



国家出版基金项目
NATIONAL PUBLICATION FOUNDATION

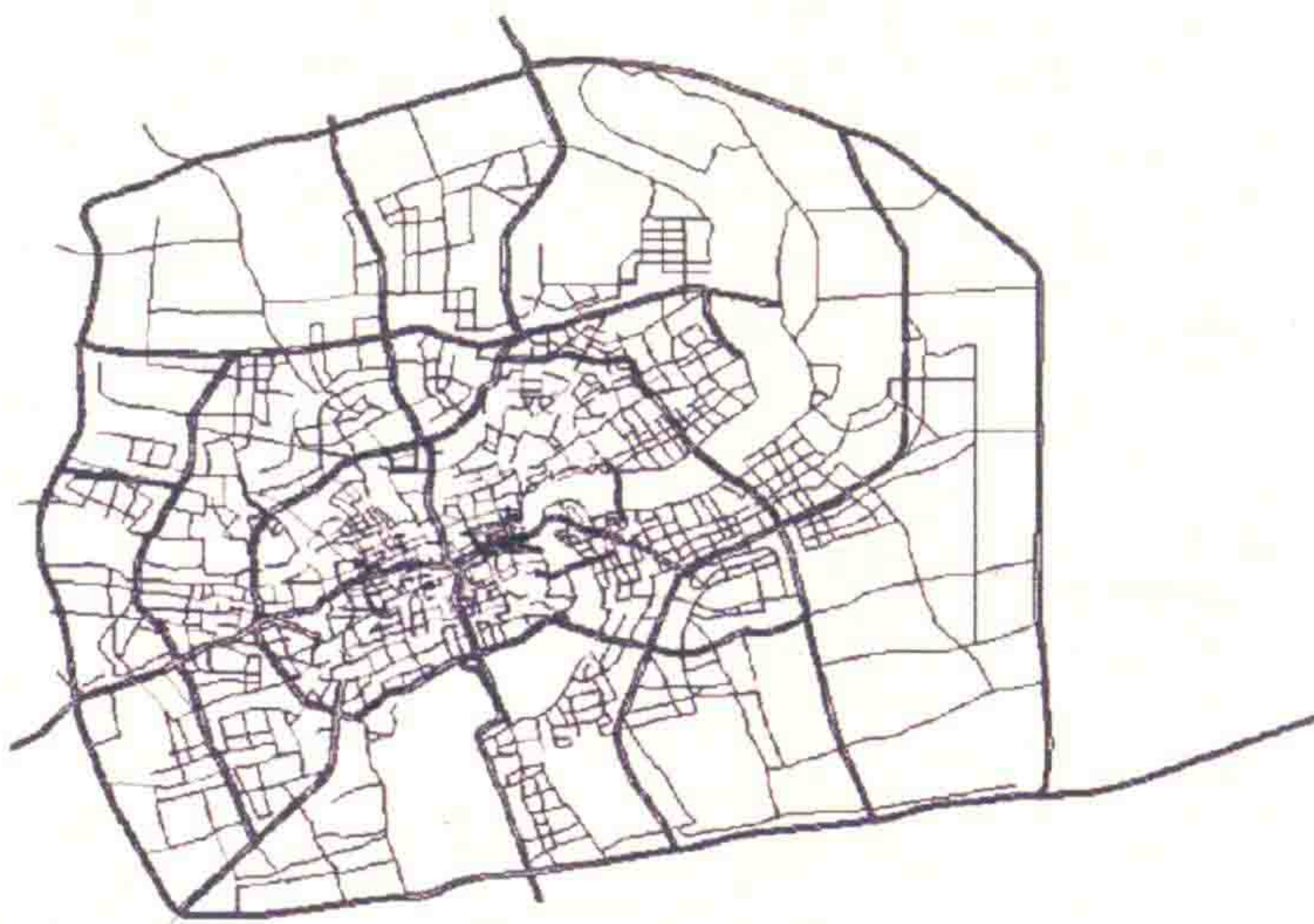
“十三五”国家重点图书出版物出版规划项目

面向未来的交通出版工程·政策与规划系列

透过大数据 把脉城市交通

杨东援 段征宇 著

同济大学出版社





国家出版基金项目
NATIONAL PUBLICATION FOUNDATION

“十三五”国家重点图书出版物出版规划项目

面向未来的交通出版工程·政策与规划系列

透过大数据把脉城市交通

杨东援 段征宇 著

 同济大学出版社
TONGJI UNIVERSITY PRESS

图书在版编目(CIP)数据

透过大数据把脉城市交通/杨东援,段征宇著.--上海:同济大学出版社,2017.3

(面向未来的交通出版工程·政策与规划系列/汪光焘主编)

