬误问题，都需要在大数据研究中，建立微观个体到宏观群体两个层面模式的关联。

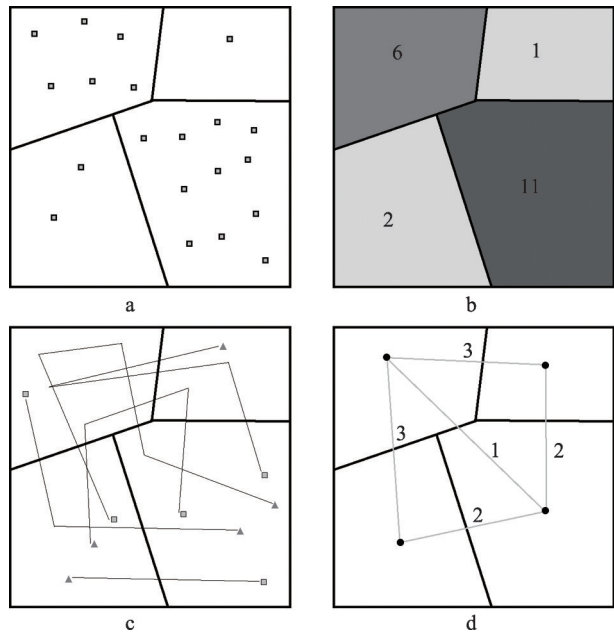


图4 个体和群体两个层面的空间分布与空间交互
 (a. 个体层面的事件点分布; b. 群体层面的分布; c. 个体层面的空间移动轨迹，其中三角符号和方形符号别表示轨迹起点和终点; d. 群体层面的空间交互模式)
 Fig. 4 Spatial distribution and interaction at both the individual and aggregate levels. (a. Individual level point distribution. b. Aggregate level distribution based on areal units. c. Individual level interaction. d. Aggregate level interaction between based on areal units.)

3 定性方法和定量方法

定性方法和定量方法是人文地理学研究中的两条重要途径。前者指通过访谈方式获

取被研究对象（如特定人群）的属性进行分析并得到结论，后者则指利用数学工具尤其是统计方法量化地理现象并构建相关定量模型（如重力模型），关美宝指出，“通过定量方法与定性方法在不同领域中的混合使用，（可以）实现对社会—文化与空间—分析隔阂的超越并形成更有洞察力的研究方法”^[41]。我们认为，大数据由于其独特的对人的空间行为模式和地理环境的感知能力，有助于支持定性方法和定量方法的混合使用。

3.1 大数据和小数据

与大数据相对应，小数据指个体化的或者小样本的数据。由于地理学很少针对个体开展研究，因此本文中小数据特指利用访谈方式获取的小样本数据，其形式既可定量，亦可定性。小数据在地理研究中的重要意义也不能被忽视，图5从数量的角度比较了大数据与小数据的区别，并给出了基于大数据与小数据的不同研究实例。我们可以用数据表做比喻描述它们的差异，即大数据体现为“行数多而列数少”，而小数据则“行数多而列数少”。通常，大数据样本量较大，但是缺乏丰富的属性内容。例如，基于手机通话数据，可以提取手机用户的移动轨迹以及社交关系，然而移动轨迹对应的出行目的等信息却无从得知。与之相反，小数据样本量较小确拥有相对全面的描述性信息，如人的年龄、性别、家庭状况、历史迁居等，如在Kwan针对穆斯林妇女的环境恐惧研究中，采用访谈方式对37名受访对象进行了调查^[42]，采集了每个个体的活动及感受的细节信息。由于大数据和小数据的差异，也有将大数据视为“薄数据（thin data）”，而小数据视为“厚数据（thick data）”的说法^[43]。

