

从政策驱动到技术践行： 大数据开辟可持续发展研究新途径

*From policy-driven to practical implementation:
big data opens up a new research paradigm
for sustainable development*



周绮凤 (1976-), 女, 博士, 厦门大学自动化系副教授。2002年起从事数据挖掘及智能系统方面的研究工作, 2014-2015年在美国佛罗里达国际大学访学, 主要研究兴趣包括机器学习、数据挖掘及其在可持续发展等领域的应用。



李涛 (1975-), 男, 南京邮电大学计算机学院、软件学院院长, 南京邮电大学大数据研究院院长。2004年7月获美国罗彻斯特大学 (University of Rochester) 计算机科学博士学位, 2004-2014年先后任美国佛罗里达国际大学 (Florida International University) 计算机学院助理教授、副教授 (终身教授)、教授 (full professor)、研究生主管 (graduate program director)。由于在数据挖掘及应用领域成效显著的研究工作, 曾多次获得各种荣誉和奖励, 其中包括2006年美国国家自然科学基金委员会颁发的杰出青年教授奖, 2010年IBM大规模数据分析创新奖, 并于2009年获得美国佛罗里达国际大学最高学术研究奖。

中图分类号: TP399

文献标识码: A

doi: 10.11959/j.issn.2096-0271.2016013

1 可持续发展及面临的难题

可持续发展是世界对发展道路的审慎选择,也是我国的基本发展战略。随着全球资源枯竭、环境恶化的日益严重以及人类文明的演进和对生态保护理念的不断深化,可持续发展已经成为全球共识,对这一领域的研究也越来越热,日渐深入。

实现可持续发展的关键是如何制定合理的、能够平衡环境、经济和社会需求的复杂决策。然而自然、社会、经济系统本身的高度复杂性、动态性以及不确定性使得实现这一最优或近似最优的决策成为一个巨大的挑战。目前,可持续发展战略的实施仍然停留在政策驱动的层面,如何落实到技术实战,成为切实实现可持续发展的瓶颈问题。

2 可持续发展的新兴研究领域

近年来出现的计算可持续性(computational sustainability)是为解决可持续发展面临的挑战而出现的一个新兴的跨学科研究领域^[1,2],其目的是综合应用计算机科学、信息科学、运筹学、应用数学、统计学等多学科交叉技术来平衡环境、经济以及社会需求,以支持可持续发展的发展。计算可持续性研究涉及能源、生态、经济、环境等众多学科,汇集了计算领域和各种具有悠久传统的可持续性问题,如生态多样性、自然资源管理、生物与环境工程、资源经济学等。

计算可持续性研究的重点是针对可持续发展问题,开发计算模型、数学模型及相关方法,以帮助解决一些与可持续发展相关的最具挑战性的问题。计算可持

续性研究涉及面极其广泛,从野生动物保护、生物多样性到社会经济需求平衡、大规模环境布署以及再生能源的管理等都有涉及。如Cornell大学的计算可持续性研究机构(Institute for Computational Sustainability, ICS)采用计算的方法,对生态保护、物种迁移、人口分布等进行研究,构建数学模型并进行优化^[3,4]。

3 大数据助力可持续发展研究

大数据时代的来临为可持续发展研究带来新的机遇和挑战。卫星技术、传感技术的发展日新月异,每天可以采集到的各类环境数据无时无刻不在增加。大数据蕴含丰富的信息和潜在的知识,给人们研究可持续发展开辟了一个以数据为驱动的全新的研究方式,将极大地促进可持续发展研究^[5,6]。

目前,数据驱动的可持续发展研究已成为一个国际研究热点,各种会议正在持续热烈地举行。近年来,在人工智能(artificial intelligence, AI)、机器学习(machine learning, ML)等国际权威学术会议上,每年都有关于可持续发展的专题研讨。图1给出了《人工智能》杂志2014年“计算可持续性(Comp Sust)”专刊发表的可持续与人工智能相关的文章主题,由此可以看到,越来越多的人工智能、机器学习、数据挖掘等方法被应用到城市规划、物种分布、政策制定、健康、农业、交通、能源、智能电网等多种可持续性问题研究中^[7-9]。这些研究进展预示着大数据将成为可持续发展研究中的重要技术支撑,基于数据驱动的可持续发展问题研究是一个很有潜力的研究方向,为解决生态、环境、经济等诸多可持续发展问题提供以数据为驱动的解决途径,具有重要意义。