“十三五”国家重点图书出版物出版规划项目

ISBN 978-7-5608-6749-6

I. ①透… II. ①杨… ②段… III. ①城市交通—研究 IV. ①U12

中国版本图书馆 CIP 数据核字(2017)第 020414 号

面向未来的交通出版工程·政策与规划系列

透过大数据把脉城市交通

杨东援 段征宇 著

出品人: 华春荣

策划: 高晓辉 陆克丽霞

责任编辑: 高晓辉

责任校对: 徐春莲

装帧设计: 陈益平

出版发行	同济大学出版社 www.tongjipress.com.cn (上海市四平路 1239 号 邮编:200092 电话:021-65985622)
经 销	全国各地新华书店、建筑书店、网络书店
排版制作	南京新翰博图文制作有限公司
印 刷	上海安兴汇东纸业有限公司
开 本	787mm×1092mm 1/16
印 张	17.75
字 数	443 000
版 次	2017 年 3 月第 1 版 2017 年 3 月第 1 次印刷
书 号	ISBN 978-7-5608-6749-6
定 价	108.00 元

版权所有 侵权必究 印装问题 负责调换

《面向未来的交通出版工程·政策与规划系列》编委会

学术顾问

邹德慈 中国工程院院士
郭重庆 中国工程院院士
郑时龄 中国科学院院士

主任

汪光焘

副主任

杨东援

编委(以姓氏笔画为序)

马 林	方守恩	叶建红	朱 洪	关志超	杨 飞
杨 轸	杨 涛	杨 超	李 健	李 锋	吴 兵
宋小冬	张晓春	陈小鸿	邵 丹	林 群	赵鸿铎
段征宇	倪桂明	徐瑞华	海德俊	惠 英	潘晓东
潘海啸	薛美根				

■ 总 序 ■

伴随城镇化和机动化进程,交通的发展目标、关注问题和对策方法等也在发生着巨大的变化。未来的交通运输工程将是工程科学、人文科学等交叉学科融合发展又具有独立特色的领域,在这一领域中相关理论与技术正处于挑战、探索和机遇并存的重要发展阶段。

中国快速城镇化、机动化所带来的前所未有的压力,由于受到资源和环境的制约,已经不可能单纯依靠交通基础设施建设来加以化解,需要构建适合中国国情的战略—政策—规划—建设有机融合的城市交通一体化对策理论体系;城市交通进入建设与管理并重的转型发展阶段,将社会管理与交通技术系统建设有机融合,通过理性供给和动态调控,引导城市交通模式进入可持续发展的轨道,是理论研究者、管理者和工程技术人员苦苦思考与探索的问题。

近年来相关技术领域的快速发展,促使交通领域中的许多技术概念正在发生巨大的变化。信息技术与控制技术的融合,促成了车联网的孕育与发展;对交通安全的社会关注,促进了车辆与交通技术设施两方面的主被动安全技术的应用;“互联网+服务”、“互联网+交通”催生了共享单车、分时租赁、网约车等新型交通服务模式;新型有轨电车、个人城市交通(PRT)等交通方式正在逐渐走进我们的生活;交通信息系统建设以及大数据理论的发展,为针对具有复杂适应特征的城市交通系统监测和战略调控创造了条件。

在此背景下,无论是交通工程研究者、工程技术人员还是行业管理者,都深感原有知识和经验面临挑战,需要重新认识我们所面对的问题,适应需求和技术环境变化进行理论创新与技术变革。

