

信息与通信技术

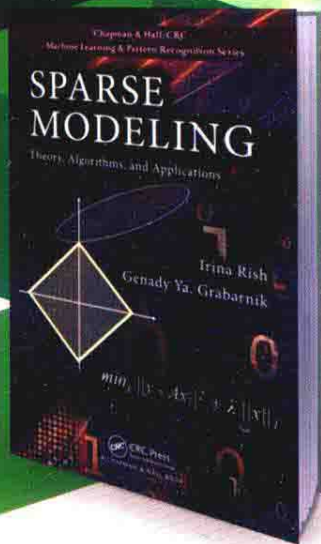
Sparse Modeling: Theory, Algorithms, and Applications

稀疏建模理论、 算法及其应用

Sparse Modeling: Theory, Algorithms,
and Applications

【美】 Irina Rish 著
Genady Ya. Grabarnik

栾悉道 王卫威 谢毓湘 魏迎梅 译



经典译丛·信息与通信技术

稀疏建模理论、算法及其应用

Sparse Modeling
Theory, Algorithms, and Applications

[美] Irina Rish 著
Genady Ya. Grabarnik

栾悉道 王卫威 谢毓湘 魏迎梅 译

电子工业出版社
Publishing House of Electronics Industry
北京·BEIJING

内 容 简 介

稀疏建模与现代统计学、信号处理、机器学习联系密切,可以实现利用相对较少的观测数据精确复原待估信号,广泛应用于图像重构、数据的参数学习模型、故障诊断、模式识别与雷达信号处理等领域。本书详细讨论了稀疏建模的相关内容,包括对稀疏解产生的问题描述、寻找稀疏解的求解算法、稀疏复原的理论成果以及应用实例等。

本书适合信息与信号处理、图像处理等专业的高校研究生以及科研机构相关研究人员使用。

Irina Rish, Genady Ya. Grabarnik: Sparse Modeling: Theory, Algorithms, and Applications

ISBN: 9781439828694

Copyright © 2015 by Taylor & Francis Group, LLC

Authorized translation from English language edition published by CRC Press, part of Taylor & Francis Group LLC; All rights reserved.

Publishing House of Electronics Industry is authorized to publish and distribute exclusively the Chinese (Simplified Characters) language edition. This edition is authorized for sale throughout Mainland of China. No part of the publication may be reproduced or distributed by any means, or stored in a database or retrieval system, without the prior written permission of the publisher.

Copies of this book sold without a Taylor & Francis sticker on the cover are unauthorized and illegal.

本书原版由 Taylor & Francis 出版集团旗下 CRC 出版公司出版,并经其授权翻译出版。版权所有,侵权必究。本书中文简体翻译版授权由电子工业出版社独家出版并限在中国大陆地区销售。未经出版者书面许可,不得以任何方式复制或发行本书的任何部分。

本书封面贴有 Taylor & Francis 公司防伪标签,无标签者不得销售。

版权贸易合同登记号 图字:01-2015-8700

图书在版编目(CIP)数据

稀疏建模理论、算法及其应用/(美)伊琳娜·里什(Irina Rish),(美)贾纳德里·亚·格拉巴尔尼克(Genady Ya. Grabarnik)著;栾悉道等译.—北京:电子工业出版社,2018.1

(经典译丛·信息与通信技术)

书名原文: Sparse Modeling: Theory, Algorithms, and Applications

ISBN 978-7-121-33356-9

I. ①稀… II. ①伊… ②贾… ③栾… III. ①数学模型—研究 IV. ①O141.4

中国版本图书馆 CIP 数据核字(2017)第 321951 号

策划编辑:马 岚

责任编辑:葛卉婷 特约编辑:姚 旭

印 刷:涿州市京南印刷厂

装 订:涿州市京南印刷厂

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:11.5 字数:261 千字 彩插:1

版 次:2018 年 1 月第 1 版

印 次:2018 年 1 月第 1 次印刷

定 价:49.00 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 zlt@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。

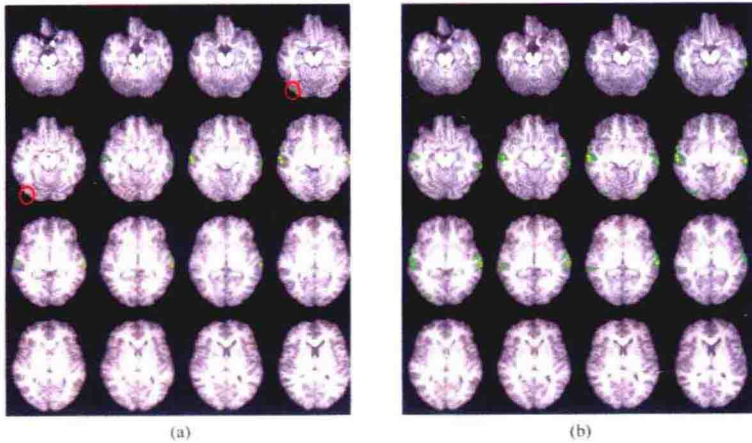


图 6.3 对于受试者 1(影像学视角), 该图显示了 PBAIC 数据集中“指令”目标变量的弹性网解(即线性模型的系数 x_i) 绝对值的脑部图像。将非零元素(激活变量)的数量固定为 1000。两幅图分别显示了(a) $\lambda_2 = 0.1$ (b)与 $\lambda_2 = 2$ 情况下的 EN 解(图)。当 λ_2 的值较大时, 非零体素的簇(Cluster)也较大, 包括部分但并非全部 $\lambda_2 = 0.1$ 时的非零体素簇。注意,(a)中高亮的(红色环)簇由具有 $\lambda_2 = 0.1$ 的 EN 所辨识, 但不能被具有 $\lambda_2 = 2$ 的 EN 所辨识