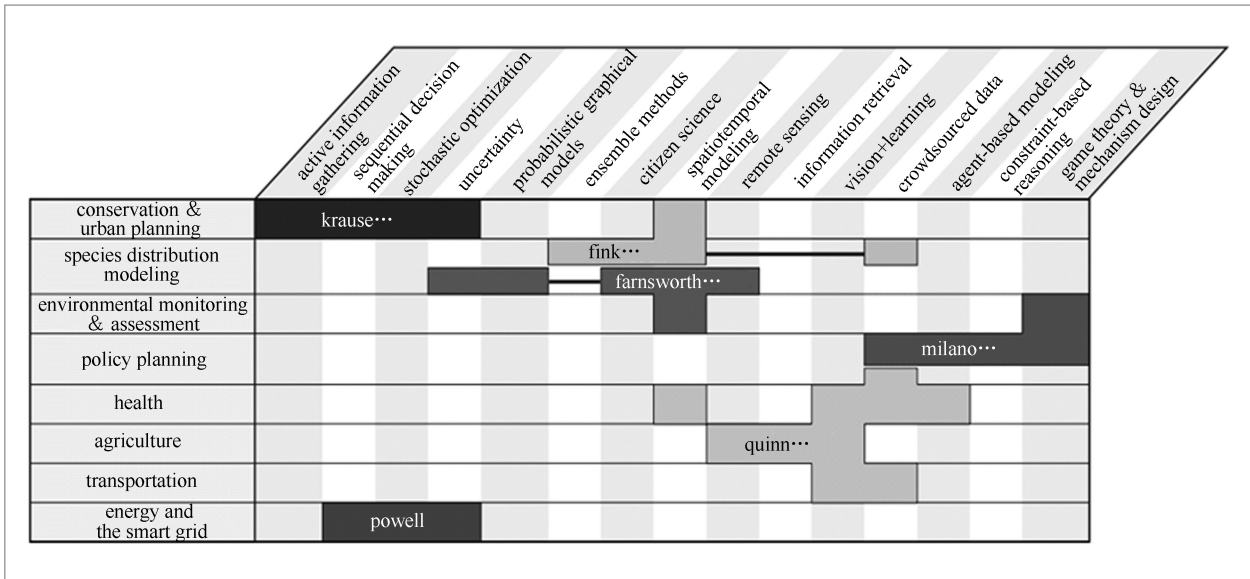


图1 2014年《人工智能》杂志 Comp Sust 专刊发表的计算可持续性相关研究主题¹

¹ <http://blog.computational-sustainability.org/>

4 基于大数据挖掘的计算可持续性研究新动向

在大数据时代，计算可持续性研究同样面临着新的机遇和挑战。一方面，大数据限制了研究者可以使用相对简单的分析技术，已有的构建和优化这些模型的方法遇到了可扩展性等挑战；另一方面，大数据所蕴含的丰富信息和潜在知识，将开辟一个以数据为驱动的全新的研究方式，可以帮助解决更加复杂、更大规模的计算可持续性研究问题。笔者所带领的研究团队，在国内率先开展基于大数据的计算可持续性研究，研究思路如图2所示。

究，研究思路如图2所示。

综合运用各种大数据处理、大数据分析以及大数据挖掘技术，解决大数据环境下计算可持续性面临的问题复杂性、计算效率、方法可扩展性等挑战，并结合可持续发展中亟待解决的焦点问题，开展可持续发展应用研究，整体研究框架如图3所示。

在该研究思路和框架下，认为大数据是应用、算法、数据和平台4个要素的有机结合，并通过理论、算法研究与实际问题、应用研究相结合，以平台建设为支撑，将可持续发展研究从现有的政策驱动真正落实到技术实施。下面以可持续发展中的一个焦点问题——建筑环境影响评价为例，简要阐述上述研究思路的具体实施^[10]。