《面向未来的交通出版工程》这套丛书正是为了适应这种需求而编纂。面对一个快速变化和发展的领域,采用滚动组织编写的方式,力求更快地反映出相关理论研究成果和实践经验。

在本丛书的出版选题中,强调如下三个基本原则。

(1) 适应发展阶段需求的理论针对性。

我国交通进入转型发展的新阶段：对生态环境和城市人居环境的关注，要求在可持续发展理念指导下，研究绿色交通和新能源车辆推广的对策问题；在城市规划进入用地总量控制内涵式发展的背景下，需要正确处理土地再开发引发建成区交通需求增长与交通基础设施难以大规模扩展的矛盾；城镇化进入城市群发展阶段，要求从城市群视野整合综合交通体系，调整中心城交通模式以适应功能布局的演化。

(2) 符合国际交通理论发展趋势的科学技术原创性。

在交通网络流分析和交通行为分析等传统领域的研究日益深化；大数据和复杂性理论的交融对交通理论变革产生了重要的推动作用；城市规划理论关注点从“位空间”转向“流空间”、地理学的关系转型和对空间中行为研究等方面所提供的成果借鉴。整个学科体系正在出现变革的趋势。

(3) 适合中国国情特点的实践探索性。

城市交通战略、政策和规划并非单纯的技术问题，而是一种科学发展观和公共政策的体现。正因为如此，在特定管理制度架构下的实践探索，具有理论和技术研究所不可替代的作用。

中国交通的快速发展迫切需要与之适应的理论与技术支持，中国交通的实践又为相关理论与技术的变革提供了成长的“土壤”。面对发展与变革，不必苛求体系的完美，研究和实践中的百花齐放，必然带来中国交通科学与技术领域的万紫千红。



2016 年 4 月

■ 前 言 ■

城市交通大数据分析的方法隐藏于目的之后,因此本书的目标专注于大数据支持下城市交通的决策判断思维和技术思路问题,舍弃了对于具体技术的详细讨论。

伴随大数据和复杂系统理论的发展,城市交通理论面临着新的变革。在大数据创建的新技术环境基础上,依托复杂性理论等指导,融合传统技术形成新的决策支持框架,会产生推动城市交通理论革命性发展的激流。

城市的快速发展将越来越多的难题扔到了决策者的面前:不断增长的需求与难以为继的资源之间的矛盾日益突出,控制需求的尝试将交通管理者和技术人员推向了社会的对立面;“专车”之类的新型服务方式突破了原有的管理理念与框架,广泛的社会响应与管理者的不知所措形成了鲜明的对照;交通规划正在步入社会参与的开放时代,不再是管理者与技术“精英”的“私家会所”。诸如此类的变化,使得决策者意识到未来的城市交通是体系作战,作为一个非常复杂的调控过程,简单依靠传统经验和理论制定城市交通对策的做法已经很难适应,交通工程正在寻求新的理论和技术突破。

正当传统技术与新任务之间的不适应困扰着研究者和工程师的时候,大数据技术将许多新的技术手段和方法展现在我们面前。移动通信、IC卡、FCD、网络POI等许多新的数据资源,借助数据驱动和大强度计算得到提升的决策支持能力,在大数据基础上实现的信息提炼与概括、自动深度学习等,在我们面前展开了一幅值得期待的未来景象。

但是技术道路绝不平坦,在初步尝试的喜悦之后,越来越多的困难呈现在人们面前。大数据的模糊性和盖然性特征引发研究者对于大数据分析与传统模型分析技术方法内在差异的思考;更加丰富的信息被塞入基于OD的理论框架,也暴露出不相适应与所受制约;将碎片化信息转化为决策依据过程中,大量“间接证据”的使用表现为一条若隐若现的探索路径;交通决策者与分析技术人员之间缺乏与大数据

体系相适应的技术语境,出现了新的沟通困难。

多年的研究经历产生了一种感受:大数据与城市交通的融合犹如一个技术领域的“围城”,圈外的研究者受到新技术展现出来优势的不断诱惑,而身陷其中的“探路者”则在苦恼如何跳出不完备数据环境的泥沼。