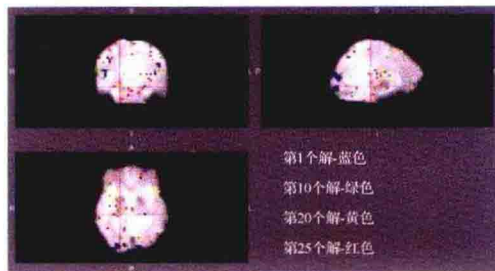
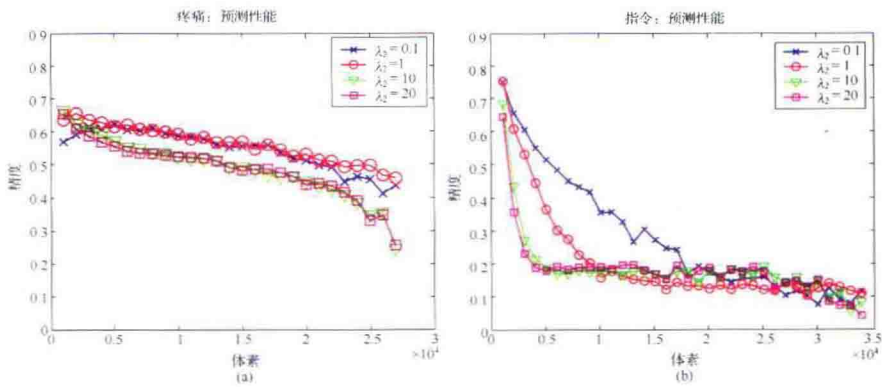
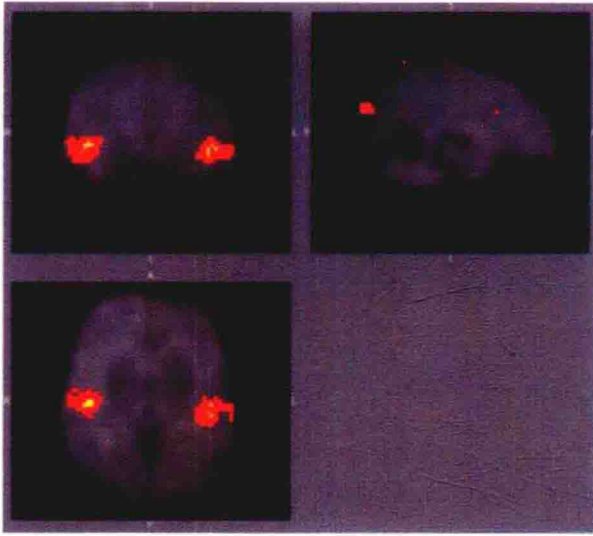
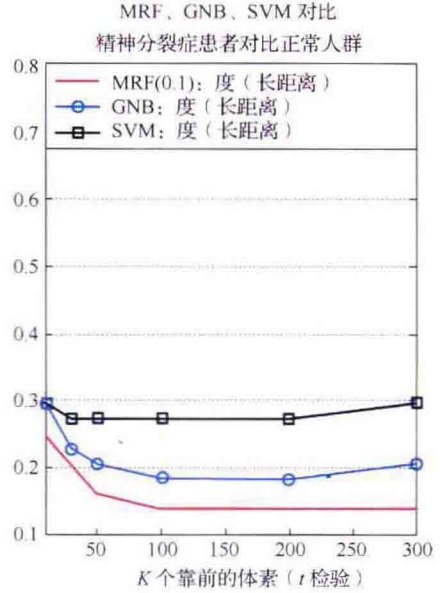


图 6.5 后续的“受限”弹性网解的预测精度。(a)疼痛感知;(b)PBAIC 中的“指令”任务。注意, 在疼痛预测情况下, 移除大量的预测体素后所得到的解中, 精度下降非常缓慢, 意味着与疼痛有关的信息高度散布于大脑中(同时可见图(c)中某些解的空间可视化)。在“指令”情况下可观测到与之相反的行为, 即前几个受限的解移除后, 精度出现急剧下降, 图 6.3 中显示了局部化的预测解



(a)



(b)

图 8.2 (a)FDR 校正的两样本 t 检验结果, 其中每一个体素的原假设都假定精神分裂症患者与正常人群之间没有区别。红/黄色表示在 α 水平下通过 FDR 校正的低 p 值区域(即 5% 的假阳性率)。注意, 正常人群这些体素的均值(标准化的)总是比精神分裂症人群要高。(b)利用 100 个排序靠前(最局区分性)的特征, 如泛函网络中的体素度, 高斯 MRF 分类器以 86% 的精度预测精神分裂症

