经典译丛

信息与通信技术

Cognitive Networked Sensing and Big Data

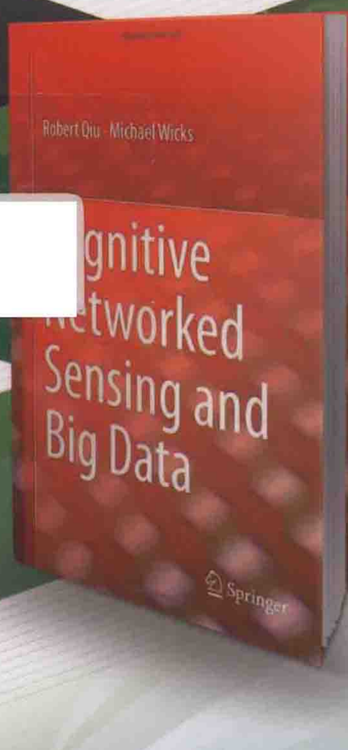
Springer

认知网络测量与大数据

Cognitive Networked Sensing and Big Data

【美】Robert Qiu（邱才明） Michael Wicks 著

宋 彬 秦 浩 刘海啸 等译



中国工信出版集团

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

经典译丛·信息与通信技术

认知网络测量与大数据

Cognitive Networked Sensing and Big Data

［美］ Robert Qiu（邱才明） Michael Wicks 著

宋 彬 秦 浩 刘海啸 等译

電子工業出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

本书系统论述了大规模网络下认知测量的基本理论及某些应用问题，基本涵盖了认知测量在理论和实际应用中各个方面的内容。全书包括随机矩阵和的性质，随机矩阵的集中不等式性质及高维大数据矩阵特征值的集中不等式性质，随机矩阵的非渐进和局部性质及渐进和全局性质。本书还详细介绍了认知测量理论在其他学科中的具体应用，包括压缩感知、矩阵填充、低秩矩阵恢复、高维协方差矩阵估计、高维信号检测、概率条件受限的优化问题求解等。最后讨论了相关理论在大数据应用中的分析方法。

本书可作为高等院校信息工程、通信工程、雷达、计算机、电子学、信息及计算科学等相关专业的本科生、研究生的认知网络课程教材或教学参考书，也可供数学、物理、生物学、系统科学等专业研究生，以及从事认知网络和大数据理论、技术、方法研究的科研工作者和工程人员参考。

图书在版编目(CIP)数据

认知网络测量与大数据 / (美)邱才明(Qiu, R.)，(美)威克斯(Wicks, M.)著；宋彬等译. —北京：电子工业出版社，2016.5

(经典译丛·信息与通信技术)

书名原文：Cognitive Networked Sensing and Big Data

ISBN 978-7-121-27551-7

I. ①认… II. ①邱… ②威… ③宋… III. ①计算机通信网－测量－研究 IV. ①TN915

中国版本图书馆 CIP 数据核字(2015)第 268075 号

策划编辑：马　岚
责任编辑：李秦华
印　　刷：三河市鑫金马印装有限公司
装　　订：三河市鑫金马印装有限公司
出版发行：电子工业出版社
　　　　　北京市海淀区万寿路 173 信箱　邮编　100036
开　　本：787×1092　1/16　印张：26.75　字数：684.8 千字
版　　次：2016 年 5 月第 1 版
印　　次：2016 年 5 月第 1 次印刷
定　　价：89.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：(010)88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：classic-series-info@phei.com.cn。

译者序

目前，大数据在认知网络测量、认知无线电、认知雷达和智能电网等相关领域研究中处于核心和基础的地位。本书以无线分布式计算和认知传感为最初构思的出发点，随着对大规模认知无线电网络的深入研究，发现“大数据”在其中扮演了非常重要的角色。因此，本书主要致力于解决以下问题：如何使用大规模网络去感知无线环境？如何处理获得的大数据？数据的维度又是如何去影响测量的精度的？为了回答这些问题，自然而然的会引入新的问题：到底需要什么样的数学工具？这些数学理论的发展现状如何？以及如何使用这些数学工具？

