



城市国际化论坛

FORUM ON URBAN INTERNATIONALIZATION

城市大数据的 分析与应用

Urban Big Data Analysis and Application

段 霞 ◎主编



中国经济出版社

CHINA ECONOMIC PUBLISHING HOUSE



城市国际化论坛

FORUM ON URBAN INTERNATIONALIZATION

城市大数据的 分析与应用

Urban Big Data Analysis and Application

段 霞 ◎主编



中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE

北 京

图书在版编目 (CIP) 数据

城市大数据的分析与应用/段霞主编.

北京: 中国经济出版社, 2016. 1

ISBN 978 - 7 - 5136 - 4062 - 6

I. ①城… II. ①段… III. ①现代化城市—城市建设—研究 IV. ①C912. 81

中国版本图书馆 CIP 数据核字 (2015) 第 273618 号

责任编辑 张梦初

责任审读 贺 静

责任印制 马小宾

封面设计 华子图文设计公司

出版发行 中国经济出版社

印 刷 者 北京艾普海德印刷有限公司

经 销 者 各地新华书店

开 本 710mm × 1000mm 1/16

印 张 16.25

字 数 320 千字

版 次 2016 年 1 月第 1 版

印 次 2016 年 1 月第 1 次

定 价 52.00 元

广告经营许可证 京西工商广字第 8179 号

中国经济出版社 网址 www.economyph.com 社址 北京市西城区百万庄北街 3 号 邮编 100037

本版图书如存在印装质量问题, 请与本社发行中心联系调换 (联系电话: 010 - 68330607)

版权所有 盗版必究 (举报电话: 010 - 68355416 010 - 68319282)

国家版权局反盗版举报中心 (举报电话: 12390) 服务热线: 010 - 88386794

编 委 会

编委会主任 王稼琼

主 编 段 霞

副 主 编 吴 康 陆 辉

编 委 (按姓氏笔画排序)

丁成日 王力丁 王世强 王连峰 王彦京

毛琦梁 史利国 申建军 乔卫兵 刘智勇

安树伟 李义虎 李尔炳 李如刚 李湛军

张贵祥 杨发喜 陈 飞 陈之昌 陈 岳

陈 剑 林利民 周 芳 祝合良 徐 君

韩 凯 缪明月 谭善勇

序

近年来,大数据已然成为网络热词,但我一直以为那是数据分析专业人士的事,与我的研究关联不大。直到2013年我接受了北京市发改委委托的“首都非核心功能疏解及在京津冀布局”课题,苦于不能获得基于地理位置的企业空间数据,因此对用手机、公交刷卡、道路监控等传感器自动生成的城市大数据产生了兴趣。当时我拜托城市研究共同体(URC)的伙伴们去探寻大数据在城市与区域规划、治理、发展中运用的规律,这样就有了2014年城市国际化论坛的主题“全球化与信息化时代城市大数据的分析与应用”,以及今天呈现在各位眼前的这本研究报告。

大数据是什么、怎么用,很多地方都能找到对于它的专业表达和各领域的描述,本书也有系统阐述。作为门外汉,我只想结合前面提到的课题谈一下对于大数据应用的两点体会。

一、权威部门掌握的政府、企业、行业数据亟待开放与共享

按照首都所承担的全国政治中心、文化中心、国际交往中心、科技创新中心这四大核心功能,根据《国民经济行业分类》国家标准,我们找出其对应的4位代码产业^①,将排除在外的其他类别列入可以疏解名单。但是有效的产业调整升级绝不是这样简单的行政安排可以实现的。市场竞争条件下影响产业聚集与扩散的因素很复杂,比如地理条件、规模效应、交通成本、消费者偏好、产业投入产出关系等。那些依托北京城市特点发展、有竞争优势的产业转移出去不仅会对当地经济增长、财政收入、就业、关联行业景气产生严重影响,而且不当的疏解还会影响企业与相关行业的生存与持续发展能力。因此,甄别出造成首都人口严重超载、交通拥堵和环境恶化的聚集区、行业和企业,才能有针对性地进行疏解。

