

大数据对人文—经济地理学研究的促进与局限

杨振山^{1*}, 龙 瀛², Nicolas DOUAY³

(1. 中国科学院地理科学与资源研究所, 中国科学院区域可持续发展分析与模拟实验室, 北京 100101;
2. 北京市城市规划设计研究院, 北京 100045; 3. 法国巴黎第七大学, 巴黎 75205, 法国)

摘 要:大数据技术的诞生不仅快速推动着社会的进步,而且也将科学研究不断引向新的高度。以人类社会经济活动为主要研究对象的人文—经济地理学与当前大数据建设和发展趋势具有高度一致性,大数据的发展对丰富和完善人文—经济地理学势必起到积极的推动作用,同时也对人文—经济地理学的学科思维和研究方法提出了新的挑战。梳理和分析了目前大数据在人文—经济地理学主要研究领域,包括城市内部空间研究、交通与消费行为、社会空间与社会网络研究中的最近进展,以及大数据对参与式研究和决策平台的作用。着重剖析了大数据对人文—经济地理学数据获取,研究思维与范式,研究内容、研究时空尺度与研究目标等方面的促进作用与存在问题,特别是由于大数据自身发展的不完善,在数据收集特别是数据属性方面还存在很大的局限,缺乏理论基础将会使得大数据与实际应用受到很大限制,同时,数据本身也不能替代研究者思维和决策过程。因此,人文—经济地理学者应该科学对待大数据所带来的机遇,弥补和丰富以往发展中的短板,即完善学科数据建设、建立大数据应用较为完善的研究方法体系,促进跨域数据整合和跨域研究,以及推进研究对象和研究目的的转变。

关键词:大数据;人文—经济地理学;研究范式

1 引言

数据与信息收集是科学研究和政府以及企业决策重要环节。自IT技术发展以来,人们获取数据的技术一日千里,同时全球各行各业,各个领域对于数据的依赖性也大幅度提高。特别是自20世纪90年代后,IT领域开展了以大数据技术、云计算和物联网为标志的又一次颠覆性的变革,这三项技术之间相互促进,其中大数据最为直接地与社会经济活动联系在一起。作为以社会经济活动为主要研究内容的人文—经济地理学,在大数据时代中既面临深化学科研究内容的机遇,又须面对调整传统研究思维的挑战。

数据是科学研究工作的重要基础,大数据的发展对科学研究来说具有不可思议的魅力。20世纪中叶,以数学模型为代表的定量研究方法从纷杂的社会经济现象中抽取典型性数据,通过数学模型的方式揭示社会经济活动的某种联系,从而使得社会

经济现象的研究方法从传统描述性(Descriptive)研究进入到解释性(Explanative)研究。人文—经济地理学者曾经经历过对数学模型的沉醉、抛弃和理性回归(杨振山等,2010),如今大数据的快速发展似乎使得研究者又将面临相似的局面。不同的是,这次将从研究范式、数据收集与特点、研究内容和研究目的发起更多疑问。

鉴于此,本文从国内外已有的与人文—经济地理学相关的案例和经验出发,结合其学科研究思维、范式、特点和任务对此加以剖析,希望能够在这一新的技术变革下引起争鸣,促进人文—经济地理学积极面对这一重要转变。

2 大数据的发展及其在人文—经济地理学中的应用

人文—经济地理学和大数据有着内在的本质联系。这是因为,以人类社会经济活动为主要研究

收稿日期:2015-03;修订日期:2015-04。

基金项目:国家自然科学基金青年一面上连续项目(41371008);中国科学院青年创新促进研究会项目(2014042)。

作者简介:杨振山(1979-),新疆博乐人,副研究员,主要从事城市与区域研究,E-mail: yangzs@igsrr.ac.cn。

引用格式:杨振山,龙瀛,Nicolas DOUAY. 2015. 大数据对人文—经济地理学研究的促进与局限[J]. 地理科学进展, 34(4): 410-417. [Yang Z S, Long Y, Douay N. 2015. Opportunities and limitations of big data applications to human and economic geography: the state of the art[J]. Progress in Geography, 34(4): 410-417.]. DOI: 10.11820/dlkxjz.2015.04.002

对象和研究内容的人文—经济地理学与当前大数据的建设和发展趋势高度一致。另外,人文—经济地理学是研究人和社会经济活动在地球表层或一定空间、时间范围内的规律和特征。与传统数据相比,大数据的收集和记录都更加强调时间和空间维度,试图在时空轨迹中,无时无刻、无处不在地记录所发生的事件,从而构成海量的数据规模(Volume)、快速的数据流转和动态的数据体系(Velocity)、多样和多元化的数据类型(Variety)和巨大的数据价值(Value)等基本特征(Laney, 2001)。从数据内容来看,当前大数据多是以人类社会经济活动为主要内容(如社交媒体、出行轨迹等)。在一定程度上,大数据技术用数据的方式反应了特定的社会经济文化现象,揭示了社会经济文化的复杂性。因此,大数据时代对于作为从空间角度研究人类社会经济活动的人文—经济地理学来说具有非常重要的意义。目前,大数据技术已经在人文—经济地理学的下述领域中得到一些初步应用。

