

大数据时代的空间分析

张晓祥¹

1 河海大学地理信息科学与工程研究所,江苏 南京,210098

摘要:近年来,大数据越来越多地被用来描述正在到来的信息爆炸的时代,也引起了产学研政府部门的高度重视。首先综述了大数据和空间大数据的特点和最新发展。其次,从GIS基础理论方法的角度回顾和评述了空间分析、空间建模和空间优化等三个地理信息科学的核心议题。在此基础上,引申出空间理论方法背后隐含的空间哲学和空间思维。最后指出了大数据时代地理信息科学发展的机遇与挑战。

关键词:大数据;空间分析;空间建模;空间优化;空间哲学

中图法分类号:P208

文献标志码:A

1 大数据与空间大数据

大数据(big data),简单来说就是指数据规模巨大到无法通过人工进行处理和解读的数据。除了数据量大以外,大数据还有诸如数据处理速度迅速、数据来源和构成复杂以及数据真实感强烈等特点,同时大数据还需要有很高的使用价值^[1]。大数据出现的同时,一定程度上解决了过去长期存在的数据源短缺问题,从而使得以前一些因为数据短缺而无法进行的分析计算变得可行。近年来,急剧增长的地理空间数据已经成为大数据流的重要组成部分,空间大数据的发展将会有效化解地理信息领域长期存在的数据瓶颈问题。

在空间数据处理速度上,伴随着高性能计算机等硬件技术的发展以及云计算和移动计算等软件技术的发展,GIS的高性能计算的发展主要分两种思路:①注重采用硬件加速方法进行复杂图形计算,如GPU图形处理技术^[2];②注重采用软件方法来实现复杂的计算分析功能,如赛博GIS技术(CyberGIS)^[1,3]。

在多源异构的数据集成上面,诸如图论(graph)、语义(semantics)和本体(ontology)等新的技术方法已经逐步应用于空间数据的自动提取和集成^[4],这些方法特别适合于非SQL的数据结构,诸如Hadoop等可以对大数据进行分布式处理的软件框架。大数据时代,常规的基于SQL的

元数据的手工提取和解译已经变得越来越不可行,图论方法则可以利用动态图进行数据模型的内部比较和数据集成;同时,在大数据时代,越来越要求数据和模型的语义注释要能够被人 and 机器所理解,空间语义研究可以使得时空数据库易于发布、提取、探索和整合;而本体方法则可以针对深入的行业应用建立各个应用行业领域中的不同尺度与粒度的对象、过程和关系,这是信息集成的基础。

在数据的真实感方面,实时快速的社会媒体(social media)等志愿者地理信息(volunteered geographic information, VGI)以及物联网数据增强了信息的现场感和真实感。在空间大数据上,除了政府、机构提供的数据以外,国外Facebook、Twitter以及国内的微博、微信等直接来自用户个体的社会媒体数据作为一类重要的与空间位置相关的信息也得到快速发展^[5-6]。

地理信息系统是一种非常有效的时空数据组织方法,空间和时间可用于基本的知识组织,并且这种组织方法容易消除信息歧义。良好建构的空间大数据无疑有着巨大的应用价值,空间大数据与云计算、移动计算、物联网、社会化媒体技术CyberGIS结合,可以用于建立下一代地理空间信息基础设施,用于空间数据集成和空间知识传播。同时,这种数据基础设施还可按时按需提供空间信息和知识,通过高性能计算进行各类常规和应急任务的技术支撑^[7-9]。一些研究还基于空间大

收稿日期:2014-02-26

项目来源:国家自然科学基金资助项目(41201394,41201117)。

第一作者:张晓祥,博士,副教授。主要从事空间分析与建模研究。E-mail: xiaoxiang@hhu.edu.cn

数据整合多样的地理信息资源建立了空间信息交换与处理平台——地理空间服务网,聚焦于建立数据、信息和知识共享与交换的平台,进行异构信息资源的整合与互操作,实现空间数据、地理信息和地学知识的网络共享与在线服务^[10]。