译者曾与本书的作者针对无线分布式计算及认知测量环境中的高维数据处理问题进行过深入的讨论，对本书的整体架构比较清楚，也为书中全面新颖的内容深深吸引。本书的作者作为长期从事无线通信技术领域的资深专家，在智能电网、大数据、无线网络与无线定位等研究方向上有着独到的见解和深刻的认识。作者还利用他们丰富的实践经验，深入浅出地剖析了大规模认知无线网络、硬件测试平台、分布式测量感知及分布式计算等方面的内容。这是一本优秀的书籍，所以我们由衷地希望将这本书推荐给广大的中国读者。

本书全面、清晰地阐述了必要的数学工具及其使用方法、认知测量环境中的高维数据处理与无线分布式计算等，在内容选取和安排上，第一部分主要介绍必备的数学基础和背景知识，包括矩阵值随机变量之和、测量的集中性、特征值及其函数的集中性、随机矩阵的局部非渐近性理论与全局渐近理论；第二部分将数学理论用于实践，针对随机矩阵，给出其在压缩感知和稀疏重构、矩阵填充、高维空间、概率约束优化以及数据集的高效处理等领域的应用。本书可作为从事无线通信研究和高等院校研究生的理论参考教材。

本书由宋彬、秦浩和刘海啸等翻译并审校，由宋彬负责全书统稿。其中秦浩负责第 1 章至第 6 章、第 10 章至第 13 章的内容，刘海啸负责第 7 章至第 9 章的内容。参与本书整理工作的还有来自西安电子科技大学综合业务网理论及关键技术国家重点实验室的博士生郭洁、李莹华和张悦等。在此，向所有为本书出版提供帮助的人士表示诚挚的谢意！

需要特别说明的是，本书译者是在尽量忠实于原书的基础上翻译的，由于专业水平和时间有限，书中翻译不妥之处，敬请广大读者及同行专家批评指正。

序　言

本书在最初构思的时候，以无线分布式计算和认知传感为出发点，但随着对大规模认知无线电网络的深入研究，我们发现“大数据”在其中扮演了非常重要的角色。因此，本书在结构和内容上做了相应的调整。从这个意义上讲，传感（sensing）可近似地认为等同于“测量”（measurement）。

我们在本书中试图回答以下几个基本问题：如何使用大规模网络去感知或测量无线环境？认知无线电有什么特殊的地方？我们又如何处理获得的大数据？数据的维度又是如何去影响测量的精度。为了回答这些问题，自然而然地会引入新的问题：到底需要什么样的数学工具？这些数学理论的发展现状如何？以及如何使用这些工具？

本书的基础是随机变量和随机过程等研究生课程。熟悉无线通信和信号处理对本书的理解会非常有帮助。并且本书可以视为相关理论的数学基础教材，作为 *Cognitive Radio Communications and Networking: Principles and Practice*(John Wiley and Sons, 2012) 和 *Introduction to Smart Grid*(John Wiley and Sons, 2014) 两本书的补充。

第 1 章主要介绍了一些必备的数学基础和背景知识，并引用了很多相关文献中的最新研究结果。这些基础知识或许对一些读者而言仍然有些困难，毕竟本书需要随机过程等研究生课程作为基础。

第 2 章至第 5 章作为本书（第一部分）的核心部分，介绍了相关数学理论，对于电子和计算机类的工科研究生可能会相对陌生。

其中，第 2 章介绍了矩阵随机变量的和。一个基本的问题就是“样本的规模到底如何影响其精确度？”我们研究的核心内容就是样本协方差（随机）矩阵，重点即伯恩斯坦类型的集中不等式。

第 3 章是本书研究内容的出发点，更详细介绍了相关数学理论。而本书将第 2 章放在这之前，主要是为了引导读者更好地理解如何处理矩阵的线性方程。其中集中不等式理论的研究是要回答下面的问题：假设给定随机向量 $\mathbf{x}$[①]，其在概率空间 $\mathcal{X}$（通常是高维的欧式空间）上取值，并且已知映射 $f:\mathcal{X}\to\mathbb{R}$，那么概率 $\mathbb{P}(|f(\mathbf{x})-\mathbb{E}f(\mathbf{x})|\geqslant t)$ 的下界会是多少？传统的概率论相关理论，对其严格准确的估计进行了大量研究。但在很多实际问题中，严格准确的近似是不可能的，而集中不等式就通过给出一个快衰落的下界实现一种“次优”的近似。本书的目标就是在传统概率理论失效的情形下，系统地研究处理这些所谓的“次优”问题。

