统计稀疏学习中的贝叶斯非参数建模方法及其应用研究

TONGJI XISHU XUEXIZHONG DE BEIYESI FEICANSHU JIANMO FANGFA JIQI YINGYONG YANJIU

何岩 著

浙江工商大學出版社

统计稀疏学习中的贝叶斯非参数建模方法及其应用研究

TONGJI XISHU XUEXIZHONG DE BEIYESI FEICANSHU JIANMO FANGFA JIQI YINGYONG YANJIU

何岩 著

浙江工商大学出版社

图书在版编目(CIP)数据

统计稀疏学习中的贝叶斯非参数建模方法及其应用研究／何岩著．—杭州：浙江工商大学出版社，2014.4

ISBN 978-7-5178-0188-7

Ⅰ．①稀… Ⅱ．①何… Ⅲ．①贝叶斯估计—数学模型—研究 Ⅳ．①O211.67

中国版本图书馆 CIP 数据核字(2013)第 319128 号

统计稀疏学习中的贝叶斯非参数建模方法及其应用研究

何　岩　著

责任编辑　王玲娜　刘　韵
责任印制　包建辉
出版发行　浙江工商大学出版社
（杭州市教工路 198 号　邮政编码 310012）
（E-mail:zjgsupress@163.com）
（网址:http://www.zjgsupress.com）
电话:0571-88904980,88831806(传真)
排　　版　杭州朝曦图文设计有限公司
印　　刷　杭州杭新印务有限公司
开　　本　710mm×1000mm　1/16
印　　张　6
字　　数　114 千
版 印 次　2014 年 4 月第 1 版　2014 年 4 月第 1 次印刷
书　　号　ISBN 978-7-5178-0188-7
定　　价　18.00 元

浙江工商大学出版社营销部邮购电话　0571-88804228

目　录
CONTENTS

第1章
绪　　论

1.1 研究背景和意义

统计学习(Statistical Learning)是一种专门研究小样本情况下机器学习规律的理论，在这种体系下的统计推理规则不仅考虑了对渐近性能的要求，而且追求在现有有限信息的条件下得到最优结果。近年来，统计学习领域的学者结合稀疏特性对传统统计学习理论和方法进行了丰富和拓展，基于稀疏的统计学习逐步成为统计学习与信息处理的重要研究方向，其在数据挖掘、内容检索、基因数据分析等诸多领域得到了广泛应用。

1.1.1 稀疏编码的生物感知基础

对于稀疏的研究，最早源于对神经科学和脑科学认知的研究成果。1954 年，Attneave 最先提出视觉感知的目标就是产生一个外部输入信号的有效表示。Barlow 在 1961 年基于信息论提出了“有效编码假设”，认为初级视觉皮层神经细胞的主要功能是去除输入刺激的统计相关性。20 世纪 60 年代末期，神经生理研究已表明了初级视觉皮层下细胞的感受野具有显著的方向敏感性，单个神经元仅对某一频段的信息呈现较强的反映，如特定方向的边缘、线段、条纹等图像特征，其空间感受野被描述为具有局部性、方向性和带通特性的信号编码滤波器，而每个神经元对这些刺激的表达则采用了稀疏编码(Sparse Coding)原则，将图像在边缘、端点、条纹等方面的特性以稀疏编码的形式进行描述。1996 年，Olshausen 和 Field 在 *Nature* 上发表论文，指出自然图像经过稀疏编码后得到的基函数类似于 V1 区内简单细胞感受野的反应特性。这种稀疏编码模型提取的基函数首次成功模拟了 V1 区内简单细胞感受野的三个响应特性：空间域的局部性、时域和频域的方向性和选择性。考虑到基函数的过完备性(基函数维数大于输出神经元的个数)，