记得曾经在网络上看到西南交通大学一位交通工程博士研究生的博客:“大数据概念这两年又要在国内玩过火了的架势,不管写项目申报还是论文报告,不扯上个大数据似乎都有点不好意思拿出手。看了几本关于大数据的专著,全是老外讲他们如何把啤酒尿布摆在一起卖出去的,细加琢磨,不禁喉头一紧后背直发凉,且看且看。维克托·迈尔-舍恩伯格和肯尼思·库克在其合著《大数据时代》的开篇便强调了大数据的三个基本特征:要全体不要抽样、要效率不要绝对精确、要相关不要因果。再想想传统交通的思维模式:数据从抽样调查中获得,模型要努力提升精度,重视机理分析和因果关系。乖乖,这真是要推倒旧世界建立新世界的节奏么。”学生所提出的问题是尖锐的,交通工程学科在大数据面前应该冷静而深刻地思考。

面对这一系列问题,需要在总结已有经验基础上梳理交通大数据研究和应用思路。研究实践也使我们深切地感受到,如果技术要求和方向得以明确,具体技术的实现不会成为致命障碍,而思路迷茫与混乱却正在阻碍大数据技术快速融入交通工程学科领域。

对于将大数据技术融入城市交通分析体系来说,有三个重要的技术障碍横亘在面前:

首先是分析技术必须实现“对盖然性”的包容。虽然都与“不确定性”具有紧密关联,但是数字化的“概率性”与证明力的“盖然性”却是有着重要的区别,证明力的盖然性评估不能简单等同于数学上的概率验算问题。概率指的是事物出现的可能性,而盖然性则是对于事实真相认知程度的表征。例如,在应用 IC 卡数据分析用户乘坐公交行为过程中,根据刷卡时间间隔判断是否属于公交换乘,可以认为是在特定时间间隔内刷卡属于“换乘行为”的概率问题,但是要根据用户一段时间(例如 1~3 个月)的公交使用特征,判断其是否属于“依靠公交通勤”的用户,就成为一个盖然性问题。由于并不能肯定用户始终使用 1 张 IC 卡,由于分析者并不能确定用户属于上班一族等多方面复杂原因,我们只能判断一部分用户肯定(或者否定)是公交通勤用户,但是存在相当一部分难以真正判定其属性。也就是说,我们并没有掌握研究对象是否属于公交通勤用户的直接证据,而是根据其内在属性模糊联系的表面行为特征做出了具有盖然性的判断。因而,力图通过大数据把脉城市交通,就遇到了分析者到底多大程度上掌握了事实真相的评估门槛,不迈过这一门槛就难以做出

对策决断。

其次必须在“相关”与“因果”之间架设桥梁。在大数据时代,相关分析由于其具有可以快捷、高效地发现问题的优势而受到广泛的关注,在商业分析、医疗诊断等领域的成功应用,使其大有替代“因果”的趋势。的确,对于复杂适应系统来说,错综复杂的联系以及行为主体的适应性演化,造成构建精巧模型完整、准确地反映系统演化规律几近奢望。但是对于承担巨大社会责任的管理者来说,面对意见分歧的尖锐争论和心中无数的社会响应,完全撇开“因果”又是一种难以接受的决策方式。正因为如此,尽管交通领域中相当程度的技术研究都围绕着将大数据纳入基于 OD 的模型体系框架而努力,但所取得的进展相当有限。事实上,大数据在时空范围内的观察广度,以及对研究对象的大样本覆盖,超越了依托小样本数据构建的模型观察能力。但是对于复杂适应系统反映基本规律的内部模型而言,基于行为的模型、基于活动的模型,以及网络交通流分析模型等具有不可替代的优势。融会二者的优势,实现宏微观融合、大样本分析与案例剖析融合,对于城市交通分析技术创新具有战略性意义。

第三是需要从趋势外延预测走向反馈基础上修正认识掌控未来。由于城市交通的复杂适应系统特性,造成未来的演化具有很强的不确定性。简单根据过去趋势与特征推演未来的线性思维预测方法,难以适应快速变化的现实。大数据环境所提供的连续追踪观测的能力,为基于复杂性理论研究城市交通创建了不可或缺的基础。问题在于这并非简单的分析技术变革,而是决策模式和管控方法,乃至城市交通规划与管理体制的变革。兼顾管理、工程和研究理论体系的建立,需要不断地实践与探索,当务之急是尽快明确理论与技术的发展方向。这种学科发展的战略抉择,并非可以简单依据理念与愿望,而是需要在城市交通系统管控中所积累经验的支持,以及案例研究的启示。