定性研究中获取的访谈数据，除了样本量较小外，根据研究的目的，在属性上通常具有较高的均质性，例如在文献[42]中，受访对象具有“穆斯林”和“女性”两个特征，而文献[44]则针对居住于单位大院的城市居民。与之相反，大数据的采集由于未经采样设计，所反映的人群通常是异质的。González发现了人群异质性对于所观测到的移动性模式的影响^[45]，Xie则进一步指出人群异质性是社会学研究需要处理的重要问题^[46]。对于地理研究而言，人群异质性使得基于大数据提取的模式较为平凡而针对性不强，这约束了大数据的应用价值；此外，样本有偏以及属性偏少的缺陷也影响了解释性模型的构建。例如，出租车轨迹数据只能反映了一个城市中的部分出行而且无法获取出行目的，使得难以基于该数据针对特定出行需求（如就医）优化城市规划从而减少出行总量。

因此，尽管目前大数据已经被广泛应用，但为了弥补大数据的上述不足，小数据的重要性依然不可忽视^[16]。从“人”的角度出发，大数据与小数据的集成需要解决大数据的人群异质性与属性信息少的问题。一条可行的途径是根据大数据所反映的空间行为模式，对人群进行聚类或根据预设规则识别出特定群组，从而得到相对均质的子集。该方

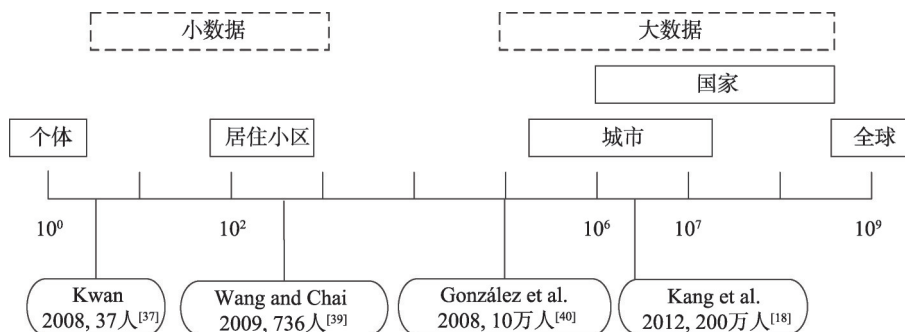


图5 不同实证研究中人群样本大小的差异，表达了大数据与小数据的区别

Fig. 5 Population sizes of different studies, indicating the different between big data and small data

法相当于增加了数据列数，从而更好地支持与小数据的集成，即可以通过传统方式对感兴趣子集收集更为丰富的属性信息。在实践中，对人群分组的依据包括空间行为模式的相似性以及社交关系强度。例如，Long利用公交卡数据提取了城市居民的极端出行行为，如早出晚归、长距离通勤等^[47]，从而识别出了部分城市低收入人群的空间移动特征；而Shi利用手机通话数据反映的个体联系，对城市居民进行分组，进而定量分析了特定人群（如由于单位集体购房的职住分离人群）及其时空移动模式^[48]。

3.2 空间与场所

空间（space）和场所（place）是理解地理环境的两条重要途径^[49,50]。空间定义了地理分析的参考框架，空间视角的分析方法注重坐标、几何、距离等精确的度量^[51]。而场所则与个人的体验有关^[52]，在GIS中对于场所多基于地名及地名间的关系等定性方式建模。大数据不仅支持空间视角的分布和交互分析手段（参见第二节），而且为理解场所提供了基础^[5]。由于大数据对于人的空间行为模式的揭示能力，使得可以从语义与情感、人群活动、空间交互等途径描述与一个场所相关联的人的体验。Sui和Goodchild认为空间和场所分别提供了“自上而下”和“自下而上”的分析地理问题的视角^[53]。如图6所示的UML类图，一个空间中可以嵌入多个场所，每个场所在定性GIS实现中拥有自己名称并且与其它场所具有不同的空间关系^[51]。然而，人文地理学研究更关注人的场所体验，在传统研究中，该体验通常基于访谈、问卷等途径获得。

大数据从语义与情感、人群活动、交互模式等三个途径提供了对场所体验的感知（Sense of place）手段。图7利用带位置的新浪微博数据，分析了与“北京大学”这一场所相关联的空间行为特征。其中，图7a展示了微博内容的关键词频率，部分反映了相应微博用户的场所情感。由于微博带有时空标记，还能够分析校园内以大学生为主体的人群的活动时间分布（图7b）和空间分布（图7c）特征，其中图7c所展示的活动热区分布与宿舍以及教学区高度吻合。图7c还展示了一个很有趣的现象，就是北京大学校园景色