(1) 城市功能区、热点区和边界鉴别

城市功能区、热点区和边界是城市空间结构的核心组成部分。其中城市功能区是实现城市功能的载体,城市热点区是某种社会经济活动的空间集聚区域,二者都是城市空间中重要地区,也是政府管理和市场行为的聚焦区域。城市增长边界对于控制城市规模,合理促进城乡协调发展具有重要意义(Nelson et al, 1995)。长期以来,上述问题的研究一般通过对规划文本的收集和土地利用分析来获得,其中土地利用分析重点在于说明城市景观的变化。在此类研究中,功能区、热点区和增长的边界具有一定的“被划定性”。随着遥感信息的丰富,研究者也在试图通过卫星图像和灯光遥感技术来识别城市的边界(Tan et al, 2010; Jiang et al, 2012),但距离以社会经济功能来刻画城市功能区和增长边界还有不小差距。

大数据技术促进了从社会经济活动来揭示城市的边界和功能区这一方面的研究。例如近来学者利用大数据,通过兴趣点(Point of Interest, POI)来获得人们对某一感兴趣地区的访问次数,从而对城市的热点区和边界加以甄别(Long et al, 2013; 秦萧等, 2013)。Long 等(2014)利用道路网络和POI数据,对中国城市的城市边界进行了地块尺度刻画,综合考虑了物理设施和人类活动对应的兴趣点。重要的是,因为大数据的采集可以是任意时段的,那么理论上对空间利用的分析就可以是动态的过

程,这对传统分析方法来说无疑是一个巨大进步。

(2) 行为和消费研究

丰富的数据获取渠道和大规模的数据量为研究交通出行和消费奠定了坚实的基础。随着移动互联网技术的广泛应用和发展,个人数据的产生、记录、存贮已经成为城市居民日常生活的一部分,特别是智能手机数据、公交智能卡数据、签到数据、出租车轨迹数据为交通和行为研究提供了丰富的数据来源,极大地推进了人文—经济地理学中行为地理和交通地理学研究(Bentley et al, 2014)。如Jun 等(2014)利用交通数据了解消费者行为;Wang 等(2014)通过电话记录比较城市和乡村居民经济行为的差异;Cai 等(2014)根据已经掌握的出租车轨迹来建立电动车充电桩方案;龙瀛等(2012)利用公交卡数据来分析人们出行的行为规律。

(3) 社会空间与社会网络分析

随着全社会人文关怀程度的提高,近年人文—经济地理学中的社会地理学得到很大发展。情感、文化、性别以及公平和民生等都成为社会地理学中重要话题。大数据在这一方面具有独特的优势,它通过社交媒体和社交网站表达和收集大众心声,成为获取上述话题研究所需数据的重要来源(秦萧等, 2013)。一些研究利用大数据来揭示城市的社会空间构成,如犯罪空间(Graham et al, 2014)。在网络和计算机技术的推动下,人们还通过获取社会行为标签(Tag)来反映人们的社会行为和网络群体的关注(Zhang et al, 2011),并在社会网络图中辨识区域所在节点的特殊性和作用。如何从认知论的角度将标签信息化成为这一类研究的关键和难点。

3 大数据研究范式的冲击与人文—经济地理学研究范式的不可替代性

由于大数据在研究思维和研究方式等很多方面与传统研究有较大不同,因此在推动人文—经济地理学研究的同时,大数据也激发了人文—经济地理学的思考。首当其冲是关于研究范式。人文—经济地理学存在人文主义、结构主义和实证主义等研究范式,从不同角度揭示和解释社会经济活动规律,以及人与自然环境之间的交互关系。受20世纪80年代以来西方新经济地理学的“文化转向”和“制度转向”的影响(刘卫东等, 2004),人文—经济地理学开始关注文化、制度,将其作为社会发展格局和动力机制的深层次原因。此后关系地理(Relation)、

演化地理(Evolution)等研究范式也为人文—经济地理学者理解空间经济要素及其相互关系提供了重要的理论基础(苗长虹等, 2007; 贺灿飞等, 2014)。在人文—经济地理学发展壮大的过程中, 理论指导发挥着关键性作用。