如果可以获得基于空间定位的企业数据(可隐去企业名称等),我们就可以搞清楚基于地理位置的企业、行业聚集和发展情况,计算该行业为首都本地发

^① 详细内容请阅读“拓展篇 未能拥有充分数据下的首都非核心功能疏解研究”。——作者注。

展和其他地区生产服务的就业人数及其空间布局情况,从而判断应疏解多大规模及疏解哪些位置上的企业或机构,预判其疏解的成本、收益与风险,为疏解非首都功能产业决策提供更科学、准确的研究依据。由于《统计法》第九条关于商业秘密受法律保护的规定,我们无法获得基于空间位置的企业数据。

这一类数据属于国家权威部门掌握的数据,包括政府、企业、行业以及其他机构拥有的各种形式的数据资源。这些都是传统意义上的权威数据平台的数据,通过统计调查手段获得。这类数据的价值在于开放与共享,开放原始数据(Raw Data)。把非涉密的统计数据开放出来,不是不要控制与管理,而是通过开放授权提供。自2009年美国提出开放数据计划(Open Data Program)后,许多城市开放给市民那些与他们生活息息相关的非机密数据,并提供公共服务和相关产品,鼓励公众和开发者应用并从中得到好处。Code for America上每天都会有志愿者们无偿提供解决城市问题的各种案例,城市管理运营成本大大降低,城市看似复杂的系统里面的经脉也因众人参与逐渐变得清晰。

二、具有异构性和稀疏性的城市大数据须多学科协同应对

城市大数据从哪里来?我们每天网聊、网购、公共交通、开车或在路上步行、手机通信、叫外卖与快递、拨打服务电话、用煤气水电等,所有的时间、地点与其他对象的联系、需求与兴趣,只要有传感器,就会被感知,不自觉地成为城市大数据的贡献者。当然,也有一些组织提供开源数据平台,让用户主动提供数据。对城市感知、捕获到的这些数据,研究者加以分析挖掘,就可以发现城市人口聚集、交通拥堵、资源消耗、环境恶化、管理规划落后的具体体现、问题严重程度,理解城市的运行规律和公众的真实需求,弥补传统规划、管理与政策的不足,使精细化的城市经营、规划、管理成为可能。

这些开放的城市数据十分丰富,但是刚接触时会觉得无从下手,因为数据属性差别很大。有时序、空间数据,点、线、面、空间图数据,也有流量、轨迹数据,文字、图像数据等,有研究者根据这些数据的属性将之称为“异构数据”。^①怎么分析与计算?这需要懂得城市科学专门知识,给数据下定义,贴标签,建立数据与概念、原理、问题以及数据之间的关联性,使之成为用之有“据”的数,才有意义。其次,这些数据虽然很庞大,但却不是全样本,网络不能覆盖或没有传感器的区域,不习惯上网、刷卡的人群不在其列。比如目前利用公交卡数据研究居民出行方式,只是针对使用公交卡人群,不用公交卡的本地居民和外地过

^① 郑宇. 城市计算概述[J]. 武汉大学学报信息科学报. 2015(4):2.

客则数据缺乏,这些也是研究居民公共交通出行方式须考虑的重要人群信息。这就是所谓的数据看似海量,实则稀疏的问题。此外,网络平台上的虚假信息也很难甄别,数据的局限性也是显而易见。

在前面讲的课题研究中,没有第一类的企业空间统计数据,也可以借助第二类的城市大数据分析聚集区产业尤其是就业人群职住情况对当地交通、人口构成、资源承载压力、公共服务等的影响,推导出疏解什么、疏解多少、在哪选址最优的方案。例如动物园批发市场搬迁问题,根据批发市场选址或靠近市场、或靠近原产地、或考虑集聚规模效益的特点,借助商家、客户以及人车轨迹等大数据可以分析:第一,有多大规模、哪些产品是提供给临近地区零售商的,这一部分无论怎么疏解,产品都会流入本地,搬迁会产生新的交通问题,可以考虑保留相应比例;第二,有多大规模、哪些类别的产品批发是提供给郊区、外地的零售商和散户的,这一部分可以疏解,应根据现有客户的来源在相应的交通干道附近选址设点;第三,有多大规模、哪些类别是基于品牌和聚集效应形成的,可以创新为网络营销模式,在原地保留批发总部,将仓储地建在河北、天津、北京郊区,既带动天津、河北经济发展与就业,又保留“动批”这个品牌。