随机矩阵的和是线性矩阵的和，但在实际应用中，我们常碰到的是非线性矩阵函数。这就促使我们研究第 4 章的内容，即测量集中性。在这一章中，我们利用的主要数学工具是矩阵的利普希茨函数，例如特征值。

① 本书英文版中数学符号较多，若逐一规范有可能出现二次错误，故在翻译版中的表示方式尽可能与原文保持一致——编者注。

第 5 章罗列了随机矩阵理论的发展，以及最新的研究成果。据我们所知，这些研究成果尚未在工程领域得到应用。尽管本章介绍的数学理论相对较为晦涩，但是我们认为如果能够充分理解本章内容，那么将对工程研究员的工作起到非常大的促进和指导作用。

第 6 章作为第 5 章的补充，使本书内容更加完整和系统，更多资料可参考 *Cognitive Radio Communications and Networking: Principles and Practice* (John Wiley and Sons,2012)。

在本书的第 7 章至第 13 章（第二部分），主要研究了这些数学工具在不同领域中的应用，重点是研究数学理论和不同实际应用之间的联系。

第 7 章重点介绍了其在压缩感知和稀疏重构领域中的应用，详细阐述了集中不等式在稀疏重构里起到的核心作用，并且可证明压缩感知领域里的约束等距性质就是集中不等式的另一种表现形式。

第 8 章介绍了矩阵填充中的相关理论应用。矩阵可以分解为特征值和特征向量，而低秩矩阵对应的特征向量则是稀疏的。

第 9 章讨论了高维空间的相关问题研究，尤其是统计学中方差矩阵估计问题，这是一个比压缩感知和低秩矩阵恢复更复杂的问题。协方差矩阵估计问题的解决，可以让我们在不用应用场合中利用相关的统计信息。

第 10 章利用协方差矩阵研究高维空间的假设检验。具体来说，主要在信息加噪模型的研究中，充分利用其低秩结构，因此我们把低秩矩阵的重构问题放在第 9 章讨论。目前现代检验理论的研究趋势，就是充分利用数据的结构特征（例如稀疏和低秩），由于理论研究的迅速发展，我们在本章中尽可能地收集了最新的研究成果。

第 11 章主要讨论有概率约束的优化问题。2003 年，Nemirovski 等人的研究表明，在过去被认为是无解的有概率约束优化问题可以转化为确定性凸优化问题，并通过现代凸优化技术求解。因此，闭式的伯恩斯坦集中不等式在其中起到了非常重要的作用。

第 12 章指出了集中不等式和面向数据的处理工作（例如低秩矩阵估计）的联系，我们在本章只是重点研究它们之间的关联。

第 13 章对本书的内容作了系统的总结，同时可以作为本书第 1 章的导言部分，引出本书其他章节的研究。从中可以看出，目前的研究中仍然存在很多问题，本书只是触及了冰山的一角。

第一作者希望感谢他在 2012 年秋季 ECE 7970 随机矩阵、集中和网络课程的学生。他们的评论很大程度上推动了这本书的进程。我们还要感谢 TTU 的博士生：Jason Bonior, Shujie Hou, Xia Li, Feng Lin 以及 Changchun Zhang，在校对方面给予的帮助。仿真的部分由 Feng Lin 完成，事实上他完成了这本书很多地方的概念和公式，特别是假设检验的部分。TTU 的 Zhen Hu 博士和 Nan Guo 博士也参与和帮助了讨论。第一作者的共同研究者 Husheng Li 教授（来自 Knoxville 的田纳西大学）提供了很多引人深思的意见。

多年来，Office of Naval Research（ONR）通过项目管理员 Santanu K. Das 博士给第一作者提供了大力的支持。我们的朋友 Paul James Browning 为这本书提供了技术支持。这项工作的部分资助来自 National Science Foundation（NSF）的两项奖助金（ECCS-0901420 和 CNS-1247778）；Office of Naval Research （ONR）的两项奖助金（N00010-10-10810 和 N00014-11-1-0006）；以及 Air Force Office of Scientific Research，地方承包商（主要合同号 FA8650-10-D-1750-Task 4）。书中的一些部分是 2012 年夏季，第一作者在挪威科技大学

（NTNU）的 Center for Quantifiable Quality of Service in Communication Systems （Q2S）担任访问学者期间完成的。在此答谢东道主 Yuming Jiang 教授。