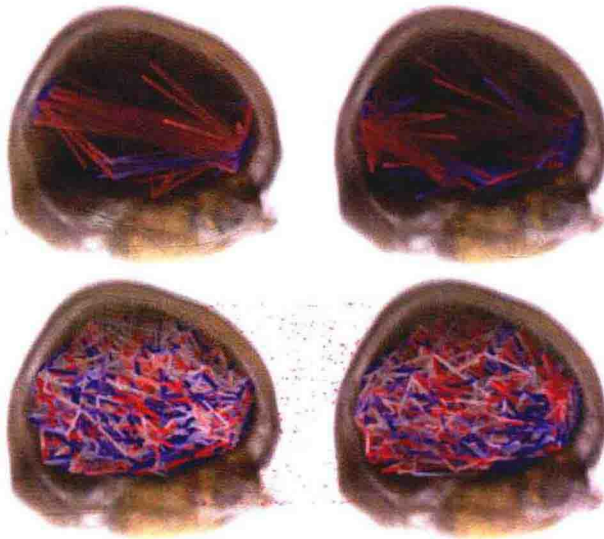


图 8.3 用于可卡因上瘾(左)与控制受试者(右)学习的结构, 顶部为借助 $l_{1,2}$ 方法进行变量选择的马尔可夫网络学习方法得到的结构, 底部为未进行变量选择的马尔可夫网络学习方面, 即标准图形化 LASSO 方法得到的结构。正交互作用显示为蓝色, 负交互作用显示为红色。注意, 顶部的结构(密度为 0.0016)比底部的结构(密度为 0.023)更为稀疏, 其中, 完全图中边的数量约为 378 000

译者序

信号稀疏表示从 20 世纪 90 年代起引起研究人员的广泛关注,成为信号处理与应用领域的热门研究问题。它通过挖掘待估信号中的稀疏先验信息构造正则化模型,能够在不完全观测下对待估信号进行高精度的重构。经过 20 多年的发展,信号稀疏表示在模型构建、求解算法与理论分析方面已形成了一套成熟的理论,并且在图像去噪与重构、多媒体处理、盲源分析、DOA 估计与医学影像复原等应用中发挥着重要作用。

本书的四位译者多年来从事信号处理领域的相关研究,一直希望能系统地总结信号稀疏表示的基础理论及其应用,因此共同翻译了此书。本书在电子工业出版社马岚老师的帮助下,顺利获得相关版权,在此特别致谢。

需要说明的是,本书的翻译得到了国家自然科学基金项目“非线性稀疏表示理论及其应用”(编号:61201337)、国家自然科学基金项目“视频内容重复检测及关联分析技术研究”(编号:61571453)、湖南省教育厅重点项目“视频内容重复检测技术研究”(编号:15A020)、湖南省教育科学规划课题“基于学习分析技术的大规模在线学习行为分析与教学策略研究”(编号:XJK17BXX010)和长沙市科技计划项目“网购平台图像快速匹配与检索系统及其应用研究”(编号:ZD1601014)的资助。

本书在翻译过程中,所有译者一直坚持忠于原书,以谨慎细致的态度开展工作,但是其中难免存在疏漏,恳请广大读者批评指正。

前 言

如果托勒密、阿加莎·克里斯蒂与奥卡姆的威廉聚在一起，他们很可能认同一个共同的思想。托勒密会说，“我们认为用最简单的假设对现象进行解释是一种很好的准则”。阿加莎可能会补充，“最简单的解释总是最适合的”。奥卡姆的威廉将可能点头同意，“如无必要，勿增实体”。该节省性原则，就是今天有名的奥卡姆剃刀原理，是渗透于从古至今所有哲学、艺术与科学领域的一个基础性思想。“至繁归于至简”（莱昂纳多·达·芬奇）。“尽量把所有事情变得简单，以致不能更简单”（阿尔伯特·爱因斯坦）。在人类历史上，先哲支持“简单性”的名言可以无休无尽，很容易写满许多页纸。但是，我们希望保持该序言简短。

该书的主题——稀疏建模，是节省性原则在现代统计学、机器学习与信号处理领域的特殊体现。在这些领域，一个基础性的问题就是由于观测成本或其他限制，需要从数量相对较少的观测中对未观测高维信号进行精确复原。图像重构、从数据中学习模型参数、系统故障或人类疾病诊断，是逆问题出现后要解决的一些例子。一般地，高维、小样本推断问题是欠定的，且在计算上是难于处理的，除非该问题具有某一特定的结构，如稀疏性。

事实上，当仅有少量变量为真正重要的变量时，真实解可以很好地由稀疏向量来近似，将剩余变量设置为零或接近零。换言之，少量最相关的变量（起因、预测因子等）通常对于解释感兴趣的现象来说是充分的。更一般地，即使原始问题并没有产生稀疏解，我们也可以找到一个到新坐标系统的映射或字典，从而实现稀疏表示。因此，稀疏结构看上去是很多自然信号固有的性质——没有该结构，认知并适应这个世界是相当具有挑战性的。

本书提供对稀疏建模简要的介绍，包括应用实例、导致稀疏解的问题描述、寻找稀疏解的算法，以及一些关于稀疏复原最新的理论成果。该书的内容基于我们几年前在国际机器学习大会(ICML-2010)上的辅导性讲座，以及2011年春季学期在哥伦比亚大学教授的研究生课程。