2 空间分析、建模与优化

1980s,美国国家基金委充分认识到 GIS 空间分析方法的巨大潜力及其与非空间分析的巨大差异,选择美国加州大学圣芭芭拉分校、纽约州立大学布法罗分校、缅因大学建立了三个国家地理信息分析中心来进行地理信息系统的空间分析基础研究。当时,三个中心在全世界范围内招聘到一批计量地理、区域科学方面的著名教授,由此展开了 1980 年代末至 2000 年的地理信息系统研究的黄金时代,地理信息系统也由单纯的技术驱动而上升为地理信息科学^[11]。Haining 教授认为,地理信息科学(GIScience)=GIS 系统软件+空间分析^[12]。从中也可以看出地理信息科学的一大特点就是空间分析,通过空间分析,GIS 可以超越单纯的地图制图和数据库管理,最终发展成为可以进行分析决策的工具。

大数据时代,地理信息系统数据极大丰富,数据已经成为一种类似水、电、气等“立等可取”的基础设施。研究人员也会发现,在 GIS 中集成空间分析程序将会有助于揭示空间分析的一些基本问题。但是,作为应用的空间分析本身却越来越多地误用和滥用^[13],因此有必要对空间分析本身进行深入的剖析。空间分析主要研究空间格局、空间位置、空间行为、空间关系、空间过程,广义的空间分析包括空间分析、空间建模、空间优化三部分^[14]。

2.1 空间分析

本节的空间分析是指狭义的空间分析,也称为空间数据分析和空间统计分析,其中最经典的就是统计方法在空间数据分析上的应用^[15]。由于空间数据通常不满足统计学上的样本独立性假设,因此空间分析与非空间分析之间存在显著差异,空间分析方法并不能简单地套用非空间分析方法。与非空间分析方法相比,空间分析有着显著的特点,即空间信息还有位置、距离、方向、邻接性、尺度等多种特征。空间分析通常相对于非空间分析显得模糊而不够精确,不同的人可能会得到完全不同的空间分析结果。

为了改变空间分析的状况,研究人员通过研

发空间分析软件产品来进行空间分析的标准化工作。以 GIS 软件 ArcGIS 9.3 版本为例,在该版本的软件工具箱中,就包括 3D 分析、地统计分析、空间统计等近 20 大类 600 多种空间分析操作,现在 GIS 已经可以非常方便地进行空间内插、空间采样、空间相关性分析和空间异质性分析等。在应用过程中,类似统计软件容易导致误用一样,空间分析软件同样会导致一些误用,一个经典的例子是在空间分析中要特别注明由于尺度问题导致的 MAUP(modifiable area unit problem,可塑性面积单元)问题^[13]。

可以想象,未来还会有更多的空间分析功能会被开发成软件模块,以便自如地进行空间点、线、面、格网数据的分析和计算。大数据时代,空间分析的应用方式也会有很大改变。如面对 PB 级别的大数据,数据本身无法全部下载到本地进行分析处理,因此需要发展一种“数据不动软件动”的处理方式进行在线空间分析。

2.2 空间建模

空间建模也称为空间过程建模。如果说空间(统计)分析是一种“黑箱”分析,那么空间建模则是一种“灰箱”分析,甚至是一种“白箱”分析。也就是说,空间建模是在了解或者部分了解空间过程“机理”的基础上建立的数学模型或数值模型,这种“过程”很多时候也是一种物理过程。在进行物理过程的空间建模时,考虑到计算的复杂性,需要考虑模型的精细程度,即建模过程没有最优解,只有近似解。GIS 软件目前已经发展成为一个功能强大的软件包,但是经典 GIS 的一个显而易见的问题是“GIS 缺乏专用的过程模型来控制系统中地理对象的状态、调控以及演化”^[16]。

1990s,美国加州大学圣巴巴拉分校的国家空间分析中心在 Goodchild 教授的领导下,于 1991、1993、1996、2000 年分别举办了 4 次 GIS 与环境建模国际研讨会,组织了多个行业领域的一大批重要专家进行 GIS 与各类专业空间过程建模的集成研究,后面三次会议还专门编辑出版了专著^[17-19],在国际上推动了这个领域的研究。1990s 末,GIS 对象-关系数据库技术和面向对象建模的方法逐渐成熟,在新的关系数据库基础上进行复杂的专业模型的建模逐步提上日程,各种行业空间建模软件的开发大量涌现,基本上都是基于 GIS 进行专业模型的软件模块开发。以水利行业为例,丹麦 Mike 系列软件、荷兰 Delft 3D、美国 ArcSWAT 等专业软件功能都是构建 ESRI ArcGIS 软件的面向对象扩展模块。