作者想感谢 Springer（UK）的编辑 Brett Kurzman，感谢他对本书的兴趣，还有 Springer（UK）的 Rebecca Hytowitz 的帮助。

第一作者想感谢他的导师们。书中很多内容都是受他们启蒙：Weigan Lin（中国，电子科技大学）在遥感方面，Zhengde Wu（中国，电子科技大学）在电磁方面，Shuzhang Liu（中国，电子科技大学）在电磁材料方面，I-Tai Lu （纽约理工学院）在无线电传输方面，Lawrence Carin （杜克大学）在物理层信号处理方面，Leopold Felsen 在散射方面，以及 Henry Bertoni 在无线传输方面。他在 GTE 实验室（现为 Verizon Wireless）以及 Bell 实验室（Alcatel-Lucent）的同事们极大地启发了他的思路。

最后，第一作者希望感谢他的妻子 Lily Liman Li 对他的关爱，鼓励以及支持，陪着他度过孤单（却很兴奋）的写作旅程——她一直陪着他。还有他的孩子 Michelle，David 和 Jackie，他们点亮了他的生命。在这个特殊的时刻，作者想记录下他的每个家人，这个夏天他的长女 Michelle 将要自己开车了，Daivd 在高中过得很开心，Jackie 也再次露出微笑。另外这个夏天也是他在工业界奋斗 8 年后重返学术界的第一个十年。写作是他最深的愿望，这驱使着他在过去 20 年中，在看电视的同时阅读数学文献和图书。最后，他很怀念 2006 年去世的母亲，Suxiao Li。他的父亲 Dafu Qiu 住在中国。他的岳父母 Lumei Li 和 Jinxue Chen 多年来和他住在一起。他们的关爱以及每日的鼓励在他的生命中起到至关重要的作用。

前　言

本书主要研究认知无线电网络数据的分析和处理，但并不局限于此，在本书中我们还更多地讨论了更一般的数学模型，以及最新的研究进展。

所谓大数据，在最初就是指我们处理的较大规模或者较大量级的数据 [1]。而制约我们研究的最大问题也正是数据的量级。随着研究发展，人们对大数据的关注点从因果关系转变为相关性，也因此引入了“概率”的讨论。进一步，人们意识到大数据在某种意义下是“杂乱的”，正是因为其量级较大，我们因此可以不去苛求其精确性。处理大数据的数学模型除了预测我们需要的，还必须降低不精确带来的风险。

在此书的写作过程中，大数据被认为是在科学和工程领域里的一次转变，如图1所示。在 2011 年 11 月，当我们完成参考文献 [2] 的写作时，意识到了大数据的划时代意义。所以在此书的第 1 页和第 1 章，都用了大数据作为标题 (参考文献 [2] 的 1.1 节)，我们对此的理解是认知无线电网络频谱感知的研究自然会引导我们对大数据的关注。在过去的 18 个月里，更深刻地认识到了数学工具的重要性，尤其是高维大数据的重要性。并且可以预见到智能电网 [3] 将会是大数据理论的一个重要应用，特别是用到了本书所介绍的数学理论。这里需要注意的是，如果没有数据“高维”的假设，本书介绍的许多数学理论是不成立的，因此，本书大量采用非渐近的集中不等式作为主要工具。

如图1所示，大数据在认知网络测量、认知无线电、认知雷达和智能电网等相关领域研究中处于核心和基础的地位。有关智能电网的相关研究可参阅参考文献 [3]。尤为注意的是，高维数据统计分析是这些课题研究的出发点，而随机矩阵是建立大数据模型最基本的工具，另外测量集中性是高维空间中独有的现象。

为了更直观地理解本书，我们首先考虑以下一个基本的问题：已知一个数据集合，用矩阵表示为

$$\mathbf{X}=\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix} \in \mathbb{C}^{m\times n}$$

其中 X_{ij} 是随机变量，比如亚高斯随机变量。这里，m,n 是有限大的常数，例如，$m=100$，$n=100$。随着随机变量 $\mathbf{X}$ 的维度趋于无穷，其谱也趋于稳定。在过去几年的研究里，局部和非渐近条件下，我们一般认为 $\mathbf{X}$ 的维度是固定，而非趋于无穷的。此时，测量集中性的性质也自然而然就具备了。相应地，特征值 $\lambda_i\left(\mathbf{X}^T\mathbf{X}\right), i=1,\cdots,n$ 也就成为我们的主要研究对象。由于特征值可看做是利普希茨函数，进而可以用 Talagrand 集中不等式处理。其隐含