“让数据说话(Data Tell the Truth)”是大数据技术发出的最强音。同时, 在数据处理上, 大数据技术具有“要全体不要抽样、要效率不要绝对精确、要相关不要因果”3个显著性特征(Mayer-Schonberger et al, 2013)。在没有理论和任何假设的基础上, 大数据技术是由最终数据结果所呈现的具体形态、格局和问题来反映真实世界。数据所揭示的内容就是真实世界本身, 而科学研究所需要进行的只是对数据技术(收集、过滤和表达等)进行完善, 对数据价值不断挖掘, 从“事物的相关性”中去预测。

上述研究范式无疑对人文—经济地理学的科学研究范式构成了重大冲击。“如果你知道一个人过去的所有社会数据, 那么你对他未来行为的预测的准确性将达到93%(Barabasi, 2010)”。大数据真的只给除数据之外的科学研究留下7%的空间吗? 人文—经济地理学的研究是否只要交给数据科学家就可以了呢? 大数据对人文—经济地理学最大的挑战可能在于“一切让数据说话”的口号。大数据研究似乎更加“单刀直入”, 希望通过全样本数据来深刻地揭示研究对象, 从而对研究内容进行“无微不至”的深刻揭示。然而, 没有任何理论假设让大数据的使用在科学研究中失去“先验性”科学设想。一个个数据结果, 一张张数据图最终还是会和特定的社会发展现象联系起来, 然后再回到问题研究的整体脉络中进行分析, 并解释其原因。大数据研究有时可能仅仅简单到数据分析(Boyd et al, 2012)。所以从这一点来说, 如果大数据分析离开原有人文—经济地理学长期积累的学科知识和理论体系, 将无法体现大数据巨大的价值(Value)。

人文—经济地理学研究往往是采用的描述、分析、解释和预测这些主要研究方式的一种或者几种。大数据在描述上具有无与伦比的优势。但是当前我们看到的的数据结果却常常很相似或有雷同, 即海量复杂数据最终生产出千篇一律的结果。理论性的先天不足又使得大数据很难在分析和解释上有所进展。另外, 依托数据进行科学预测也需要根据先验知识或者参照某一理论来进行。由此看来, 数据是大数据的魅力所在, 但过度强调则可能会使大数据自身发展丧失其巨大价值(Value)得到挖掘的机会。人文—经济地理学在这方面与大数据

据技术的发展是互补性关系, 利用大数据海量的数据规模(Volume)、快速的数据流转和动态的数据体系(Velocity)、多样和多元化的数据类型(Variety), 人文—经济地理学无论是检验和深化既有理论体系, 还是发展新的学科内容都具有无限的可能性, 同时也将有助于大数据本身巨大价值的发挥。

4 大数据对人文—经济地理学数据源的革新与使用局限

人文—经济地理学研究, 尤其是定量研究常常依赖于官方数据。通常, 官方数据是自上而下的数据结构, 往往不能满足研究和应用的需要, 如房地产研究常常需要实时房价信息。因此, 人文—经济地理学中很多研究问题, 特别是周期性、长期性研究受到很大的局限。

在信息化、数字化、网格化和云计算等技术的辅助支撑下, 大数据充分地利用空间技术或地学技术, 如遥感和地理信息系统和空间定位技术等, 实时对研究兴趣点(Point of Interest)进行数据捕捉。收集的数据既不同于一贯采用的以“人”为媒介的调查方法所获得的数据, 也不是官方发布的“硬数据”, 而是通过“机器”实时收集的人文—经济地理学“一手数据”和“软数据”。因此, 大数据技术为人文—经济地理学者提供了另一数据收集手段(甄峰等, 2014)。在“机器”的帮助下, 数据数量和种类都极大地丰富了。

但是, 随之也会产生一些问题。首先是数据的使用“寿命”大大缩短了, 如交通流、车辆拥堵情况等都被具体定格化。统计资料可能需要每个月、每年(如经济统计资料)或几年(如普查)发布一次, 这些官方资料由于其发布时间的稳定性、综合性和权威性, 通常可以被看作是人文社会经济领域的“硬数据”。而通过大数据方式采集的数据只是时空的一个片段, 是“硬数据”的一个个子集或细化单元。当然, 如果可以对大数据进行长期的记录来分析其规律性, 那么一个个子集描述的情况远比综合(Aggregate)、代表性描述准确得多, 信息也丰富得多。但是即便如此, 大数据中的单个记录(Record)远没有传统的统计资料中的记录那么有意义。

其次, 尽管大数据宣称是用全样本在做数据分析, 但其实调查群体存在很大限定性, 或者说存在数据有偏性。如社交网络数据是基于对使用社交网络人群的调查, 对于不用社交网络的人来说, 如

大多数老年人、未入学的少年儿童和因贫困没有能力购买社交电子产品的人来说,大数据技术显然在目前是无能为力的。再如获取公交刷卡记录的前提是对使用公共交通者的出行规律进行调查,对于通勤范围很短,没必要使用公共交通工具的人来说,大数据的使用便存在一定局限性。而针对数据有偏性,一些研究已经开始积极采取对应方法,例如加强对局部群体的行为特征研究,或者是结合其他数据对同一问题进行研究,来保证研究结果的稳定性(Long et al, 2014)。此外,对人文—经济地理学者而言,一个很大的遗憾是大数据只记录了空间兴趣点的空间信息,其他社会和经济属性甚少,大大削弱了大数据在人文—经济地理学中的应用潜力。