物理过程建模和空间建模理念的提出是 GIS 领域很重要的一个问题。现在提的时空 GIS 问题实际上是一种动态的时空 GIS 问题,是相对于以前的静态或准静态 GIS 而言的。国土、规划领域也做时空 GIS,其实很多时候是一种静态的 GIS 时空问题,但水、空气等流体的物理现象变化比较快,动态性强,原有的基于图层管理的时空构架难以适应这种空间建模的需要。空间过程模型有很多种类,除了常见的水、气的流动,还可以用于传染病的传播、技术扩散、空间交互建模以及随机模拟等研究。

过去,空间数据的收集工作量巨大,人们总是通过减少模型复杂度来尽量使用较少的参数和数据进行预测建模,这从一定程度上影响了模型的精度。大数据的出现为复杂空间过程建模的深入研究提供了可能,特别是对一些严重依赖背景信息的迭代计算模型,实时的数据获取和实时的建模计算为这种问题提供了解决方法。另外,大数据还可以聚合展现空间对象在复杂模型中的具体行为,从微观尺度上弥补宏观模型构建上的错误和不足。

2.3 空间优化

如果说,空间分析可以描述空间位置、空间关系、空间距离、空间梯度、空间格局等解释性信息;空间建模可以进行对历史的精确反演、对未来的情境分析,那么空间优化则可以对目前的现状进行评价、调控和决策。

空间优化实际上与地理学有着千丝万缕的联系,在经典的经济地理学研究中,有着诸如 von Thunen 地租关系理论(农业区位论)、Weber 的耗费最小化(工业区位论)、Christaller and Losch 的中心地理论等基础地理学理论。这些区位论研究通常需要应用空间优化方法,以解释和理解地理格局,并设定最佳位置。空间优化建立在丰富的学科传统和现实意义基础上,特别是地理学的计量革命以外,空间优化逐渐发展成为地理学的一个重要分支^[14]。

空间优化问题包括三个组成部分:目标、所要作出的决策、约束条件。这三者在某种程度上有时混杂在一起,或明确或隐含地反映感兴趣的地理问题,具体求解可以采用精确解或近似解。研究目标通常与问题的背景相关,需要满足一定的限制条件,如要确保费用最小或收益最大等。各类应用问题的目标通常以一个或多个目标函数来表达。决策变量与决策本身密切相关,需要根据实际问题确定参加求解的决策变量。空间约束实

际上把空间关系和空间属性作为限制条件加入原有的优化方程,以共同组成方程组求解^[20]。

大数据时代,数据的精度越来越高,空间数据也越来越丰富,这对于复杂优化问题的求解是一个难得的机遇。优化问题的建模并不是一个简单的工作,尤其是在空间背景下的建模更为困难。对于一些问题,限制条件可以有多种,空间优化问题本身可能更为复杂。空间优化并不总能得到精确解,有时候近似解也是可以接受的,因而启发式算法在空间优化领域有着广泛的应用。在实际应用中,空间优化对于交通建模、位置建模、商业地理学、医学地理学、政治地理学、学区划分、土地利用规划、城市规划、经济社会发展规划等有着重要的应用价值,可用于区域空间综合评价、资源空间优化配置、突发事件优化管理以及空间决策支持等^[14, 20]。

3 空间哲学和空间思维

3.1 空间哲学

美国加州大学圣芭芭拉分校地理系的 Simonett 教授提出的 Simonett 魔方对该校地理系的崛起有重要意义^[21]。

Simonett 教授特别注重以定量方法来研究地理问题,打破过去地理系自然地理、人文地理的学科领域的限制来招聘教授,UCSB 地理系当时招聘的每一个教授都必须要有进行学科交叉研究的能力,即必须在 Simonett 魔方里找到自己的三维坐标(X, Y, Z),也即必须有地理学基础、专业建模能力和量化分析能力,缺一不可。魔方的一侧表示以空间分析、遥感、计算机制图学、地理信息系统为核心的空间分析与建模技术;魔方的另一侧表示各类专业建模领域,如海洋建模、水文学建模、大气建模、植被/土壤建模、行为地理学、区域/城市建模、自然资源管理等;两者之间的连接则通过统计建模、数学建模、数值建模来实现。

哲学作为关于世界观的学说,是理论化、系统化的世界观,是自然知识、社会知识、思维知识的概括和总结,是世界观和方法论的统一。空间研究同样需要坚持一定的空间哲学,Simonett 魔方是经过实践检验的空间哲学,到目前为止还有重要的学术意义和研究价值。