第1章从引导性示例开始，对稀疏建模关键的最新进展进行了概述。第2章对优化问题进行了描述，该优化问题涉及常用于强化稀疏的工具，如 l_0 与 l_1 范数约束。必要的理论结果在第3章与第4章中进行介绍，第5章讨论了若干用于寻找稀疏解的著名算法。然后，在第6章与第7章讨论了大量的稀疏复原问题，分别将基本的问题形式扩展到更为复杂的结构性稀疏形式与不同的损失函数。第8章介绍了特殊的稀疏图模型，如稀疏高斯马尔可夫随机场，该模型是稀疏建模中热门且快速发展的子领域。最后，第9章研究了字典学习与稀疏矩阵分解。

注意，本书并不能对所有与稀疏有关的最新进展进行全部研究。事实上，仅仅一本书

不可能对这个快速发展的领域全面涉猎。然而，我们希望本书能够作为稀疏建模的入门书籍，激励读者继续学习本书之外的内容。

最后，我们想感谢以不同方式对本书作出贡献的人。Irina 感谢 IBM 沃特森研究中心的同事 Chid Apte、Guillermo Cecchi、James Kozloski、Laxmi Parida、Charles Peck、Ravi Rao、Jeremy Rice 与 Ajay Royyuru，感谢他们这些年来给予的鼓励与支持。同时，其他同事与朋友的想法也有助于本书的成稿，包括 Narges Bani Asadi、Alina Beygelzimer、Melissa Carroll、Gaurav Chandalia、Jean Honorio、Natalia Odintsova、Dimitris Samaras、Katya Scheinberg 与 Ben Taskar。Ben 于去年去世，但他仍然活在我们的记忆与他优秀的工作中。

感谢 Dmitry Malioutov、Aurelie Lozano 与 Francisco Pereira，他们阅读了手稿，并提出了很多有价值的建议，对本书改进帮助很大。还要特别感谢本书的编辑 Randi Cohen，他使我们一直保持积极性并耐心地等待本书完稿。最后，感谢我们的家人，是他们的爱、支持与耐心成为我们灵感的无限源泉。我们不得不承认该书花费了比预期长的时间(多了几年)。因此，Irina(很高兴地)输掉了与她女儿 Natalie 关于谁先出版一本书的赌约。

目 录

第 1 章 导论	1
1.1 引导性示例	3
1.1.1 计算机网络诊断	3
1.1.2 神经影像分析	4
1.1.3 压缩感知	6
1.2 稀疏复原简介	7
1.3 统计学习与压缩感知	9
1.4 总结与参考书目	9
第 2 章 稀疏复原: 问题描述	11
2.1 不含噪稀疏复原	11
2.2 近似	13
2.3 凸性: 简要回顾	13
2.4 问题(P_0)的松弛	14
2.5 l_q -正则函数对解的稀疏性的影响	15
2.6 l_1 范数最小化与线性规划的等价性	16
2.7 含噪稀疏复原	17
2.8 稀疏复原问题的统计学视角	20
2.9 扩展 LASSO: 其他损失函数与正则函数	22
2.10 总结与参考书目	24
第 3 章 理论结果(确定性部分)	26
3.1 采样定理	26
3.2 令人惊讶的实验结果	27
3.3 从不完整频率信息中进行信号复原	29
3.4 互相关	30
3.5 Spark 与问题(P_0)解的唯一性	32
3.6 零空间性质与问题(P_1)解的唯一性	34
3.7 有限等距性质	35
3.8 最坏情况下精确复原问题的平方根瓶颈	36

3.9	基于 RIP 的精确重构	37
3.10	总结与参考书目	40
第 4 章	理论结果(概率部分)	41
4.1	RIP 何时成立?	41
4.2	Johnson-Lindenstrauss 引理与亚高斯随机矩阵的 RIP	42
4.2.1	Johnson-Lindenstrauss 集中不等式的证明	42
4.2.2	具有亚高斯随机元素的矩阵的 RIP	43
4.3	满足 RIP 的随机矩阵	46
4.3.1	特征值与 RIP	46
4.3.2	随机向量, 等距随机向量	47
4.4	具有独立有界行的矩阵与具有傅里叶变换随机行的矩阵的 RIP	47
4.4.1	URI 的证明	50
4.4.2	一致大数定律的尾界	52
4.5	总结与参考书目	54
第 5 章	稀疏复原问题的算法	55
5.1	一元阈值是正交设计的最优方法	55
5.1.1	l_0 范数最小化	56
5.1.2	l_1 范数最小化	57
5.2	求解 l_0 范数最小化的算法	58
5.2.1	贪婪方法综述	61
5.3	用于 l_1 范数最小化的算法	63
5.3.1	用于求解 LASSO 的最小角回归方法	63
5.3.2	坐标下降法	66
5.3.3	近端方法	67
5.4	总结与参考书目	72
第 6 章	扩展 LASSO: 结构稀疏性	73
6.1	弹性网	73
6.1.1	实际中的弹性网: 神经成像应用	77
6.2	融合 LASSO	83
6.3	分组 LASSO: l_1/l_2 罚函数	84
6.4	同步 LASSO: l_1/l_∞ 罚函数	85
6.5	一般化	86
6.5.1	块 l_1/l_q 范数及其扩展	86
6.5.2	重叠分组	87