再次,我们还需要清楚地记得,人文—经济地理学包含定量和定性两种方法。从开始的实证主义描述到计量经济革命,再到20世纪90年代后期空间经济的兴起,以及社会文化地理的百花齐放,人文经济地理最终走向了定性和定量研究相结合的局面,以提高解决复杂科学问题的能力,具体方法包括描述、解释、格局和模式识别、机理分析与趋势预测等。而大数据更多地是定量研究,需要定性研究来配合使得研究更加丰富和完善。即使就定量研究而言,提高大数据的分析能力,还需要新的理论和分析方法。可以想象,目前的统计分析工具并不完全适应全样本数据。为此,一些研究者在探索新的方法,特别是利用数据分析方法来促进大数据在人文—经济地理学中的应用,如通过对数据分布中异常值的揭示来发现城市的个性(Jiang, 2015)。

最后,大数据对科学研究也带来了另一个严肃的问题,即科学隐私。所有的人完全暴露在各种数据采集器之下,自身的行为、特征等完全不受自身意愿掌控,被采集并加以分析,甚至有可能会被公布。另外,在利用大数据进行决策的时候,也注意到数据分析是对以往情况的总结,如果作为判断和惩罚人们潜在行为的标准,也是有悖伦理道德的。因此,科学研究在使用大数据时应该有度。

5 大数据技术对人文—经济地理学研究内容和时空尺度的影响

大数据的基本核心研究内容就是数据,包括数据获取、组织管理和分析等等各种技术。这一点与人文—经济地理学所关注的社会经济活动既有联系,又有所不同。数据的发出者是人,因此大数据

直接推动了人文—经济地理学对人的关注,如人的社交网络、交通行为等。这样,大数据技术的发展客观上推动了人文—经济地理学对人的研究转向,使得研究更加关注人的活动,如游客满意度分析等(Xiang et al, 2015)。但需要注意的是,数据本身对于“人”的特征的记录很少,因此,大数据仍然不能完全替代人文—经济地理学的数据调查方法来达到对人的认识。

其次,研究数据的丰富和精准化使得人文—经济地理学将面临尺度收缩问题。人文—经济地理学关注既定空间的社会经济活动。空间在研究过程中经过了高度的概括,如城市、区域或国家尺度等,相应地形成了不同的学科体系。城市研究只关注城市内部空间,区域研究将城市高度抽象为点,而国家或全球尺度研究则将国家或者区域抽象成点。这样,人文—经济地理学实际上存在空间尺度和时间尺度收缩问题(Scaling Issue)。通俗地讲,就是存在“只见森林,不见树木”或是“只见树木,不见森林”的情况。比如区域交通流分析往往是城市间或者省区间被加总的的数据,个体数据(特别是一些异常值或溢出值)被过滤掉了,或是省区间交通流向的数据无法进一步细化到城市与城市间、城市与乡村间。大数据可以从多角度、多维度对一个具体的研究空间对象进行分析,从整体(群体)和个体两个层面对研究对象进行深入。在这一情况下,随着空间尺度的缩放,事物特征逐级展现出来。时间尺度上,人文—经济地理学研究,如交通流量,往往用一段时间的整体情况来进行分析,无法反映出时间片段,如季节性、周期性变化等。所以同样地,时间尺度的缩放可以帮助研究者更加准确地理解事物特征。

随着尺度缩放问题的解决,一些“细化”的问题得到揭示。这些“细化”问题可能会具有重要的研究价值。如社会经济活动中的旅游活动,往往存在季节性特征;两个城区间交通总量一致的情况下,也可能存在明显的城市与城市、城市与乡村之间的差异,这些都构成了人文—经济地理学研究地区分异规律的重要内容。一些溢出值或异常值也可能在某些情况下具有重要的研究意义或者研究价值,反映了有别于总体特征规律的特殊情况。随着大数据技术的推进,人文—经济地理学可能要同时对时空尺度缩放所产生的问题进行细致深入的研究。

另外,大数据大大推进了人文地理学的跨界研究。与以往的数据收集不同,很多数据的收集以人为核心,采集了社会和自然的信息,反映了人与自

然、人与环境的相互关系,如健康、污染暴露等。从这一点来说,大数据有望推进人文地理学的跨界研究,更加深入地研究人地关系/人与自然的关系(Human-Nature Relationship) (Crain et al, 2014)。如Google公司利用网页访问的大数据,比美国疾病中心提前约一周的时间预测了传染病在美国的传播路径(Ginsberg et al, 2009);有学者利用大数据对环境变化检测和人类健康影响做出了评价等(Vitolo et al, 2015)。大数据的采集可能对新的研究话题,或是过去受到数据收集限制的话题做出积极的贡献,如地缘政治的研究。