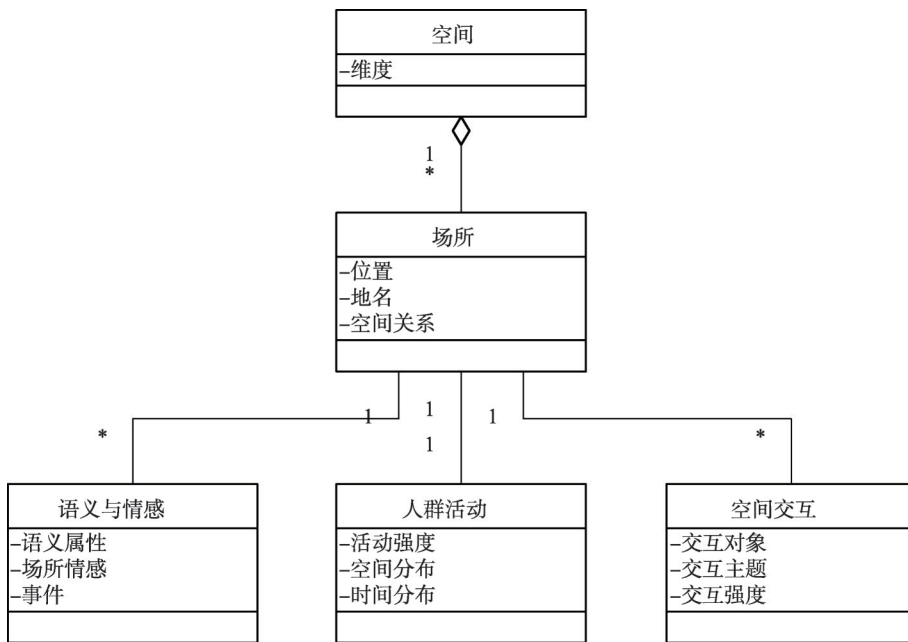


图6 空间、场所以及与场所相关联的人的体验

Fig. 6 A class diagram of place, representing the concepts of space, place, and human experiences and their relationships

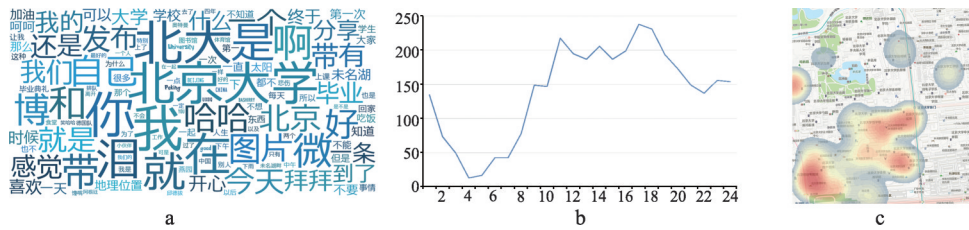


图7 利用微博数据对一个特定场所(北京大学)相关联空间行为的表达

(a. 利用微博数据提取得到的主题词; b. 在北京大学校园范文内发布微博时间的分布曲线, 其中横轴为时间, 间隔为1小时, 纵轴为每个小时发布的微博数量, 该曲线反映了校园活动的时间特征; c. 北京大学校园内发布微博的空间分布)

Fig. 7 Representing the activities and experience associated with a particular place, the campus of Peking University, using social media data (Sina Weibo). (a. Word cloud of Weibo contents submitted in the campus. b. Temporal activity variation of posting Weibo. c. Spatial distribution of all Weibo posts inside Peking University)

最为漂亮的未名湖区微博签到分布量并不高。由于数据时间范围为学期中间, 校园内的签到微博主要由大学生发布。一方面大学生学业紧张, 使得他们活动范围主要集中于宿舍和教室; 另一方面本校学生对于未名湖景色习以为常, 缺少在该区域发布微博的动机。根据文献的研究, 旅游过程中的签到数量会偏高^[30], 因此, 可以想象到暑假期间, 由于有大量游客参观北大校园, 未名湖区的签到数量将会大大增加。以上通过微博所反映的时空行为模式, 都体现了以大学生为主的相对均质的人群对于其生活的校园这样一个场所的体验。尽管信息通讯技术提高了人们的移动性和联系性, 降低了场所与个体行为之间的耦合关系^[35], 大数据所提供的群体体验依然有助于研究者理解一个场所的特征。