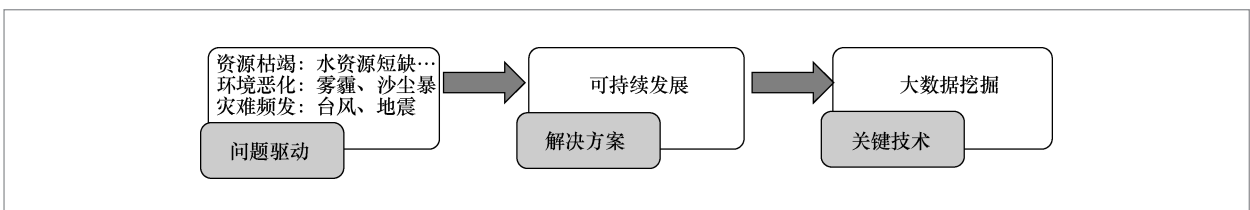


图2 基于大数据的可持续发展研究思路

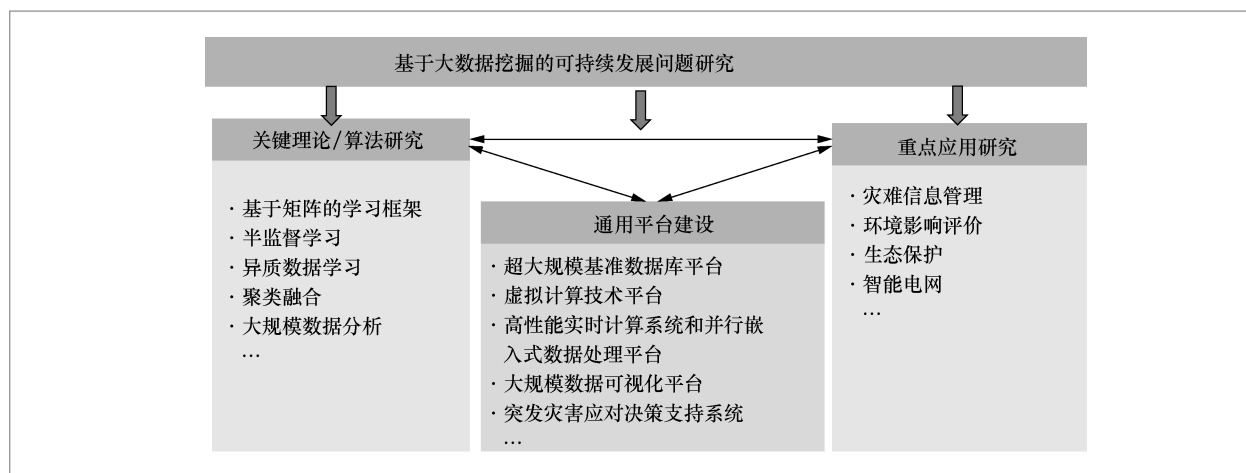


图3 基于大数据的计算可持续性研究框架

气候变化与环境污染的首要因素是温室气体 (green housegases, GHG) 的排放, 而建筑行业 and 建筑物正是产生GHG的主要来源之一^[11]。在我国, 建筑环境影响更是城市化加速发展过程中无法回避的重要议题。目前, 我国的建筑业仍处于高投入、高消耗换取高增长的发展模式, 给我国的生态环境带来了巨大影响^[12]。近年来, 我国碳排放和建筑扬尘等导致雾霾天气频发, 严重影响了人们的正常生产和生活。开展建筑环境影响评价有助于缓解日

益突出的建筑发展与环境恶化之间的矛盾, 具有重要意义。

然而, 目前广泛使用的环境影响评价方法代价高昂, 且仅能从建筑生态学的角度给出局部或简化的评价结果, 无法充分利用海量、高维、异构的建筑环境影响评价数据。对此, 基于计算可持续性的研究理念, 采用大数据挖掘技术, 提出了基于数据驱动的建筑环境影响评价模型 (如图4所示)。在该模型中, 利用基于约束的特征选择研究解决环境热点分