大数据时代,空间数据的存储、管理问题得到了良好的解决,高性能计算也使用户感觉不到计算的存在。在空间应用研究中,需要特别注意把握用空间明确的模型(spatially explicit models)来理解人与环境的交互,这从根本上说是一个哲

学问题。在这种空间哲学中,对于自然界来说,就是用物理公式来描述自然过程;对于人类社会来说,就是懂得如何就地球资源的使用进行决策。但是这些自然界的规律与现实社会的运行方式对人类来说还有大量的未知领域。对于未知世界的探索,空间哲学的探讨和研究是首要的问题。

3.2 空间思维

Simonett 魔方实际上也是一种空间思考方式,也即无论是空间分析、空间建模还是空间优化,实际上贯穿着一种“空间思维”。在人类生活中,空间思维实际上是无处不在的,从幼儿园学习的简单的空间概念到大学里学习的抽象的空间表达,无不体现一种空间思维,这种空间思维突出地表现为空间知识、空间方法和空间工具的掌握和应用^[22]。从另一个意义上,空间思维实际上是基于地理数据和地理信息,运用空间分析、空间建模和空间优化等理论方法获取认识和改造现实世界的知识、智慧和能力。GIS技术的发展让空间思维越来越容易使用,Google Earth 的风靡实际上也是空间思维的胜利。

可以预计,大数据是未来十年空间分析研究的新的驱动力,空间分析将更多地依赖从复杂、多变的数据流中实时获取,更多的空间分析也将依赖高性能计算变得实时和准实时。一些研究人员甚至认为大数据是继试验研究、理论推导、计算仿真后的“第四种科学范式”^[23],将显著加快和促进科学、工程领域的创新步伐。在这种情景之下,作为科学研究本质的“空间思维”将显得尤其重要。

4 结 语

1990s 以来,以地理信息系统、遥感、全球定位系统为代表的空间信息技术获得飞速发展。当前,应对全球变化和实现可持续发展已经成为当今世界的两大重要议题,地理信息技术对于这两大议题的解决有着重要的应用前景。在全球变化和可持续发展研究中,涉及地理学、经济学、心理学、人类学、社会学等多个领域,人地关系也即人与自然的关系是其中的核心。

全球变化和可持续发展是针对地球和局部的系统建模,地球本身作为一个开放的复杂巨系统,涉及要素众多,时空变化剧烈,建模过程非常复杂。地理信息科学的一大优势就是可以在分析过程中集成不同领域专家的智慧,通过定量的空间分析、建模和优化,以期让各个行业的专家达成共识,最终做出决策。

大数据的应用将显著促进地理信息技术的发展和应用,空间大数据更多地需要处理数据的空间性、时间性、流动性、多元性等多种因素混合导致的大数据的复杂性,未来在通过计算、可视化和分析等方法进行减少数据维度的研究仍充满挑战。但更重要的是,作为地理信息科学自身,需要重视空间分析、空间建模和空间优化等基本理论方法,需要坚持类似 Simonett 魔方一样的空间哲学和空间思维。

参 考 文 献

- [1] Anselin L. From SpaceStat to CyberGIS: Twenty Years of Spatial Data Analysis Software[J]. *International Regional Science Review*, 2012, 35(2): 131-157
- [2] Yang C, Wu H, Huang Q, et al. Using Spatial Principles to Optimize Distributed Computing for Enabling the Physical Science Discoveries[J]. *The National Academy of Sciences*, 2011, 108(14): 5 498-5 503
- [3] Wang S A. CyberGIS Framework for the Synthesis of Cyberinfrastructure, GIS, and Spatial Analysis [J]. *Annals of the Association of American Geographers*, 2010, 100(3): 535-557
- [4] Brisaboa N R, Luaces M R, Places Á S, et al. Exploiting Geographic References of Documents in a Geographical Information Retrieval System Using an Ontology-Based Index[J]. *Geoinformatica*, 2010, 14(3): 307-331
- [5] Goodchild M F. Citizens as Sensors: The World of Volunteered Geography[J]. *GeoJournal*, 2007, 69(4): 211-221
- [6] Goodchild M F, Glennon J A. Crowd Sourcing Geographic Information for Disaster Response: A Research Frontier[J]. *International Journal of Digital Earth*, 2010 3(3): 231-241
- [7] Zhang T, Tsou M H. Developing a Grid-Enabled Spatial Web Portal for Internet GIServices and Geospatial Cyberinfrastructure[J]. *International Journal of Geographical Information Science*, 2009, 23(5): 605-630
- [8] Yang C, Raskin R, Goodchild M F, et al. Geospatial Cyberinfrastructure: Past, Present and Future [J]. *Computers, Environment and Urban Systems*, 2010, 34(4): 264-277
- [9] Wright D J, Wang S. The Emergence of Spatial Cyberinfrastructure[J]. *The National Academy of Sciences*, 2011, 108(14): 5 488-5 491
- [10] Gong Jianya, Geng Jing, Wu Huayi, et al. The