目前,这些数据“孤岛”大多掌握在运营商手里,政府需要搭建城市大数据的应用平台,建立大数据交易市场。如果城市大数据应用能突破上述局限和问题,在感知技术和分析方法上更成熟,就可以为更加精细、科学的城市规划、管理和决策提供依据,这将有更多的市民来领养行道树、消防栓、人行道扫雪任务,参与城市问题的解决,创造新的就业岗位,使城市生活在减少政府财政压力的同时变得更美好。

大数据无处不在且时刻涌现。我们有理由相信大数据的应用将使城市变得更加繁荣、精细而和谐。请记住马克·安德森的名言“Software is eating the world”,主动寻找跨界合作伙伴,掌握城市大数据研究规律。

还好,我们在路上。

段 霞

2015年11月26日

目 录

序	1
---------	---

上篇 大数据在城市规划与管理中的应用

第 1 章 大数据时代的来临	3
1.1 技术进步催生数据革命	3
1.2 从大数据到开放的城市数据	4
1.3 大数据与城市规划管理	8
第 2 章 应用理论与方法	10
2.1 城市规划与管理的经典理论	10
2.2 大数据时代的新理论与新方法	12
第 3 章 变革与发展	22
3.1 空间尺度的变革:从宏观分析到微观模拟	22
3.2 时间尺度的变革:从静态截面到动态连续	23
3.3 研究粒度的变革:从以地为本到以人为本	24
3.4 研究方法的变革:从单一团队到开源众包	24
3.5 研究目的的变革:从普世研究到问题导向	25
第 4 章 应用案例	26
4.1 全国城市地块尺度建成区界定	26
4.2 地块尺度模拟全国城市扩张过程	27
4.3 全国街道尺度人口对 PM _{2.5} 的暴露评价	28
4.4 全国城市增长边界评价	29

4.5 全国地块尺度人口空间化与属性合成	30
4.6 结论	31

中篇 大数据在城市公共治理中的发展运用

第 5 章 国家与城市治理视域中的大数据	35
5.1 理解大数据	35
5.2 大数据在现代治理视域的意义	36
5.3 现代治理视域中大数据存在的局限	37
5.4 大数据在现代治理中的应用	40
5.5 结语	41
第 6 章 互联网时代的社会治理	43
6.1 从传统城市管理走向国际化的城市治理	43
6.2 从社会管理到社会治理	43
6.3 城市国际化与社会治理	45
第 7 章 大数据在社会治理中的作用与功能	47
7.1 教育领域的应用	47
7.2 城市减灾中的应用	53
7.3 新闻媒体的应用	58
7.4 社会组织的应用	60
7.5 大数据技术创新社会治理	62
第 8 章 中国社会组织的公开透明度分析	66
8.1 2014 年 GTI 指数总体得分情况	66
8.2 不同地域的公益组织 GTI 指数得分情况	70
8.3 不同领域的公益组织 GTI 指数得分情况	75
8.4 不同成立时间的公益组织 GTI 指数得分情况	79
8.5 不同登记注册性质的公益组织 GTI 指数得分情况	81
8.6 数据分析结论	84
第 9 章 应用案例	86
9.1 北京奇点公益信息技术服务中心	86