大数据同样可以用于感知场所的空间范围。由于场所与人的体验有关, 属于Fiat对象而其边界通常具有含糊性^[54]。基于大数据中具有相似语义的点数据, 可以构造聚合点集 (clustered point set) 模型以表达场所的空间范围^[55], 该模型能够较好处理场所边界含糊性^[56]。此外, 由于该方法“自下而上”的特征, 可以防止特定观察尺度的丢失, 从而在一定程度上规避了MAUP问题^[57]。如前所述, 这种自下而上的思路同样适合于人群的聚类, 即基于人的时空行为特征, 聚合得到相对均质的群组。由于人群和场所也是定性研究方法的两个重要主题, 从大数据中识别出有意义的人群和场所无疑有助于与定性方法的结合。这种结合使得我们不仅能够观察到人群与场所耦合的模式, 而且可以探究其机理和演化过程。

4 结论

随着大数据时代的到来, 地理学 (尤其是人文地理学) 研究拥有了全新的社会感知手段。就如同显微镜的出现, 使我们能够观察到之前看不见的对象 (如细胞等), 从而促进了生物学发展那样, 大数据使得我们能够观察海量个体的细节空间行为并感知地理环境特征, 从而为地理学打开了一条“从人到地”的研究路径。目前, 在人地关系研究中, 地理环境的概念外延已经拓展到建成环境 (built environment) 乃至行为环境 (behavioral environment) ^[58], 大数据所提供的空间行为及环境感知能力, 无疑可为研究人地关系这一地理学传统命题提供有力支撑。

大数据独特的“大样本”、“细粒度”的特点, 促使我们在分析处理大数据过程中深入思考地理学的基本问题。本文针对基于大数据感知空间分布和空间交互, 以及其中蕴

舍的距离、尺度等问题进行了讨论,指出量化距离衰减效应时及进行相关分析时需要考虑人群异质性,而基于个体层面的分布和交互进行自下而上的聚合,可以在空间分析中处理尺度效应。这种基于大数据的聚合方法形成了对人和空间的解构,能够识别特定属性的人群和场所,考虑到后两者是定性地理研究中的主要关注对象,我们认为大数据有助于实现定性方法和定量方法的混合使用。

尽管大数据作为一个整体概念,对于地理学研究的意义毋庸置疑。然而,在具体的研究及应用中,也要注意一些问题。首先,就目前得到广泛关注的手机等地理空间大数据而言,大都存在代表性不足以及语义信息偏少的问题^[5]。例如,基于出租车数据,只能获取城市居民特定出行目的的信息,其中会过低估计通勤出行的比例^[59]。换言之,大数据并非是全数据,研究中需要对于数据的代表性要有充分认识。解决大数据的上述两点不足,一个可能的途径是集成多源数据以及基于调查问卷等途径获取的小数据。其次,大数据主要是提供了一种观察手段并且具有一定预测能力,依然需要去构建模型,解释其中蕴含的行为模式和地理规律。否则,其预测结果可能会不可靠,就像Google流感预测失效那样^[43, 60]。在这个过程中,尽管有许多合适的定量空间分析方法可以用于分析处理大数据,但是同样不能忽视场所视角的定性方法。因此,对致力于地理空间分析方法的地理信息科学研究者而言,需要重视GIS的地理学根基。我们认为,大数据及其所提供的社会感知手段,有助于重新审视地理学的基本概念和基本问题;而在应用中为了避免大数据的不足,同样需要对地理学基本理论深刻而全面的把握。

致谢: 感谢隋正伟博士、康朝贵博士、李栋博士提供相关数据并协助制图;论文准备过程中王法辉教授、柴彦威教授、吴志峰教授、马修军副教授、李栋博士提供了完善意见,感谢以上专家及审稿人的建设性意见。