因此,大数据对人文—经济地理学的研究内容的深化起到了重要的作用。对于“人”的关注,时空尺度研究和人文—自然综合研究等都是人文—经济地理学近年来新的重要研究方向。人文—经济地理学和大数据技术应该在这些方面进行更加深入的交叉和互补。

6 科学问题和实践应用的相互促进

人文—经济地理学大数据技术发展的土壤更多地来源于实践需求。一个重要的事例就是在大数据的基础上,结合云计算和物联网技术发展的智慧城市(Batty, 2012)。智慧城市是在高度信息化状态下,充分利用大数据和各项技术对未来城市建设的新尝试。它试图从城市尺度(包括社区、主要功能区和城市其他尺度),利用现代信息技术,对现代城市发展模式进行改进,逐渐从目前的经济领域开始向人们的日常生活和空间布局推进(Allwinkle et al, 2011)。其内容包括民生、环保、公共安全、城市服务、工商业活动等方面(谢耘, 2012)。大数据将充分运用其信息和通信技术手段感测、分析、整合城市运行核心系统的各项关键信息,但如何与人类智慧型发展所需的各种条件,以及如何与城市政策决策过程有效衔接起来,目前还处于研究的起步阶段(Abb et al, 2012; Anthopoulos et al, 2012)。由于技术方式、产业内容、地方功能的重要转变,智慧城市和大数据应用将对城市规划产生重要影响。

人文—经济地理学同样与社会实践紧密相连,但更多地是对科学问题的研究。科学研究与实践可以相互促进。例如,人文—经济地理学在研究中力求建立一种参与式(Participatory Approach)的研究方法,这一研究方法通常在研究者的主导下,通过访谈的形式来完成。随着访谈的难易变化,参与

者的数量和参与次数可能会受到一定的限制,调查结果对真实情况的反映存在局限。另外,在城市规划中,规划的制定很大程度上依赖于政府官员对城市发展的定位,以及规划者和研究者对城市的认识和判断,规划和研究制定缺乏公众的参与(王广斌等, 2013)。这样规划的结果往往就是重视空间物质形态表现,缺少人文关怀;重视经济发展,缺少人居环境考虑,规划的制定和执行与当前提倡的以人为本和精细化管理要求相去甚远。大数据的兴起为参与式研究和规划搭建了新的平台。通过互联网公共平台建设,大数据可以轻松地将研究者的研究问卷或问题提交在平台上,公众可以对这些问题作出回答甚至是修改。同时,政府工作也将更加透明,一方面受公众监督,需要快捷地对居民需求作出反应;另一方面利用丰富的社会资源降低管理成本,提高管理效率。香港尽管高楼林立,但仍有70%的面积是绿地,其中有40%是政府保护性绿地,但在经济的刺激下,非法侵占绿地情况时有发生,给政府监管带来很大的麻烦。公共参与式管理为这一问题提供了解决途径。2005-2009年,关于侵占绿地20485起投诉中,有93%来自于公众,形成了特殊意义上“港人治港”的“市民地图”。

或许大数据和人文—经济地理学在针对社会经济活动中具有不同的分工,二者在实践应用和科学问题研究中不可相互替代。大数据侧重于数据的收集、分析与加工过程,虽然对揭示问题具有一定的帮助,但不能完全替代科学工作者对科学问题的认知和把握。另外,大数据缺少了决策分析这一重要环节(Analytis et al, 2014),因此在数据分析之后,常常需要研究者结合其他情况进行更加综合的分析和判断。

7 结论:以大数据来推进人文—经济地理学发展

信息技术正在成为社会发展和建设的主要关键性动力,深刻地影响和改变着今后的社会发展。人文—经济地理学及相关学者从研究到实践都积极参与其中,适应并积极响应这一新趋势带来的机遇和挑战。目前大数据技术刚刚崭露头角,因此本文探讨的许多问题还都没有最终答案,希望将来会有更多、更深入的思考。另外,鉴于在大数据技术的应用和探索中,中国基本处于与国际同步阶段,所以也希望中国将来通过研究的深化,在这一方面

能够处于国际人文地理学界领先地位。

7.1 充分利用新的数据源,完善学科数据建设

数据(包括定量和定性数据)制约着人文—经济地理学研究的深度和广度(宋长青等, 2004)。大数据为人文—经济地理学提供了新的重要数据获取方式,在数据的空间和时间精度上进行了很大的推进,从而使得人文地理学在广度(新内容)和深度(研究尺度,研究问题的深入)的发展上成为可能。随着数据类型的丰富和数据源的多样化,特别是长时间、多尺度、个体记录在大数据技术上的初现端倪,数据建设将成为人文—经济地理学中一个重要问题。人文—经济地理学者从被动地接受二手数据变为主动地获取和持续记录自己感兴趣的研究数据,并按照研究内容和研究目的设计数据类别和属性,获取数据的方式发生了重大变化。