9.2 公益慈善信息化义工联	88
9.3 益云社会创新中心	90
9.4 新浪财经	95
9.5 北京益盟网络科技有限公司(新公益网)	98

下篇 大数据在城市和区域发展中的应用

第 10 章 新的范式与研究成果	104
10.1 基于居民时空行为的研究新范式	104
10.2 区域发展研究中的应用	106
第 11 章 机遇与挑战	109
11.1 大数据利用带来的新机遇	109
11.2 正在面临的问题与挑战	113
第 12 章 作用与功能	115
12.1 大数据在城市与区域规划中的作用与功能	115
12.2 大数据在城市建设中的功能与作用	117
12.3 大数据在城市治理中的作用与功能	119
12.4 大数据在城市发展中的作用与功能	121
12.5 大数据在区域政策、区域治理中的作用与功能	123
第 13 章 应用案例	127
13.1 城市实验室	127
13.2 智慧城市建设	131
13.3 大数据背景下的区域治理——以新加坡为例	137
第 14 章 未来发展趋势	144
14.1 大数据下城市规划变革与创新	144
14.2 跨学科融合和研究方法改进	145
14.3 大数据产业发展	146
14.4 区域品牌的大数据营销	147

14.5 城市与区域的数据化治理	147
14.6 大数据与新型工业化	148
14.7 大数据与区域经济增长转型	150

拓展篇

第 15 章 未能占有充分数据下的首都非核心功能疏解研究	153
15.1 首都非核心功能的基本内涵及疏解的可行性	155
15.2 非核心功能产业疏解原则与行业特点	165
15.3 京、津、冀三地城市功能特点与产业调整布局原则	186
15.4 功能疏解(产业转移)的条件与推进重点	201
15.5 主要结论与政策性建议	220
附录 北京市主要产业活动的成本—收益分析	227
参考文献	235



城市国际化论坛
FORUM ON URBAN INTERNATIONALIZATION

上篇 大数据在城市规划 与管理中的应用^①

本篇系统阐述了大数据在城市规划与管理中的应用进展。首先,介绍了大数据发展的时代背景以及大数据、开放数据和开放的城市数据在相关领域的定义、内涵和外延应用,并系统阐述了大数据引入城市规划与管理的必要性;其次,总结归纳了大数据在城市规划与管理领域应用的理论基础和代表性技术方法,并从空间、时间、研究粒度、研究方法、研究目的等方面提出大数据时代城市规划与管理的若干变革;最后,给出了研究团队的代表性研究案例。

^①龙瀛. 北京城市规划设计研究院, 北京 100045; 吴康. 首都经济贸易大学城市经济与公共管理学院, 北京 100070; 王碧宇. 苏黎世联邦理工学院, 苏黎世 8092; 王江浩. 中国科学院地理科学与资源研究所, 北京 100101; 李栋. 中国城市规划设计研究院, 北京 100044.

第1章 大数据时代的来临

1.1 技术进步催生数据革命

1.1.1 技术的发展

ICT(Information and Communication Technologies)是指信息与通信技术,以计算机为核心的 ICT 技术借助网络平台飞速发展,已经成为人们生活中不可或缺的重要部分。ICT 技术既包括传统的电话、书信、电报、电视等信息传递与通信技术,也包括新兴的智能手机、卫星通信、互联网技术、云计算、物联网等(申悦等,2011),除软硬件技术以外,ICT 技术也包括利用信息产品为社会提供服务的信息服务业。ICT 技术的发展使人们的信息交换与沟通变得更加便利与频繁。

根据《中国互联网络发展状况统计报告》,截至 2015 年 6 月,我国网民规模已达到 6.68 亿人,互联网普及率高达 48.8%,其中城镇地区的互联网普及率高达 64.2%。

随着智能手机的逐渐普及,移动端逐渐取代计算机成为互联网发展的核心,截至 2015 年 6 月,我国手机网民数量已经高达 5.94 亿人,网民中使用手机上网的比例接近 90%^①。这说明随着移动终端技术的不断发展,手机功能不断增加,用户体验逐渐提升,手机上网即时、方便、高效等优势逐渐发挥,得到了网民的青睐。

由于手机具有私人、即时、便携、个性化等特点,手机网络的发展促进了信息服务业的蓬勃发展,并带动了一系列传统行业的变革,如医疗、教育、餐饮、银行、家政服务等。目前,已经出现了互联网金融、移动医疗等新兴领域。伴随着这些新兴信息服务,手机所承载的信息量呈几何式增长,并且产生大量与用户相关的数据。可以说随着物联网、互联网、云计算、三网融合等 ICT 技术的迅猛发展,数据开始迅猛增长,并为众多领域带来了变革的机遇与挑战,数据革命已