6.6	应用	88
6.6.1	时间因果关系建模	88
6.6.2	广义加性模型	88
6.6.3	多核学习	89
6.6.4	多任务学习	90
6.7	总结与参考书目	91
第7章	扩展 LASSO: 其他损失函数	92
7.1	含噪观测情况下的稀疏复原	92
7.2	指数族、GLM 与 Bregman 散度	93
7.2.1	指数族	94
7.2.2	广义线性模型	95
7.2.3	Bregman 散度	96
7.3	具有 GLM 回归的稀疏复原	98
7.4	总结与参考书目	104
第8章	稀疏图模型	106
8.1	背景	106
8.2	马尔可夫网络	107
8.2.1	马尔可夫性质: 更为仔细的观察	108
8.2.2	高斯 MRF	110
8.3	马尔可夫网络中的学习与推断	110
8.3.1	学习	110
8.3.2	推断	111
8.3.3	例子: 神经影像应用	111
8.4	学习稀疏高斯 MRF	115
8.4.1	稀疏逆协方差选择问题	116
8.4.2	优化方法	117
8.4.3	选择正则化参数	123
8.5	总结与参考书目	126
第9章	稀疏矩阵分解: 字典学习与扩展	128
9.1	字典学习	129
9.1.1	问题描述	129
9.1.2	字典学习算法	130
9.2	稀疏 PCA	132
9.2.1	背景	132

9.2.2 稀疏 PCA: 合成视角	134
9.2.3 稀疏 PCA: 分析视角	136
9.3 用于盲源分离的稀疏 NMF	136
9.4 总结与参考书目	138
后记	140
附录 A 数学背景	141
参考文献	154

第1章 导论

如何从有限的观测中推断未被观测到的高维“世界状态”(state of the world), 这个问题经常出现在广泛的实际应用中。例如, 寻找基因中引发某疾病的子集; 定位与某一心理状态存在关联的大脑区域; 诊断大规模分布式计算机系统性能瓶颈; 使用压缩观测值重构高质量的图像等。更一般的例子包括, 从任意一类信号的含噪编码中对信号解码, 以及在高维但小样本的统计情况下估计模型参数。

图 1.1 描述的就是这种基本的推断问题, 其中 $\mathbf{x} = (x_1, \dots, x_n)$ 和 $\mathbf{y} = (y_1, \dots, y_m)$ 分别代表一个未被观测的 n 维的世界状态和它的 m 个观测结果。观测结果的输出向量 \mathbf{y} 可以看成是输入向量 \mathbf{x} 的一个含噪函数(编码)。一种常用的推断(解码)方法是, 给定观测结果 \mathbf{y} , 找到使某种损失函数 $L(\mathbf{x}; \mathbf{y})$ 最小化的 \mathbf{x} 。例如, 常见的极大似然法, 就是旨在找到一个使观测结果的似然 $P(\mathbf{y}|\mathbf{x})$ 最大化, 即使得负对数似然损失最小化的参数向量 \mathbf{x} 。

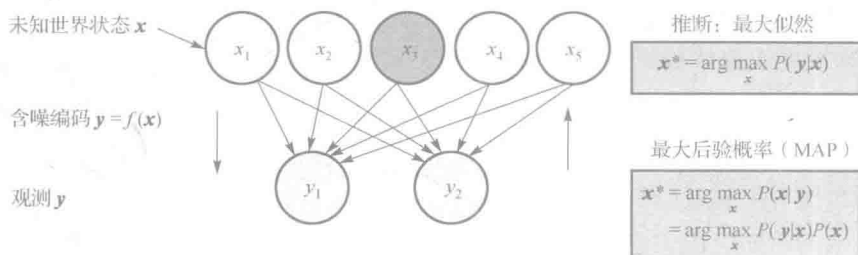


图 1.1 是否能够从低维、含噪的观测结果 \mathbf{y} 中复原高维信号 \mathbf{x} ? 出人意料的是, 答案是肯定的, 不过前提是 \mathbf{x} 具有某种特定的结构, 如(充分的)稀疏度, 并且映射 $\mathbf{y} = f(\mathbf{x})$ 保留了足够的信息来重构 \mathbf{x}

然而, 在许多实际问题中, 未被观测到的变量在数量上远远多于观测值, 因为对后者的观测可能很耗费成本, 同时还受到具体问题的限制。例如, 在计算机网络诊断、基因网络分析和神经影像应用中, 未知变量(如网络元素、基因或大脑体系素的状态)的总数可以达到数千个, 甚至数十万个, 相比之下观测结果或者样本的总数通常只有数百个。因此, 上述最大似然表达式变为欠定情况。同时, 为了限制可能的解空间, 必须引入额外的反映特定域性质或假设的正则化约束。从贝叶斯概率的角度来说, 正则化可以看作是对未知参数 \mathbf{x} 施加先验 $P(\mathbf{x})$, 然后再最大化后验概率 $P(\mathbf{x}|\mathbf{y}) = P(\mathbf{y}|\mathbf{x})P(\mathbf{x})/P(\mathbf{y})$, 这一问题将在下一章中详细讨论。