在数据库的建设中,数据(源)之间的可比性成为比较研究的关键。如Facebook在中国和俄罗斯都不是主流社交媒体工具。受文化、制度和技术发展水平的影响,各地大数据技术应用等都会有所不同,发达与落后国家和地区会有形成“数字鸿沟”的潜在风险。

数据的重要性,潜在的一个疑问就是,对于研究者来说是否存在“数据为王”的现象?近期由于受数据获取性的限制,我们认为短期内会存在这一情况。但长期来看,由于开源数据的发展趋势,以及大数据本身需要与理论和应用的结合,思想的高度最终会决定大数据的命运和研究者的成败。在这一方面,韩国政府正在推进面向研究和政策制定者、通过物联网技术从环境中提取大数据,从而推动以大数据为支撑的绿色IT发展(Baek et al, 2015)。

7.2 建立大数据应用较为完善的研究方法体系

大数据可能是除定量和定性之外的第三大类研究方法^①(Delyser et al, 2003; Gray et al, 2007)。一方面,它将定量数据的收集发挥到极致;另一方面对于人的感觉、情感、经验、体验、信仰、价值、思想和创造性也具有极强的数据收集和分析能力(如舆情)。大数据强调了数据密集型科学研究(Data-extensive Inquiries)并有效地利用已经建立的多种计算分析工具,从数据的角度揭示问题。从数据类型的广度、丰度、深度(时间和空间尺度)来说,大数据有望成为人文—经济地理学中第三类研究方法。

当然建立这一研究方法并不是一蹴而就的。

一是大数据本身需要规范化;二是从适用范围、分析方法和模型等方面还需建立起一套完整的方法。目前对大数据的应用需要十分谨慎。由于研究个体特征等诸多问题,大数据分析结果的代表性尚存在很多问题。在有限条件下,回溯模拟大数据分析结果与其他数据源结果的相关性,在此基础上进行更大程度的推广可能是一个办法和途径。如研究者曾利用Google公司页面访问查询,对2003-2008年美国流行病传播进行分析,结果表明这一指标和美国疾病预防控制中心的检测结果间相关系数高达0.97(Ginsberg et al, 2009)。

7.3 促进跨域数据整合和跨域研究

人文—经济地理学和大数据技术属于不同域(Domain),目前很多最新的研究进展并不是由人文—经济地理学者完成的。技术创造者会受到应用性和研究目标的困扰,人文—经济地理学者将面对数据收集和分析的问题,这就要求人文—经济地理学者与大数据技术人员间进行通力合作。

另外,人文—经济地理学者也可以预期,未来大数据除了记录人类社会经济活动外,还将纳入以往的自然过程(如气候变化)数据。如何整合这些跨域数据(Cross-domain),以及融合社会经济发展—自然过程信息处理技术成为研究人地关系中新的技术难点和重点。

7.4 推进研究对象和研究目的的转变

人和人类社会经济活动作用于地表,构成社会经济变化的重要内容。客观上,人文—经济地理学,特别是中国的人文—经济地理学更多地关注后者。随着大数据技术对人无时无刻的记录和参与式方法平台的夯实,以及人文关怀的日益重要,以人为主要研究对象和以提高人的生活质量为研究目的将是未来人文地理学的重要方向。当然,这一转变还需大数据技术本身对调查对象的属性进行更有力、更明确的界定说明,同时也离不开与传统调查法的相互验证。但可以相信,在大数据的推动下,人文—经济地理学研究的应用性也将会有力地推进社会管理过程中的公众参与度,公共权力的透明度与城市决策过程的公开化。

参考文献(References)

- 贺灿飞, 郭琪, 马妍, 等. 2014. 西方经济地理学研究进展[J]. 地理学报, 69(8): 1207-1223. [He C F, Guo Q, Ma Y, et al. 2014. Progress of economic geography in the West: a literature review[J]. Acta Geographica Sinica, 69(8):