参考文献(References)

- [1] Batty M. Big data, smart cities and city planning. *Dialogues in Human Geography*, 2013, 3(3): 274-279.
- [2] Goodchild M F. The quality of big (geo)data. *Dialogues in Human Geography*, 2013, 3(3): 280-284.
- [3] Kitchin R. Big data and human geography: opportunities, challenges and risks. *Dialogues in Human Geography*, 2013, 3(3): 274-279.
- [4] Miller H J, Goodchild M F. Data-driven geography. *GeoJournal*, 2015, 80(4): 449-461.
- [5] Liu Yu, Liu Xi, Gao Song, et al. Social sensing: A new approach to understanding our socio-economic environments. *Annals of the Association of American Geographers*. 2015, 105(3): 512-530.
- [6] Liu Yu, Wang Fahui, Xiao Yu, et al. Urban land uses and traffic "source-sink areas": Evidence from GPS-enabled taxi data in Shanghai. *Landscape and Urban Planning*, 2012, 106:73-87.
- [7] Liu Xi, Kang Chaogui, Gong Li, et al. Incorporating spatial interaction patterns in classifying and understanding urban land use. *International Journal of Geographical Information Science*, 2016, 30(2): 334-350.
- [8] Pei Tao, Sobolevsky S, Ratti C, et al. A new insight into land use classification based on aggregated mobile phone data. *International Journal of Geographical Information Science*, 2014, 28(9): 1988-2007.
- [9] Liu Xi, Gong Li, Gong Yongxi, et al. Revealing travel patterns and city structure with taxi trip data. *Journal of Transport Geography*, 2015, 43: 78-90.
- [10] Ratti C, Sobolevsky S, Calabrese F, et al. Redrawing the map of Great Britain from a network of human interactions. *PLoS ONE*, 2010, 5(12): e14248.
- [11] Cai Bofeng, Wang Jinnan, Long Ying, et al. Evaluating the impact of odors from the 1955 landfills in China using a bottom-up approach. *Journal of Environmental Management*. 2015, 164: 206-214.
- [12] Crooks A, Croitoru A, Stefanidis A, et al. #Earthquake: Twitter as a Distributed Sensor System. *Transactions in GIS*, 2013, 17(1): 124-147.
- [13] Li Linna, Goodchild M F, Xu Bo. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr.