的意义是，大量随机变量的和高概率上是一个常数，并且我们可以在亚高斯随机变量的范畴里同时处理标准高斯和伯努利类型的随机变量。

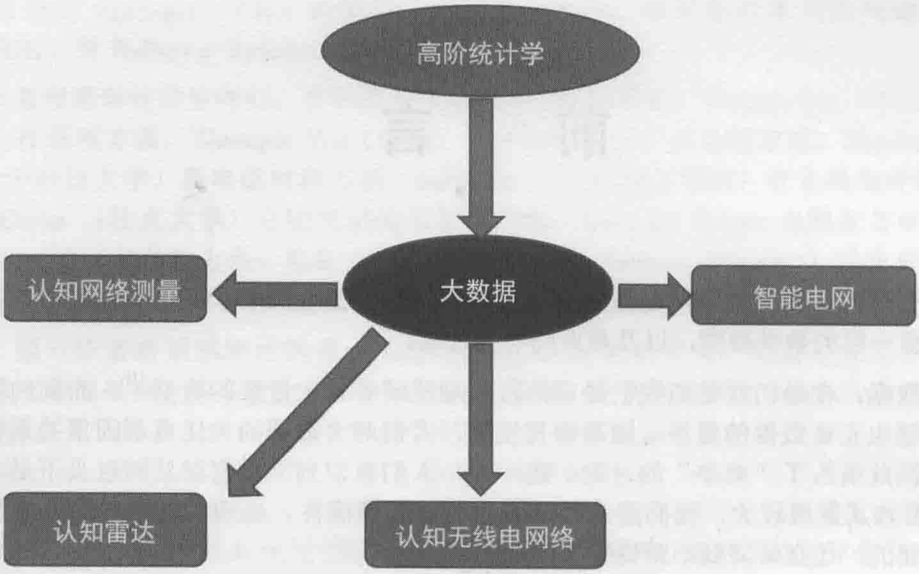


图 1 大数据图景

定理 0.0.1 (Talagrand 集中不等式) 对于每一个 $\{-1,1\}^n$ 空间上的积概率 $\mathbb{P}$，考虑一个凸的利普希茨函数 $f:\mathbb{R}^n\to\mathbb{R}$，其中利普希茨常数为 L。令 $X_1,...,X_n$ 表示取值在 $\{-1,1\}$ 上的独立随机变量，定义 $Y=f(X_1,...,X_n)$，而 $\mathbb{M}Y$ 为其中值，那么，对于任意的 $t>0$，我们有

$$\mathbb{P}\left(|Y-\mathbb{M}Y|\geqslant t\right)\leqslant 4e^{-t^2/16L^2} \tag{1}$$

且 Y 满足

$$\operatorname{Var}(Y)\leqslant 16L^2,\qquad \mathbb{E}[Y]-16L\leqslant \mathbb{M}[Y]\leqslant \mathbb{E}[Y]+16L \tag{2}$$

值得注意的是，对于任意给定的随机矩阵 $\mathbf{X}\in\mathbb{R}^{n\times n}$，以下均为其利普希茨函数：

$$(1)\lambda_{\max}(\mathbf{X});(2)\lambda_{min}(\mathbf{X});(3)\operatorname{Tr}(\mathbf{X});(4)\sum_{i=1}^{k}\lambda_i(\mathbf{X});(5)\sum_{i=1}^{k}\lambda_{n-i+1}(\mathbf{X})$$

其中 $\operatorname{Tr}(\mathbf{X})$ 具备 $L=1/n$ 的利普希茨常数，而 $\lambda_i(\mathbf{X}),i=1,...,n$ 的利普希茨常数为 $L=1/\sqrt{n}$。故 $\operatorname{Tr}(\mathbf{X})$ 的方差的上界为 $16/n^2$，而 $\lambda_i(\mathbf{X}),i=1,...,n$ 的方差的上界为 $16/n$，其相差 $1/n$ 倍。例如，当 $n=100$ 时，它们可相差 20 dB 之多，而在假设检验问题中，方差恰恰起到了极其重要的作用。

目 录

第一部分 理 论

第二部分 应 用