^① 第 36 次中国互联网络发展状况统计报告。

然在人们生活的各个领域不断上演(王元卓等,2013)。

1.1.2 数据革命

2007年,已故图灵奖获得者吉姆·格雷(Jim Gray)在加州山景城召开的NRC-CSTB上,作了《第四科学研究范式:密集型数据挖掘》的演讲,提出e-Science是科学范式的一次革命。格雷提出,近年来试验、理论等各种不同渠道正在产生大量数据,这些数据背后蕴含了大量的信息知识,对其的挖掘研究成为一种新的科学范式(刘红等,2013)。

21世纪是信息化的世纪,ICT技术发展所带来的海量数据和信息量无疑是一座重要的宝库。阿里巴巴集团董事局主席马云在“2014(第十四届)中国年度管理大会”上提出,阿里巴巴并不是一个网络购物平台,而是一个数据中心。2014年11月11日,淘宝网上购物交易额高达571亿元,由此带来的大量用户数据与信息催生了一系列中国消费者行为的研究与分析,淘宝数据就是一个数据中心。在大数据时代,这些网络平台已经远远不局限于为用户提供服务,更为重要的是为决策者提供海量数据和信息。

目前,各个国家都非常重视大数据的发展与研究,希望在此数据革命之际掌握先机。2012年3月,美国白宫科技政策办公室发布《大数据研究和发展计划》,并成立大数据高级指导小组,将大数据研究提高到国家战略层面。欧盟也将斥资25亿美元投资大数据行业,并制定了包括“2020战略”“释放欧洲云计算服务潜力”战略等一系列战略保证欧盟在数据革命中不落下风^①。目前,我国也在积极制定与大数据相关的国家战略,并已经召开了贵阳国际大数据博览会等相关会议。2015年8月31日,国务院发布《关于印发促进大数据发展行动纲要的通知》,提出大数据能够推动经济转型发展、重塑国家竞争优势、提升政府治理能力,将大数据置于很高的战略地位。这些战略与措施足见各国对数据革命、大数据时代的重视。

1.2 从大数据到开放的城市数据

1.2.1 大数据(Big Data)

目前关于大数据的概念学界还未形成统一定义。最经典的定义莫过于维克托在《大数据时代》一书中的描述,大数据就是指样本等于总体,也就是说数据量之大已经包括研究范围内的所有总体(维克托等,2013)。因此,按照这个

^① <http://www.mofcom.gov.cn/article/i/jyjl/m/201412/20141200826137.shtml>.

说法,传统的随机抽样等方法将不复存在,未来的研究将基于所有总体,以避免抽样带来的偏差。全球领先的咨询公司——麦肯锡对大数据同样给出了一个经典的定义:大数据是指那些数据规模超过一般数据软件捕捉、存储、管理和分析能力的信息。对大数据,麦肯锡没有给出一个具体的数据,因为他们认为随着技术的进步,大数据的规模也会不断发展(Manyika 等, 2011)。全球第一家信息技术研究和分析公司 Gartner Group 认为:大数据是大量、高速、多样的信息,它提供高效、创新的信息形式来提高理解和决策过程^①。IDC 认为,大数据技术描述了新型数据技术与结构,用于低成本地从海量丰富的数据中提取有用信息,并使得高速捕捉、挖掘、分析成为可能^②。IBM 对大数据的定义是,大数据具有三个属性:大量(Volume)、多样(Variety)和高速(Velocity),这是大数据的“3V”特征。由此可见,大量、多样和高速是每个定义中必提及的关键点,也是大数据的核心特征^③。

国内学者对于大数据的概念也有自己的定义。李国杰等(2012)认为大数据是指无法在可容忍的时间内用传统 IT 技术和软硬件工具对其进行感知、获取、管理、处理和服务的数据集合。他在 3V 特征的基础上又加入了 Value(价值巨大但密度很低)这一属性。王珊等(2011)认为与传统数据相比,大数据主要具有三个变化:数据量由 TB 级升至 PB 级,并仍在爆炸式增长;分析需求由常规分析转向深度分析;数据平台由高端服务器逐渐转向中低端硬件构成的大规模机群平台以更好地储存大数据。