或许, 关于问题结构所做的最简单、最常见的假设之一是解为稀疏的。换言之, 通常假定在特定情况下只有一个相对较小的变量子集是真正重要的。例如, 在一个系统中只出现少量的并发故障; 只需要少量非零傅里叶系数就能满足对各种不同信号类型的准确表示; 通常, 只有一小部分预测变量(如基因)与响应变量(疾病或特征)最相关, 学习某一精确的预测模型也只需要一小部分预测变量。在所有这些例子中, 我们所寻求的解可以看成

是一个仅有少量非零坐标的稀疏高维向量。这一假设与哲学中的节省性原则一致，而这一原则通常称为奥卡姆剃刀，由中世纪著名哲学家奥卡姆的威廉提出，但或许可以追溯至时代更久远的亚里士多德和托勒密。在奥卡姆之后还出现了许多关于节省性原则的阐述，其中包括艾萨克·牛顿的著名表述：“寻求自然事物的本因，只须找到真实而又足以解释其现象的即可，无须更多。”

融合了节省性假定的统计模型称为稀疏模型。这种模型在科学应用中特别有用，例如发现基因数据或神经影像数据中的生物标志物，其中预测模型的可解释性(如辨别最相关的预测因子)是最本质的。另一个受益于稀疏性的重要领域是信号处理，信号处理的目标是在达到高重构精度的同时最小化信号采集成本；后文也将说明，对稀疏性的利用可以极大地提高信号处理中的成本效率。

从历史角度看，稀疏信号复原问题的描述可以追溯至1943年，或者可能更早。当时首先提出了组合分组测试问题(Dorfman, 1943)。这一问题背后的原始动力是为了设计一个高效的测试方案，从一大批人(如100 000人左右)的血液样本中识别出人数相对较少的那一部分感染者(如10人左右)。如果一一检测每个人的血液样本则成本太高，因此可以对受试者进行分组，将每组的血液样本进行组合，对组合样本进行测试就可以知道该组中是否至少有一人患有疾病。按照图1.1中的推断方案，可以将第 i 个人的健康状况表示为布尔变量 x_i ，如果此人身体健康则 $x_i = 0$ ，否则 $x_i = 1$ 。一组受试者 G_j 的测试结果或观测结果 y_j 是关于该组中变量的逻辑“或”函数。即当且仅当所有的 $x_i = 0, i \in G_j$ 时， $y_j = 0$ ；否则， $y_j = 1$ 。给定人群中患病者数量的上界，即 \mathbf{x} 稀疏度的限制，分组测试的目标是以最少的检测次数识别出所有患病者(即非零 x_i)。

许多其他的诊断应用中也出现了类似的问题表述，例如，在计算机网络故障诊断中，网络节点(如路由器或链路)要么运转正常，要么出现故障，分组测试相对应的是端到端处理(也称为网络探测)，其检查路由选择表确定的特定元素子集(Rish et al., 2005)(下一节将更为详细地探讨网络诊断问题，不过将专注于连续型情况而非布尔型情况，其中“硬故障”将被放宽为性能瓶颈或时间延迟)。总的来说，分组测试历史悠久，在各种各样的实际问题上都已得到成功应用，包括DNA库筛选、多址接入控制协议和数据流等诸多例子。关于分组测试的更多细节，可以参考(Du and Hwang, 2000)这本经典专著以及其中列举的参考文献，另外还有近年发表的文献(Gilbert and Strauss, 2007; Atia and Saligrama, 2012; Gilbert et al., 2012)。

在组合分组测试领域兴起半个世纪后，即过去几十年中，稀疏信号复原受到了新的广泛关注。现在，对稀疏信号复原的研究主要专注于连续信号和观测，并且发展出了强化稀疏性的独特方式，如利用 l_1 范数正则化。1986年，针对带限反射地震图的线性反演(反卷积)，提出了基于 l_1 范数的优化方法(Santosa and Symes, 1986)。1992年，针对图像处理中的去噪问题提出了与 l_1 范数紧密相关的总变分正则函数(Rudin et al., 1992)。1996年，统计学文献中出现了一篇影响重大的关于LASSO(最小绝对值收缩与选择算子)或者说 l_1 范数正则化线性回归的论文(Tibshirani, 1996)，引领了今天稀疏回归在一系列广泛的实际问

题上的主流应用。大约同一时间,信号处理领域也出现了本质上与 LASSO 等价的基追踪方法(Chen et al., 1998),并且(Candès et al., 2006a)和(Donoho, 2006a)的突破性理论成果引起了压缩感知的兴起,相比标准的香农-奈奎斯特采样定理,压缩感知对精确且计算效率高的稀疏信号复原所需要的观测次数呈指数降低。近年来,在信号处理及相关领域,压缩感知吸引了大量的关注,并且产生了一系列理论成果、算法和新兴应用。

本书主要聚焦于连续稀疏信号,并跟踪现代稀疏统计建模和压缩感知的发展。显然,本书无法涵盖这些发展迅速的领域的所有方面。因此,我们的目标是提供对关键概念合理的介绍,并总结近期稀疏建模和信号复原领域的主要成果,如稀疏回归、稀疏马尔可夫网络和稀疏矩阵分解中的常见问题形式,以及稀疏建模的基本理论,最新的算法与一些实际应用等。本书会从一些启发性实际问题的综述开始,引出对稀疏信号复原的描述。