^①有学者将之称为第四研究范式(fourth paradigm),或偶然出现的eScience。

- 1207-1223.]
- 刘卫东, 陆大道. 2004. 经济地理学研究进展[J]. 中国科学院院刊, 19(1): 35-39. [Liu W D, Lu D D. 2004. The development of economic geography in China: progress and suggestions[J]. Bulletin of the Chinese Academy of Sciences, 19(1): 35-39.]
- 龙瀛, 张宇, 崔承印. 2012. 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, 67(10): 1339-1352. [Long Y, Zhang Y, Cui C Y. 2012. Identifying commuting pattern of Beijing using bus smart card data[J]. Acta Geographica Sinica, 67(10): 1339-1352.]
- 苗长虹, 魏也华. 2007. 西方经济地理学理论建构的发展与论争[J]. 地理研究, 26(6): 1233-1246. [Miao C H, Wei Y H. 2007. Development and debates in the theoretical construction of economic geography[J]. Geographical Research, 26(6): 1233-1246.]
- 秦萧, 甄峰, 熊丽芳, 等. 2013. 大数据时代城市时空行为研究方法[J]. 地理科学进展, 32(9): 1352-1451. [Qin X, Zhen F, Xiong L F, et al. 2013. Methods in urban temporal and spatial behavior research in the big data era[J]. Progress in Geography, 32(9): 1352-1451.]
- 宋长青, 冷疏影. 2004. 国家自然科学基金资助下的中国人文地理学发展[J]. 地理学报, 59(增刊1): 8-10. [Song C Q, Leng S Y. 2004. Features of recent human geography researches granted by National Natural Science Foundation of China[J]. Acta Geographica Sinica, 59(S1): 8-10.]
- 王广斌, 范美燕, 王捷, 等. 2013. “智慧城市”背景下的城市规划创新[J]. 上海城市规划, (2): 11-14. [Wang G B, Fan M Y, Wang J, et al. 2013. Innovation of urban planning against the background of 'smart city'[J]. Shanghai Urban Planning Review, (2): 11-14.]
- 谢耘. 2012. 智慧城市的拼图模式[J/OL]. 2012-10[2015-03-25]. 智慧城市在中国, 试刊号: 30-33. <http://www.digitalchina.com/zhszczg/zhszczg.pdf>. [Xie Y. 2012. Zhihui chengshi de pintu moshi[J/OL]. 2012-10[2015-03-25]. Smart City in China, Prototype Ed.: 30-33. <http://www.digitalchina.com/zhszczg/zhszczg.pdf>.
- 杨振山, 蔡建明. 2010. 空间统计学进展及其在经济地理研究中的应用[J]. 地理科学进展, 29(6): 757-768. [Yang Z S, Cai J M. 2010. Progress of spatial statistics and its application in economic geography[J]. Progress in Geography, 29(6): 757-768.]
- 甄峰, 秦萧, 王波. 2014. 大数据时代的人文地理研究与应用实践[J]. 人文地理, 29(3): 1-6. [Zhen F, Qin X, Wang B. 2014. Human geography research and practical application in big data era [J]. Human Geography, 29(3): 1-6.]
- ABB, European House-Ambrosetti. 2012. Smart cities in Italy: an opportunity in the spirit of the renaissance for a new quality of life[R/OL]. [http://www02.abb.com/db/db0003/db002698.nsf/0/af0569d84fa29925c1257a71004dc253/\\$file/SmartCitiesReport10Points.pdf](http://www02.abb.com/db/db0003/db002698.nsf/0/af0569d84fa29925c1257a71004dc253/$file/SmartCitiesReport10Points.pdf).
- Allwinkle S, Cruickshank P. 2011. Creating smarter cities: an overview[J]. Journal of Urban Technology, 18(2): 1-16.
- Analytis P P, Moussaid M, Artinger F, et al. 2014. "Big data" needs an analysis of decision processes[J]. Behavioral and Brain Sciences, 37(1): 76-78.
- Anthopoulos L, Vakali A. 2012. Urban Planning and smart cities: interrelations and reciprocities[M]. Berlin, Germany: Springer.
- Baek H, Park S-K. 2015. Sustainable development plan for Korea through expansion of green IT: policy issues for the effective utilization of big data[J]. Sustainability, 7(2): 1308-1328.
- Barabasi A L. 2010. The hidden pattern behind everything we do[M]. Boston, MA: Dutton Books.
- Batty M. 2012. Smart cities, big data[J]. Environment and Planning B, 39(2): 191-193.
- Bentley R A, O'Brien M J, Brock W A. 2014. Mapping collective behavior in the big-data era[J]. Behavioral and Brain Sciences, 37(1): 63-76.
- Boyd D, Crawford K. 2012. Critical questions for big data[J]. Information, Communication & Society, 15(5): 662-679.
- Cai H, Jia X, Chiu A S F, et al. 2014. Siting public electric vehicle charging stations in Beijing using big-data informed travel patterns of the taxi fleet[J]. Transportation Research: Transport and Environment, 33: 39-46.
- Crain R, Cooper C, Dickinson J L. 2014. Citizen science: a tool for integrating studies of human and natural systems [J]. Annual Review of Environment and Resources, 39: 641-665.
- Delyser D, Sui D. 2013. Crossing the qualitative- quantitative divide II: inventive approaches to big data, mobile methods, and rhythm analysis[J]. Progress in Human Geography, 37(2): 293-305.
- Ginsberg J, Mohebbi M H, Patel R S, et al. 2009. Detecting influenza epidemics using search engine query data[J]. Nature, 457: 1012-1014.
- Graham G, Mehmood R. 2014. The strategic prototype "crime-sourcing" and the science/science fiction behind it[J]. Technological Forecasting and Social Change, 84: 86-92.
- Gray J, Szalay E. 2007. Escience: a transformed scientific method[J/OL]. 2007-01-11[2015-03-25]. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_jim_gray_transcript.pdf.
- Jiang B. 2015. Head/tail breaks for visualization of city structure and dynamics[J]. Cities, 43: 69-77.
- Jiang L, Deng X, Seto K C. 2012. Multi-level modeling of urban expansion and cultivated land conversion for urban hotspot counties in China[J]. Landscape and Urban Planning, 108(2-4): 131-139.
- Jun S P, Park D H, Yeom J. 2014. The possibility of using search traffic information to explore consumer product attitudes and forecast consumer preference[J]. Technological Forecasting and Social Change, 86: 237-253.