- Technology on Geospatial Service Web and Its Development[J]. *Journal of Geomatics Science and Technology*, 2013, 30(4): 353-360(龚健雅, 耿晶, 吴华意, 等. 地理信息资源网络服务技术及其发展[J]. 测绘科学技术学报, 2013, 30(4): 353-360)
- [11] Goodchild F M. Geographical Information Science [J]. *International Journal of Geographical Information Science*, 1992, 6(1): 31-45
- [12] Haining R P. Spatial Data Analysis: Theory and Practice [M]. Cambridge: Cambridge University Press, 2003
- [13] Fotheringham A S, Rogerson P A. GIS and Spatial Analytical Problems[J]. *International Journal of Geographical Information Systems*, 1993, 7(1): 3-19
- [14] Wang Jingfeng. Spatial Analysis [M]. Beijing: Science Press, 2006(王劲峰. 空间分析 [M]. 北京: 科学出版社, 2006)
- [15] Fotheringham A S, Brunson C M, Charlton M. Quantitative Geography: Perspectives on Spatial Data Analysis [M]. London: Sage Publication Ltd, Sage, 2000
- [16] Torrens P M. Process Models and Next-generation Geographic Information Technology [J]. *GIS Best Practices: Essays on Geography and GIS*, 2009, 31(2): 63-75
- [17] Goodchild M F, Parks B O, Steyaert L T. Environmental Modeling with GIS [M]. Oxford: Oxford University Press, 1993: 488
- [18] Goodchild M F, Steyaert L T, Parks B O, et al. GIS and Environmental Modeling: Progress and Research Issues [M]. Hoboken: John Wiley & Sons, 1996: 504
- [19] Clarke K C, Parks B E, Parks B O, et al. Geographic Information Systems and Environmental Modeling [M]. New Jersey: Prentice Hall Press, 2001: 306
- [20] Tong D, Murray A T. Spatial Optimization in Geography [J]. *Annals of the Association of American Geographers*, 2012, 102(6): 1 290-1 309
- [21] Keith C C, Baumgart S R. The Department of Geography at the University of California, Santa Barbara: History, Curriculum, and Pedagogy [J]. *Yearbook of the Association of Pacific Coast Geographers*, 2004, 66: 95-113
- [22] Goodchild M F, Janelle D G. Toward Critical Spatial Thinking in the Social Sciences and Humanities [J]. *GeoJournal*, 2010, 75(1): 3-13
- [23] Tony H, Stewart T, Kristin T. The Fourth Paradigm: Data-Intensive Scientific Discovery [EB/OL]. <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>, 2009

Spatial Analysis in the Era of Big Data

ZHANG Xiaoxiang¹

¹ Institute of Geographical Information Science and Engineering, Hohai University, Nanjing 210098, China

Abstract: In recent years, Big Data has been increasingly used to refer the coming era of information explosion. Industry, academia, research and government sectors all have paid more attention to big data. This paper provides a comprehensive review on the characteristics of and recent advances in big data and spatial big data. Specifically, three core concepts in GISciences, spatial analysis, spatial modeling and spatial optimization, are addressed and discussed from the theoretical and methodological perspectives. Further, the spatial philosophy and spatial thinking underlying the spatial theories and methodologies are investigated. This paper ends with a discussion of the opportunities and challenges for the GISciences in an era of big data.

Key words: big data; spatial analysis; spatial modeling; spatial optimization; spatial philosophy

Author: ZHANG Xiaoxiang, PhD, associate professor, specializes in spatial analysis and modeling. Email: xiaoxiang@hhu.edu.cn

Foundation support: The National Natural Science Foundation of China, Nos. 41201394, 41201117.