1.1 引导性示例

1.1.1 计算机网络诊断

分布式计算机系统和网络管理中的核心问题之一,是对各种故障和性能退化进行快速、实时的诊断。然而,在大规模系统中,要监测每一个组件(即每一个网络链路、每一个应用、每一个数据库事务等)需要耗费巨大的成本,甚至是不可行的。一种可供选择的方法是利用端到端处理或探测,如 ping 和 traceroute 命令,或者利用端到端应用级测试,来收集相对较少的总体性能观测数据,然后再推断出每个组件的状况。在系统管理领域中,专注于通过间接观测来诊断网络问题的研究方向称为网络断层扫描,与医学影像技术类似,只是后者从不同器官的层析成像的图像来作出推断,并以此诊断健康问题。

特别地,考虑网络性能瓶颈的识别问题,如识别导致异常的端到端延迟的网络链路。这一问题在(Beygelzimer et al., 2007)等书中有所讨论。假定 $\mathbf{y} \in \mathbb{R}^m$ 是一个观测到的端到端处理延迟向量, $\mathbf{x} \in \mathbb{R}^n$ 是一个未观测到的链路延迟向量, \mathbf{A} 是一个路由矩阵,如果端到端测试 i 经过链路 j ,那么 $a_{ij} = 1$,否则为 0。这一问题如图 1.2 所示。通常,假定端到端延迟满足以下含噪线性模型:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \boldsymbol{\varepsilon}. \quad (1.1)$$

其中, $\boldsymbol{\varepsilon}$ 为观测噪声,反映了除链路延迟之外,其他造成端到端延迟的潜在原因,以及可能的非线性影响。重构 \mathbf{x} 的问题可以看成是一个普通最小二乘(OLS)回归问题,其中 \mathbf{A} 为设计矩阵, \mathbf{x} 为通过最小化最小二乘误差获得的线性回归系数,即最小化最小二乘误差等价于,在高斯噪声 $\boldsymbol{\varepsilon}$ 的假定条件下最大化条件对数似然 $\log P(\mathbf{y}|\mathbf{x})$,

$$\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2$$

因为测试次数 m 通常比组件数量 n 小很多,所以重构 \mathbf{x} 的问题是欠定的,即不存在唯一解。因此,需要施加一些正则化约束。对于网络性能瓶颈诊断问题,有理由期望在任意特定时刻,造成事务延迟的故障链路只有很小一部分,而其余的大多数链路都运转正常。换句话说,可以假定能够通过一个稀疏向量很好地近似 \mathbf{x} ,其中仅有一小部分坐标具有较大

的值。本书稍后几章将集中讨论如何在上述问题中强化稀疏性，以及从少量观测值中进行稀疏解复原的问题。

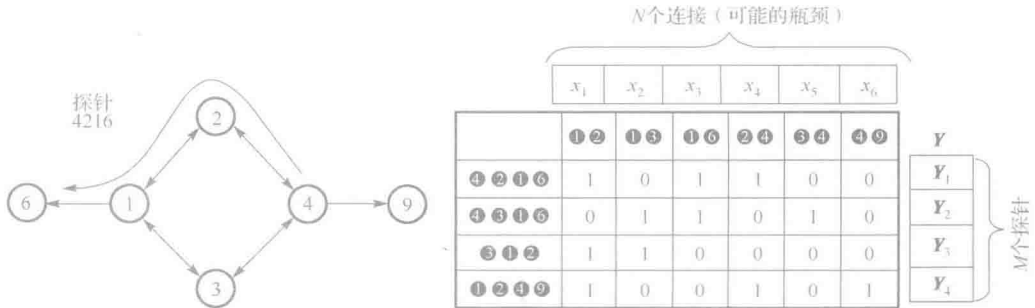


图 1.2 稀疏信号复原示例：利用端到端测试手段或探测来诊断计算机网络中的性能瓶颈

1.1.2 神经影像分析

现在，我们展示一个不同类型的应用实例，这一例子来源于医学影像领域。具体来说，就是根据大脑影像数据，如功能性磁共振成像 (fMRI)，来预测人的心理状态问题。过去十年中，人们对基于神经影像的心理状态预测进行了大量的研究，这一领域融汇了统计学、机器学习和神经科学的知识。心理状态可以是理性的，如查看一幅图或阅读一句话 (Mitchell et al., 2004)；也可以是感性的，如玩虚拟现实电子游戏时感到开心、焦虑或烦恼 (Carroll et al., 2009)。其他的例子还包括预测一个人感受到的疼痛程度 (Rish et al., 2010; Cecchi et al., 2012)；或者学习一种能够识别特定精神疾病的分类模型，具体的疾病包括精神分裂症 (Rish et al., 2012a)、老年痴呆症 (Huang et al., 2009) 或药物成瘾 (Honorio et al., 2009)。