- Laney D. 2001. 3D data management: controlling data volume, velocity, and variety[R]. Application Delivery Strategies, META Group.
- Long Y, Han H, Tu Y, et al. 2014. Evaluating the effectiveness of urban growth boundaries using human mobility and activity records[R/OL]. 2014-01-04[2015-03-25]. <http://www.beijingcitylab.com/working-papers-1/wp41-60/>.
- Long Y, Liu X. 2013. How mixed is Beijing, China: a visual exploration of mixed land use[J]. *Environment and Planning A*, 45(12): 2797-2798.
- Long Y, Shen Y. 2014. Mapping parcel-level urban areas for a large geographical area[R]. ArXiv preprint arXiv, 2014a, 1403.5864.
- Mayer-Schonberger V, Cukier K. 2013. Big data: a revolution that will transform how we live, work and think[M]. London, UK: John Murray Publishers.
- Nelson A C, Duncan J B. 1995. Growth management principles and practices[M]. Chicago, IL: Planners Press.
- Tan K C, Lim H S, Matjafri M Z, et al. 2010. Landsat data to evaluate urban expansion and determine land use/land cover changes in Penang Island, Malaysia[J]. *Environmental Earth Sciences*, 60(7): 1509-1521.
- Vitolo C, Elkhatab Y, Reusser D, et al. 2015. Web technologies for environmental big data[J]. *Environmental Modelling & Software*, 63: 185-198.
- Wang H, Kilmartin L. 2014. Comparing rural and urban social and economic behavior in Uganda: insights from mobile voice service usage[J]. *Journal of Urban Technology*, 21(2): 61-89.
- Xiang Z, Schwartz Z, Gerdes J H, et al. 2015. What can big data and text analytics tell us about hotel guest experience and satisfaction[J]. *International Journal of Hospitality Management*, 44: 120-130.
- Zhang Z, Zhou T, Zhang Y. 2011. Tag-aware recommendersystems: a state-of-the-art survey[J]. *Journal of Computer Science and Technology*, 26(5): 767-777.

Opportunities and limitations of big data applications to human and economic geography: the state of the art

YANG Zhenshan^{1*}, LONG Ying², Nicolas DOUAY³

(1. Key Lab of Regional Sustainable Development and Modelling, Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China; 2. Beijing Institute of City Planning, Beijing 100045, China; 3. Université Paris Diderot, Paris 75205, France)

Abstract: The technology of "big data" has profoundly changed our life and society, and advanced scientific research. By taking social and human activities as main data source, this technology is of great potential of applications in human and economic geography. Drawing on recent progress in research, this article analyzes the new applications of big data to the research of urban hotspots, functional areas and boundaries, transportation and consumption behaviors and social geography. Based on these analyses, this article articulates the roles of big data in enriching data sources, adding new research themes, bringing new research paradigms, and stimulating the research of coupling to human-spatial research in human and economic geography. However, the technology of "big data" still needs improved, especially the "bias" issue in collection and the attributes of data. It also needs appropriately positioned in the application in human and economic geography because big data cannot replace the data that are collected from field work, or be applied without proper theoretical grounding and hypothesis, and replace the independent thinking of researchers and decision processes. These factors limit the application of big data, which requires more efforts on big data infrastructure development as well as exploration of human and economic geography. Acknowledging the opportunities and roles of big data application, human and economic geographers should emphasize the following to advance the research of this field: exploring new data sources and paying closer attention to database construction in human and economic geography, establishing a research paradigm towards big data applications, facilitating cross-disciplinary and cross-domain research to strengthen the study of human-nature relations, and emphasizing the research towards human behaviors and demands.

Key words: big data; human and economic geography; research paradigms