- Cartography and Geographic Information Science, 2013, 40(2): 61-77.
- [14] Yang Wei, Mu Lan, Shen Ye. Effect of climate and seasonality on depressed mood among twitter users. *Applied Geography*, 2015, 63: 184-191.
- [15] Gong Jianhua, Lin Hui. Exploring human-oriented GIS. *Geomatics and Information Science of Wuhan University*, 2006, 31(8): 704-708. [龚建华,林琤.面向地理环境主体GIS初探.武汉大学学报(信息科学版). 2006, 31(8): 704-708.]
- [16] Zhen Feng, Wang Bo. Rethinking human geography in the age of big data. *Geographical Research*, 2015, 34(5): 803-811. [甄峰,王波.“大数据”热潮下人文地理学研究的再思考.地理研究, 2015, 34(5): 803-811.]
- [17] Getis A. Spatial interaction and spatial autocorrelation: A cross-product approach. *Environment and Planning A*, 1991, 23(9):1269-1277.
- [18] Fischer M M, Reisman M, Scherngell T. Spatial interaction and spatial autocorrelation. In: Anselin L, Rey S J (ed.) *Perspectives on Spatial Data Analysis*, 2010: 61-79.
- [19] Ullman E L. Geography as spatial interaction. *Annals of Association of the American Geographers*, 1954, 44: 283-284.
- [20] Hartshorne R. The concept of geography as a science of space, from Kant and Humboldt to Hettner. *Annals of the Association of American Geographers*, 1958, 48(2): 97-108.
- [21] Castells, M. *The rise of the network society. The information age: Economy, society and culture.* 1996, vol. 1, Blackwell, Oxford.
- [22] Tobler W. A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 1970, 46(2):234-240.
- [23] Goodchild M F. The validity and usefulness of laws in geographic information science and geography. *Annals of the Association of American Geographers*, 2004, 94(2):300-303.
- [24] Sui D. Tobler's first law of geography: A big idea for a small world? *Annals of the Association of American Geographers*, 2004, 94(2), 269-277.
- [25] Miller H J. Tobler's First Law and spatial analysis. *Annals of the Association of American Geographers*, 2004, 94(2): 284-289.
- [26] Reades J, Calabrese F, Ratti C. Eigenplaces: Analysing cities using the space - time structure of the mobile phone network. *Environment and Planning B: Planning and Design*, 2009, 36(5): 824-836.
- [27] Kang Chaogui, Liu Yu, Ma Xiujun, et al. Towards estimating urban population distributions from mobile call data. *Journal of Urban Technology*, 2012, 19(4):3-21.
- [28] Dodds P S, Harris K D, Kloumann I. M., et al. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE*, 2011, 6(12): e26752.
- [29] Fotheringham A, O'Kelly M E. *Spatial Interaction Models: Formulations and Applications.* 1989, Kluwer.
- [30] Liu Yu, Sui Zhengwei, Kang Chaogui, et al. Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PLoS ONE*, 2014, 9(1): e86026.
- [31] Zhao Z, Li S, Liu J, et al. The distance decay of similarity in climate variation and vegetation dynamics. *Environmental Earth Sciences*, 2015, 73(8): 4659-4670.
- [32] Liu Yu, Tong Daoqin, Liu Xi. Measuring spatial autocorrelation of vectors. *Geographical Analysis*, 2015, 47(3):300-319.
- [33] Weber J, Kwan Mei-Po. Bringing time back in: A study on the influence of travel time variations and facility opening hours on individual accessibility. *The Professional Geographer*, 2002, 54(2): 226-240.
- [34] Couclelis H. Misses, near-misses and surprises in forecasting the informational city. In: Miller H J (ed.) *Societies and Cities in the Age of Instant Access.* The GeoJournal Library, vol. 88, Berlin, Germany: Springer. 2007: 71-83.
- [35] Miller H J. Place-based versus people-based geographic information science. *Geography Compass*, 2007, 1(3): 503-535.
- [36] Cairncross F. *The Death of Distance: How the Communications Revolution will Change Our Lives.* Harvard Business Press, 2001.
- [37] Kang Chaogui, Zhang Yi, Ma Xiujun, et al. Inferring properties and revealing geographical impacts of intercity mobile communication network of China using a subnet data set. *International Journal of Geographical Information Science*, 2013, 27(3): 431-448.
- [38] Longley P, Goodchild M F, Maguire D J, et al. *Geographic Information Science and Systems (4th Edition).* 2015, Wiley.
- [39] Openshaw S. *The Modifiable Areal Unit Problem.* 1983, Geo Books, Norwick.
- [40] Armhein C. Searching for the elusive aggregation effect: Evidence from statistical simulations. *Environment & Planning A*, 1995, 27(1): 105-119.
- [41] Kwan Mei-Po. Beyond binaries: Reflections on hybrid geographies. *Progress in Human Geography*, 2013, 32(9): 1307-