在典型的“测心术”fMRI 试验中，受试者执行一项特定的任务或暴露在某种刺激当中，此时磁共振扫描器记录受试者的血氧水平依赖 (BOLD) 信号，这种信号反映整个大脑中神经活动的变化。所产生的某时间段的全脑扫描与任务或刺激相关联，并形成了一系列三维图像，每个图像大约有 10 000 ~ 100 000 个子卷或体素，而时间点的个数或者重复时间 (TR) 一般只有数百个。

如前所述，典型的试验范例的目标是理解与特定任务或刺激相关的心理状态的变化，而现代多元 fMRI 分析的核心问题之一是能否在给定一系列大脑图像的情况下预测这些心理状态。例如，在一项近期的疼痛感知研究中 (Baliki et al., 2009)，受试者的背部通过接触式探头与快速变化的热源相连接，受试者依据热源的变化在一个连续尺度上对他们的痛感程度进行评定。在另一项 2007 匹兹堡脑部行为解译比赛 (匹兹堡 EBC 小组, 2007) 有关的试验中，任务是预测受试者在玩电子游戏时的心理状态，包括感觉烦恼或焦虑、听从指令、观测别人的表情或完成一项游戏内的特定任务等。

给定一个 fMRI 数据集，即所有体素的血氧水平依赖信号 (体素活性) 的时间序列，以及对应代表任务或刺激的时间序列，便可以将预测任务构造为一个线性回归问题，其中，各

个时间点将被视作独立同分布的样本。当然，这是一个根本不现实的简单化假设，不过出人意料的是，这一假设非常适用于预测目的。体素活性级别对应于预测因子，而心理状态、任务或刺激则是被预测的响应变量。更具体地说，用 A_1, \dots, A_n 表示 n 个预测因子的集合， Y 代表响应变量， m 是样本的数量；那么就有 $A = (a_1 | \dots | a_n)$ 对应于一个 $m \times n$ 维 fMRI 数据矩阵，其中，对于所有 m 个实例，每个 a_i 都是第 i 个预测因子值的 m 维向量，而 m 维向量 y 对应于响应变量 Y 的值。这一点如图 1.3 所示。

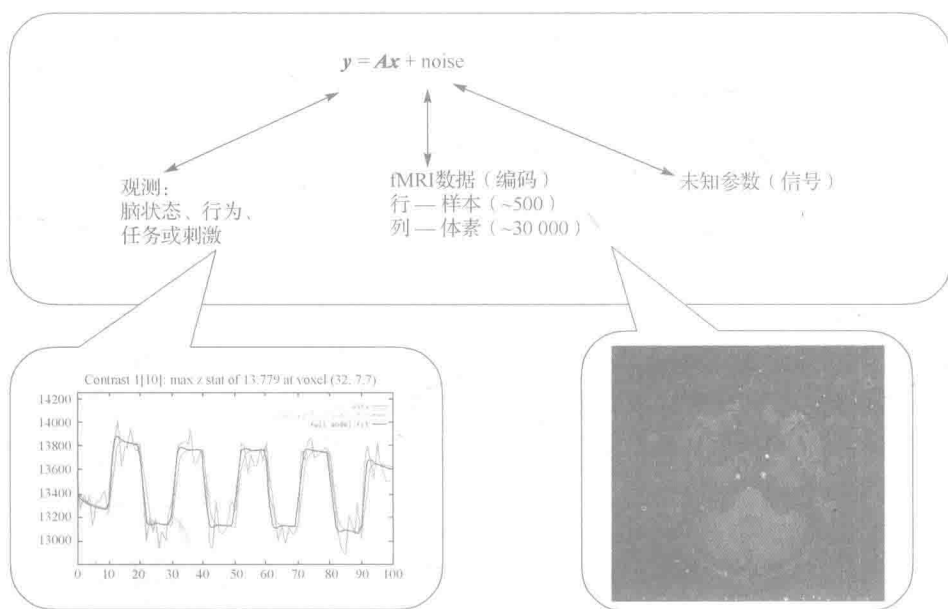


图 1.3 从 fMRI 数据中预测心理状态，视为同步变量选择的线性回归。目标是找到一个 fMRI 体素的子集，它表示与特定心理状态最相关（如最具预测性）的大脑区域

前面已经提到，在生物应用（包括神经影像）中，统计模型的可解释性常常与该模型的预测性同样重要。要改进模型的可解释性，一个常用的方法是变量选择，即选择一个与响应变量最为相关的小的预测变量子集。在上述神经影像应用中，最关键的目标之一是找到与给定任务、刺激或心理状态最为相关的大脑区域。而且，变量选择以及更为一般的降维方法能够避免对高维小样本数据过拟合（这样的数据在 fMRI 和其他生物应用中很常见），从而显著改善模型的泛化精度。

变量选择的一个简单方法是对每个预测变量进行独立评估，评估时使用一元的相关性测度，如变量和响应之间的相关性，或者两者之间的互信息。这一方法在机器学习领域称为基于滤波的方法。举个例子，广义线性模型（GLM）的传统的 fMRI 分析方法（Friston et al., 1995）可以看成是基于滤波的变量选择，因为该方法本质上是计算每个体素和任务或刺激之间的个体相关性，然后识别相关性超出特定阈值的那些大脑区域。然而，这样利用大量一元的方法虽然非常简单，但却具有明显的弊端，即完全忽视了多元交互，并可能导致