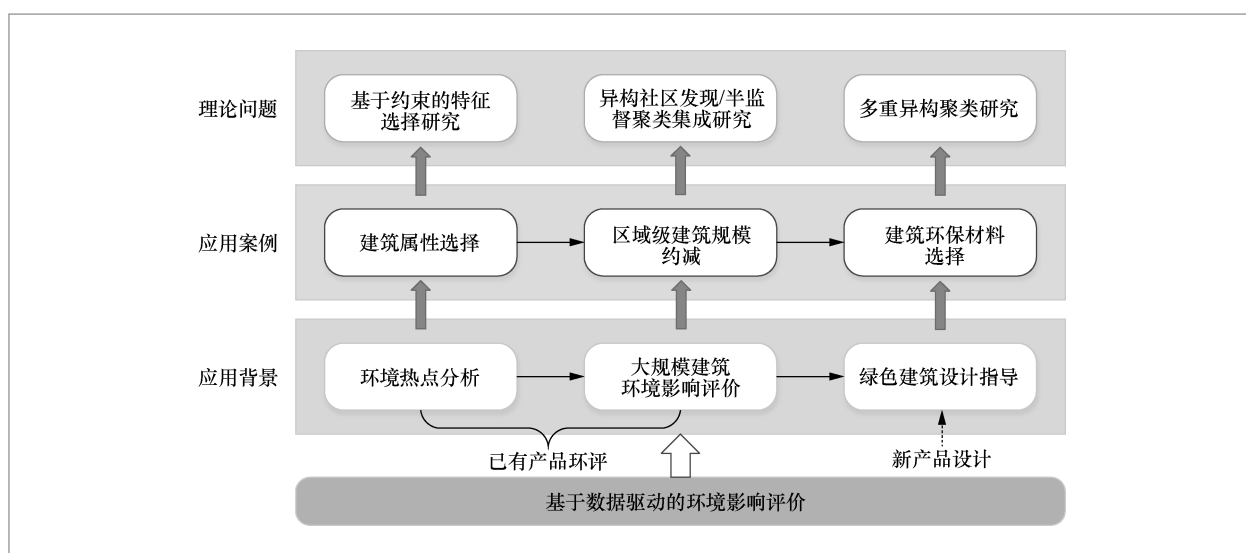


图4 基于数据驱动的建筑环境影响评价模型

析问题；利用异构社区发现、半监督聚类集成研究降低大规模建筑环境影响评价的代价；利用多重异构聚类研究辅助绿色建筑设计的指导。

在该研究中，建筑环境影响评价数据处理的难题被转化为一系列特征选择、分类、聚类的数据挖掘问题，再通过构建大规模分布式可持续发展数据处理平台，实现任务的分解与集成以及大规模算法的高效求解，从而利用大数据挖掘技术填补现有环境影响评价问题中数据处理的缺陷，并大大降低已有评估方法所需的时间和代价。

5 结束语

在计算可持续性研究的框架下，可持续发展的关键问题最终可以转化成计算和信息科学领域的决策和优化问题。大数据技术使得计算可持续性研究中大规模、动态、复杂问题的建模和求解可以以数据驱动的方式来解决，从而极大地提升了计算可持续性研究的效力并扩展了其研究范围，进一步地，将可持续发展问题从人们普遍认为的政策驱动，真正落实到技术践行。

参考文献:

- [1] GOMES C P. Computational sustainability: computational methods for a sustainable environment, economy, and society[J]. *The Bridge*, 2009, 39(4):5-13.
- [2] FRENKEL K A. Computer science meets environmental science[J]. *Communications of the ACM*, 2009, 52(9):23.
- [3] CONRAD J, GOMES C, HOEVE W J V, et al. Connections in networks: hardness of feasibility versus optimality[C]//*Proceedings of the Fourth International Conference on the Integration of AI and OR Techniques Constraint Programming*, Brussels, Belgium, May 23-26, 2007. Berlin: Springer Berlin Heidelberg, 2007:16-28.
- [4] BARRETT C B, LITTLE P, CARTER M. *Understanding and Reducing Persistent Poverty in Africa*[M]. London:Routledge, 2008.
- [5] 周绮凤, 李涛. 大数据与计算可持续性[J]. *南京邮电大学学报(自然科学版)*, 2015, 35(5): 20-31.
ZHOU Q F, LI T. Big data and computational sustainability[J]. *Journal of Nanjing University of Posts and Telecommunications(Natural Science)*, 2015, 35(5): 20-31.
- [6] 李涛. *数据挖掘的应用与实践*[M]. 厦门: 厦门大学出版社, 2013.
LI T. *Data Mining Where Theory Meets Practice*[M]. Xiamen: Xiamen University Press, 2013.
- [7] KRAUSE A, GOLOVIN D, CONVERSE S. Sequential decision making in computational sustainability via adaptive submodularity[J]. *AI Magazine*, 2014, 35(2):8-18.
- [8] FINK D, HOCHACHKA W M, SORTE F A L, et al. Documenting stewardship responsibilities across the annual cycle for birds on US public lands[J]. *Ecological Applications*, 2015, 25(1): 39-51.
- [9] MILANO M, O' SULLIVAN B, GAVANELLI M. Sustainable policy making: a strategic challenge for artificial intelligence[J]. *AI Magazine*, 2014, 35(3): 22-35.
- [10] ZHOU Q, ZHOU H, ZHU Y, et al. Data-driven solutions for building environmental impact assessment[C]//*Proceedings of IEEE International Conference on Semantic Computing (ICSC)*, Anaheim, CA, USA, February 7-9, 2015. Piscataway: IEEE Press, 2015: 316-319.
- [11] COMSTOCK M, GARRIGAN C, POUFFARY S,

- et al. Building Design and Construction: Forging Resource Efficiency and Sustainable Development[R]. Technical Report, United National Environmental Program (UNEP), 2012.
- [12] 2012-2016年中国建筑行业市场全景调研及投资价值分析研究报告[R/OL]. (2012-07).<http://www.chinabgao.com/report/372466.html>.
- Research report of the research and investment value analysis of the chinese construction industry market in 2012-2016[R/OL].(2012-07). <http://www.chinabgao.com/report/372466.html>. □