总体来看,大数据对于城市交通所带来的机遇绝不仅是一种辅助技术手段,它使我们有可能拥有对复杂适应系统多方位、深入持续观测的战略技术手段,从而改变传统的城市交通规划与管理方法。同时,大数据又不是与传统分析技术的对立和排斥,毕竟绝大多数情况下大数据只能告诉我们“是什么”,却难以回答“为什么”。大数据的价值在于支持管理大智慧,在战略层介入管理决策才无愧于“第四范式”的称谓。

在上述认识的基础上,本书围绕如下问题的思考而展开:

(1) 什么是交通大数据? 交通大数据强调的是充分利用相关数据资源,更加全面、持续、完整地观察研究对象。从这一观点出发,我们不仅需要重视数据量大的资

源,也要重视全面观察问题不可或缺的数据资源,更需要突破“盲人摸象”的技术瓶颈。

(2) 城市交通为什么需要大数据? 由于城市交通所具有的复杂适应系统特征,必须深入观察系统的演化过程,有效地进行相应的调控,促使其循可持续发展的轨迹进化。从这个角度来说,大数据提供的是对一种战略层面的诊断机制、监测机制、预警机制的支持,同时是提升政府远见与智慧的重要技术工程。

(3) 城市交通领域大数据应用与其他领域的差异是什么? 城市交通领域中的主要决策都会对城市产生重大影响,需要尽力减少战略误判和对策失误。因此必须将大数据领域的关联分析等技术与本领域具有一定积累的因果分析技术有机融合,我们面对的并非一个简单的“信息挖掘”或者“信息融合”技术应用,而是对具有复杂适应系统特征的社会工程的“证-析”。

(4) 能否将大数据技术植入城市交通传统技术框架? 由于所采用的数据并非针对传统技术概念的定制数据,如果不突破传统技术概念框架的束缚,将很难发挥新数据资源的优势,反而会深陷与传统技术指标难以直接对应的困境。采用居民活动空间概念替代 OD 概念,将对于流量的关注拓展为对交通流内在构成结构的分析等,带我们进入了一个新的广阔应用空间。

(5) 城市交通大数据应用的技术主力是谁? 交通工程传统上培养了三种类型的人员:交通工程师、交通规划师和交通模型师,面向未来需要培养一种新类型的技术人员——交通数据分析师。他们需要与其他三类人员共同组成未来的技术分析核心,才能够应对巨大的挑战。

(6) 融合大数据技术与传统技术的框架是什么? 基于证据的决策判断将大数据分析的成果,与仿真和模型分析技术产生的成果融合在一起。实现这一技术构想的难点,在于如何使用非定制数据所产生的间接证据,以及如何处理判断过程所面临的盖然性。

(7) 多源数据基础上信息融合的技术特点是什么? 城市交通大数据涉及移动通信信令数据、IC 卡数据、车辆牌照检测数据、定点检测器数据、网络 POI 数据等规格各异、参照系不统一,从不同角度表征研究对象的数据,在数据层、特征层和决策层有效地进行信息融合,是避免陷入“盲人摸象”窘境的关键。

(8) 如何在宏微观分析之间架设桥梁? 一个嵌套式框架提供了研究的整体模板,在宏微观数据之间建立链接键,通过探索性研究正确地提出问题,基于聚类分析研究问题结构,保证案例研究避免“选择性偏差”,以及在此基础上设计精准对策的“靶向”行动方法等,均为有可能影响全局的关键技术环节。

(9) 如何建立城市交通战略调控的技术框架? 由于这绝非城市交通大数据分析研究领域可以单独解决的问题, 现阶段只是形成了一些粗糙的想法, 建立战略态势观测体系、适时响应对策机制、感知-认知-洞察的技术分析体系等, 是战略调控技术框架中有待验证和充实的重要环节。

本书分为两个基本部分: 第一部分由第 1, 2, 3 章构成, 论述城市交通领域大数据思维的主要技术概念; 第二部分由第 4, 5, 6, 7 章构成, 讨论大数据分析中的主要技术如何在城市交通领域中应用。

对于信息技术人员来说, 通过本书可以了解如何在城市交通这一舞台上展现大数据的魅力; 对于交通工程师来说则希望能够帮助他们增强大数据思维; 而对于交通数据分析师来说, 希望通过本书能够引导他们建立一个技术框架。

尽管在交通数据分析领域进行了十余年的摸索, 但是面对思维方式差异很大的学科跨度, 以及有限的实践经验积累, 要写好本书仍然感到力不从心。好在虽然仅是“一家之言”, 仍然能够引发关注与讨论, 聊以宽慰自身付出的努力, 以及助手、学生和同事的信任。

本书成果来源于国家自然科学基金资助项目, 项目编号 51478350, 在此表示感谢!