1315. [关美宝. 超越地理学二元性: 混合地理学的思考, 地理科学进展, 2013, 32(9): 1307-1315.]
- [42] Kwan Mei-Po. From oral histories to visual narratives: Re-presenting the post-September 11 experiences of the Muslim women in the United States. *Social and Cultural Geography*, 2008, 9(6): 653-669.
- [43] Krenchel M., Madsbjerg C. Your big data is worthless if you don't bring it into the real world. 2014. <http://www.wired.com/2014/04/your-big-data-is-worthless-if-you-dont-bring-it-into-the-real-world/>
- [44] Wang Donggen, Chai Yanwei. The jobs-housing relationship and commuting in Beijing, China: the legacy of Danwei. *Journal of Transport Geography*, 2009, 17(1): 30-38.
- [45] González M C, Hidalgo C A, Barabási A B. Understanding individual human mobility patterns. *Nature*, 2008, 458, 779-782.
- [46] Xie Yu. Population heterogeneity and causal inference. *Proceedings of the National Academy of Sciences of the USA*, 2013, 110(16): 6262-6268.
- [47] Long Ying, Liu Xingjian, Zhou Jiangping, et al. Early birds, night owls, and tireless/recurring itinerants: An exploratory analysis of extreme transit behaviors in Beijing, China. 2015, arXiv:1502.02056.
- [48] Shi Li, Chi Guanghua, Liu Xi, et al. Human mobility patterns in different communities: a mobile phone data based social network approach. *Annals of GIS*, 2015, 21(1): 15-26.
- [49] Wainwright J, Barnes T. Nature, economy, and the space-place distinction. *Environment and Planning D: Society and Space*, 2009, 27: 966-986.
- [50] Agnew J. Space and place. In: Agnew J, Livingstone D. (ed.) *Handbook of Geographical Knowledge*. 2011, Sage, London.
- [51] Goodchild M F. Space, place and health. *Annals of GIS*, 2015, 21(2): 97-100.
- [52] Tuan Yi-Fu. *Space and Place: The Perspective of Experience*. 1977, University of Minnesota Press, Minneapolis.
- [53] Sui D, Goodchild M F. The convergence of GIS and social media: challenges for GIScience. *International Journal of Geographical Information Science*, 2011, 25(11): 1737-1748.
- [54] Smith B, Varzi A. Fiat and bona fide boundaries. *Philosophy and Phenomenological Research*, 2000, 60(2): 401-420.
- [55] Gao Song, Li Linna, Li Wenwen, et al. Constructing gazetteers from volunteered Big Geo-Data based on Hadoop. *Computers, Environment and Urban Systems*, in press. (doi: 10.1016/j.compenvurbsys.2014.02.004)
- [56] Liu Yu, Yuan Yihong, Xiao Danqing, et al. A point-set-based approximation for areal objects: A case study of representing localities. *Computers, Environment and Urban Systems*, 2010, 34(1): 28-39.
- [57] Mu Lan, Wang Fahui. A scale-space clustering method: Mitigating the effect of scale in the analysis of zone-based data. *Annals of the Association of American Geographers*, 2008, 98(1): 85-101.
- [58] Golledge R G. The nature of geographical knowledge. *Annals of the Association of American Geographers*, 2002, 92(1): 1-14.
- [59] Gong Li, Liu Xi, Wu Lun, et al. Inferring trip purposes and uncovering travel patterns from taxi trajectory data. *Cartography and Geographic Information Science*. 2016, 43(2): 103-114.
- [60] Lazer D, Kennedy R, King G, et al. The parable of Google Flu: traps in big data analysis. *Science*, 2014, 343(6176): 1203-1205.

Revisiting several basic geographical concepts: A social sensing perspective

LIU Yu

(Institute of Remote Sensing and Geographical Information Systems, Peking University, Beijing 1000871, China)

Abstract: Recently, various big data are drawing more and more attention in geographical research and many scholars have conducted lots of empirical studies using mobile phone data, social media data, taxi data, and so forth. Social sensing, a newly proposed concept, represents the capability that big data reveal socio-economic geographical features by capturing the spatial behavior patterns of a large population. Given that the term "environment" in human-environment relation studies has been expanded to include the behavioral environment, social sensing techniques provide us a new approach to understanding human-environment relations. Additionally, the emergence of social sensing helps us to rethink several fundamental issues in geographical studies. This article revisits two groups of core concepts: spatial distribution and spatial interaction, and qualitative method and quantitative method. Given that big data can measure distributions and interactions at both individual and aggregate levels, we can quantify the underlying distance and scale effects from the observed patterns. To tackle space and population heterogeneity, clustering methods can be introduced to decompose a space and/or a population into relatively homogeneous human groups and places. Considering that human groups and places are essential to qualitative studies, we argue that social sensing offers an opportunity to integrate big data and survey-based small data, and consequently, qualitative and quantitative methods. Obviously, the second merit makes it possible to construct hybrid geography.

Keywords: big data; social sensing; spatial distribution; spatial interaction; qualitative method; quantitative method