在城市规划领域,大数据也有一些特定的定义。叶宇等(2014)认为大数据是一个庞大的概念集合,用于指代各种规模巨大到无法通过手工处理来分析解读信息的海量数据。他认为大数据具有双重内涵:第一,大数据的数据样本量足够大;第二,大数据本身并非一个新概念,而是传统小样本数据分析研究方法在样本数量上的扩展,认为大数据可以减弱目前研究对传统抽样方法的依赖,从而做到样本的全覆盖。实际上,与传统数据相比,大数据主要呈现出精度高(以单个的人或设施为基本单元)、覆盖广(不受行政区域限制)、更新快(每月、每日、甚至每分钟更新)等特点。它不仅意味着更大的数据量,更反映了数据背后关于人群行为、移动、交流等活动的丰富信息,与新型城镇化时期“以人为本”“存量更新”“自下而上”等城市规划理念不谋而合,被认为是促进城市规划科学化与城镇治理高效化的有力工具,也因此为相应的学术研究、规划实践和商

① <http://www.gartner.com/it-glossary/big-data/>.

② http://www.infosysblogs.com/bigdata/2013/02/what_actually_is_big_data.html.

③ <http://www-01.ibm.com/software/in/data/bigdata/>.

业咨询带来了新的发展契机。

1.2.2 开放数据

开放数据(Open Data)是指对公众开放的数据。茅明睿(2014)认为开放数据是指经过挑选与许可的数据,这些数据不受著作权、专利权以及其他管理机制所限制,可以开放给社会公众,任何人都可以以任何目的自由出版使用。从经济学角度讲,开放数据应该是公共物品,无论其来源如何,它都始终具有非竞争性和非排他性,即所有公众都可以在不付出任何代价的情况下共享。根据这个定义,开放数据不仅指大数据,还包括传统的权威数据来源以及任何公众在任何渠道和平台发布的任何数据,如气象爱好者在气象论坛发布的自己观测的数据。目前越来越多的公众数据、网页数据等完全对外开放,网络信息已经成为一个巨大的数据库,遗憾的是,政府、权威机构和部分网站掌握的大量可靠数据还具有数据壁垒。因此,有些研究者认为开放数据重点在于政府和权威机构数据的开放,如曹凌(2013)认为开放数据是指公共机构产生、收集或支付的所有信息,包括地理信息数据、统计资料、气象资料等,有政府资助的研究项目的数据,并包括数字图书,这些公共数据可以随时访问和咨询,也可以重新再利用。目前,针对这一现象,开放数据运动应运而生,民间组织、学术机构、商业公司、政府等呼吁开放政府、开放知识、开放获取、开放应用程序(Application Programming Interface, API)接口等(茅明睿,2014)。

开放政府及权威机构数据,首先意味着数据的规模变化,为使用者提供的数据集应是跨越某个政府或某个机构的数据,可覆盖用户的所有需求;其次意味着开放政府数据接口方式,即政府已创建的数据,可直接支持专业技术人员创建新的应用,成为随时调用的创新资料、生产资料(曹凌,2013)。目前,美国和欧盟为了积极推动大数据的研究,已经斥巨资大力推动开放可靠平台的数据,而中国政府在这方面也在积极努力,国务院积极推动政府数据资源共享开放工程,到2018年将建成国家政府数据统一开放平台;上海市成为中国地方政府开放数据的先驱,上海经信委2014年5月发布的《2014年度上海市政府数据资源向社会开放工作计划》,要求2014年底之前开放28个市级部门的190项重点领域数据内容,涵盖产业发展、教育科研、交通管理、民政服务、养老医疗、法律法规等领域。然而根据《2014年我国大数据发展分析报告》显示,我国目前数字化数据资源总量仅为美国的7%,欧洲的12%,相较国外的开放数据平台,我国数据开放还有很长的路要走。