编者

2016 年 10 月

目 录

CONTENTS

总序

前言

1	大数据“热潮”中的“冷”思考	1
1.1	大数据的热潮	2
1.1.1	大数据技术的发展	2
1.1.2	大数据呈现出促进研究模式变革的萌芽	3
1.1.3	交通领域大数据技术的应用	3
1.2	交通工程基础理论的发展反思	8
1.2.1	网络交通流分析理论的奠基作用	8
1.2.2	交通行为分析理论的发展	10
1.2.3	面向未来发展的期待	12
1.3	大数据面前的追求与困惑	14
1.3.1	对新研究范式的关注	14
1.3.2	大数据并非完备的信息环境	16
1.3.3	引发困扰的原因剖析	19
1.4	大数据在城市交通分析中的价值	21
1.4.1	大样本所提供的优势	21
1.4.2	连续追踪获取的信息	22
1.4.3	不同测度的观察与探索	24
1.4.4	多种观察角度的互补	27
1.5	城市交通大数据分析的思考	28
1.5.1	问题导向的技术需求	29

1.5.2 基于大数据的“感知-认知-洞察”	31
1.5.3 “多维一体”的问题表征	32
参考文献	33

2 建立在复杂适应系统概念上的城市交通监测 39

2.1 正确面对城市交通演化的或然性	40
2.1.1 城市交通的非线性到底意味着什么	40
2.1.2 城市交通的战略调控	42
2.1.3 面对或然性的适时响应对策模式	43
2.2 将监测融入新的技术概念框架	45
2.2.1 复杂适应系统的基本概念	45
2.2.2 技术概念变换后的系统监测任务	47
2.3 基于聚类及分类的行为主体模式表征	49
2.3.1 行为主体主观属性的差异表征	49
2.3.2 个体行为特征的差异辨识与表征	51
2.3.3 新型服务模式的用户响应	55
2.3.4 交通服务提供主体的行为检测	58
2.4 城市交通系统宏观状态监测	60
2.4.1 通过模式划分简化数据	60
2.4.2 城市空间连接条件的度量	62
2.5 关于涌现的思考	67
2.5.1 发现路网交通状态的耦合特征	67
2.5.2 城市空间拓展过程中的聚落斑块特征	69
2.5.3 城市群空间联系结构的变化	72
参考文献	76

3 技术的传承与变革 77

3.1 信息化创造的交通决策支持数据资源	78
3.1.1 逐步完善的交通系统状态监测网络	78
3.1.2 信息服务系统所采集的“电子脚印”	81
3.1.3 互联网中的语义信息利用	83
3.2 新的数据资源与传统技术概念的嫁接	85
3.2.1 我国城市交通大数据的应用实践	86

3.2.2	基于传统技术概念的检测手段拓展	89
3.2.3	利用相关性架设连接传统技术概念的桥梁	91
3.3	适应数据环境的技术概念变革	94
3.3.1	大数据环境下基于证据的决策分析技术	94
3.3.2	宏观与微观融合的嵌套型技术分析框架	96
3.3.3	寻找更加适合的表述方式	97
3.3.4	借助聚类把握差异	98
3.3.5	通过关联寻找通向认知之门	101
3.3.6	通过比较研究探寻因果关系	102
	参考文献	103

4	特征提取、聚类分析与对象表征	105
4.1	“纷繁复杂”中的“规律性”	106
4.1.1	个体多样性背后隐藏的群体特征	106
4.1.2	不完备信息条件下的分类辨识	111
4.1.3	揭示时间变化中隐藏的规律	114
4.2	行为主体的多维特征属性	120
4.2.1	不同形式的活动特征表达	120
4.2.2	基于出行链的活动模式划分	125
4.3	基于属性特征的聚类分析	129
4.3.1	通过聚类实现研究对象细分	129
4.3.2	惯常行为模式的聚类分析	131
4.3.3	基于关联属性的聚类分析	134
4.4	类别划分基础上的比较研究	137
4.4.1	类别之间的特征比较	137
4.4.2	基于类别划分的位置点分布对比	139
4.5	将属性特征转化为跨界沟通的数据语言	145
4.5.1	作为跨界沟通桥梁的数据模型	145
4.5.2	借助多源流模式框架的工作协同	146
	参考文献	148

5	关联分析与相关性研究	149
5.1	透过关联认识联系	150

- 5.1.1 空间特征间的联系 150
- 5.1.2 通过联系走向深度思考 152
- 5.1.3 交通分区的空间关联 155
- 5.2 基于大数据的个体属性-空间关联分析 159
 - 5.2.1 基于牌照数据对车辆使用类别的区分 159
 - 5.2.2 类别结构空间分布的讨论 162
- 5.3 借助关联关系的问题转换 165
 - 5.3.1 可拓学思维与问题转换 165
 - 5.3.2 借助关联属性进行异常事件判别 167
- 5.4 基于关联分析的问题概括与归纳 169
 - 5.4.1 数据基础上提出问题与界定分析任务 170
 - 5.4.2 提升数据分辨率创建相关分析条件 174
 - 5.4.3 基于关联特征的问题分类 178
- 参考文献 183

- 6 信息融合与构建证据集合 185**
 - 6.1 信息融合与证据理论的技术整合 186
 - 6.1.1 依托间接证据的判断 186
 - 6.1.2 证据体系框架内的信息融合 187
 - 6.1.3 基于证据理论的判断综合 188
 - 6.2 通过数据资源互补进行信息可信度判别 190
 - 6.2.1 对移动通信数据提取轨道使用信息的可信度检验 190
 - 6.2.2 通过多源数据提高公交乘车位置信息质量 194
 - 6.2.3 借助数据融合的公交通勤人群辨识 196
 - 6.3 情报决策过程中的信息融合 200
 - 6.3.1 证据提炼过程中的情报决策 200
 - 6.3.2 多源情报比较基础上的真实性评估 201
 - 6.3.3 基于多源数据的关联分析 210
 - 参考文献 216

- 7 大数据与小样本数据的嵌套分析 217**
 - 7.1 探索性研究构建任务框架 218
 - 7.1.1 通过核心概念明确问题 218

7.1.2 通过小样本分析深化认识 219

7.1.3 宏微观嵌套的公交客户管理分析框架 223

7.2 基于 IC 卡数据的公交乘客宏观结构分析 224

7.2.1 用于公交使用行为模式判别的特征指标 224

7.2.2 公交 IC 卡用户的组群划分 225

7.3 问卷调查基础上的微观机理分析 228

7.3.1 基于用户忠诚度的机理分析框架 228

7.3.2 测量模型中变量关系的设定 229

7.3.3 针对厦门市公交通勤用户的问卷调查 230

7.3.4 通过结构方程模型探寻微观机理 235

7.4 宏微观数据结合明确改善重点 240

7.4.1 在 IC 卡数据与问卷调查数据间建立链接 240

7.4.2 宏微观数据结合的组群划分 242

7.4.3 根据组群时空分布结构确定重点改造工作对象 244

参考文献 246

8 结论与思考 247

8.1 收获——得到深化的认识 248

8.2 成果——渐趋成熟的技术 250

8.3 确认——经过验证的结论 252

8.4 思考——未来之路 253

附件 1 表格索引 255

附件 2 附图索引 258

致谢 265

1

大数据“热潮”中的“冷”思考

某些观念有时会以惊人的力量给知识状况带来巨大的冲击。由于这些观念能一下子解决许多问题,所以,它们似乎将有希望解决所有基本问题,澄清所有不明了的疑点。每个人都迅速抓住它们,作为进入某种新实证科学的法宝,作为可以用来建构一个综合分析体系的概念轴心。这种“宏大概念”突然流行起来,一时间把几乎所有的东西都挤到一边。

——苏珊·朗格^[1]