Olshausen 和 Field 在 1997 年提出了一种超完备基的稀疏编码算法，利用基函数和系数的概率密度模型成功地模拟了 V1 区简单细胞感受野。

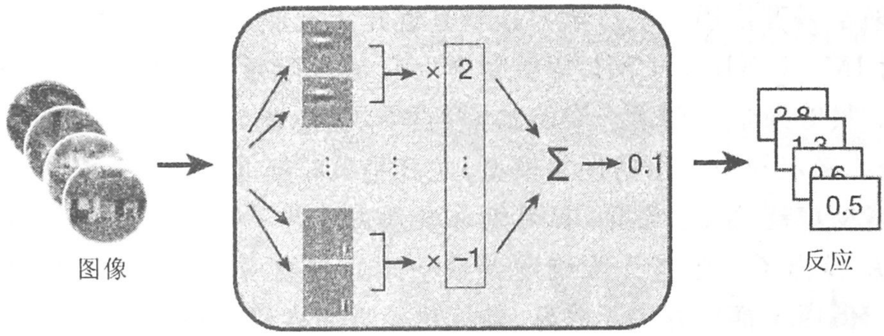
近年来，人们从神经生物学机理模型和计算机科学可计算模型等角度对稀疏编码理论进行了广泛的研究，并对生物视觉、脑科学的发展产生了重要的影响。Kay K. N. 和 Gallant 等从神经生理学机制上揭示了稀疏表达作为一种广泛的视觉先验，精确地定位于人类大脑视觉皮层多个功能区（如 V1、V2 区），并在视觉认知和推理过程中发挥着重要作用，例如图 1.1 显示 Kay K. N. 等人对图像识别的建模过程。这个过程的第一阶段是模型估计，即对每个测试者观看一组自然图像时产生的功能磁共振成像（FMRI）数据进行记录，再根据这些数据为每类图像构建一个定量的感受野模型，称为相对感受野模型（receptive-field model）。该模型基于 Gabor 滤波金字塔，并依照细胞感受野的三个特性进行描述。第二个阶段是图像识别，让每个测试者观看另外一组与先前测试图片不同的自然影像，并记录当时的功能磁共振成像数据。然后通过第一阶段构建的相对感受野模型来计算这组自然图像，预测每一张图片的功能磁共振成像数据，将预测数据与实际测量数据相对比，选取最相近的预测数据，从而得到测试者观看的图片。这些研究强调人类的认知和推理过程，不仅需要依据完整的信息输入，更需要依据视觉输入中的很少一部分典型特征，即依据某种稀疏编码求解，这为解决视觉认知问题提供了重要的生理学模型借鉴。

1.1.2 稀疏编码的信号表达基础

稀疏作为一种重要的数据编码与表达方式，不仅在人类的视觉认知机理上具有明确的理论依据，而且在信号表达与重建的理论方面得到了严格证明和推导。Donoho，Tao，Candès 和 Baraniuk 等提出的压缩感知（Compressive Sensing，CS）理论，从信号表达的角度证明了稀疏表达是高维信号（比如音频、视频等）在特定基向量（比如傅里叶基、小波基等）或“字典”上的一种自然表达。可压缩信号的少量随机线性投影即包含了重构和处理的足够信息，利用信号可压缩的先验知识和少量全局的线性测量可以获得精确的信号重建。在压缩感知理论基础上发展的约束优化求解策略为信号的稀疏表达提供了近似最优的可计算模型。同时，通过学习生成的自适应过完备冗余字典对稀疏表示求解的促进作用，引发研究者对字典学习算法的大量研究。2008 年 Candès 证明了如果随机正交模型条件成立，则能够以高概率恢复稀疏矩阵，从而从理论上证明矩阵填充（Matrix Completion）的可解性。基于上述理论的证明，目前统计稀疏学习已经广泛应用在信号压缩、图像处理、模式识别、机器学习等领域。

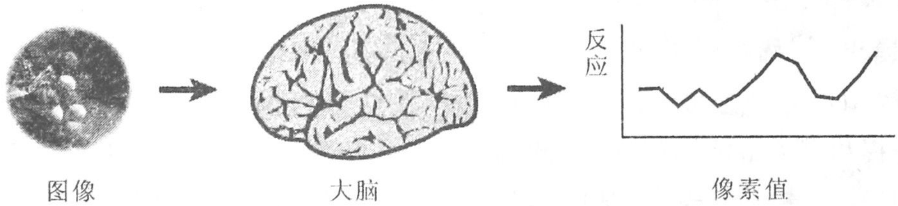
第一步，模型估计

为图像的每个像素估计一个感受野模型。

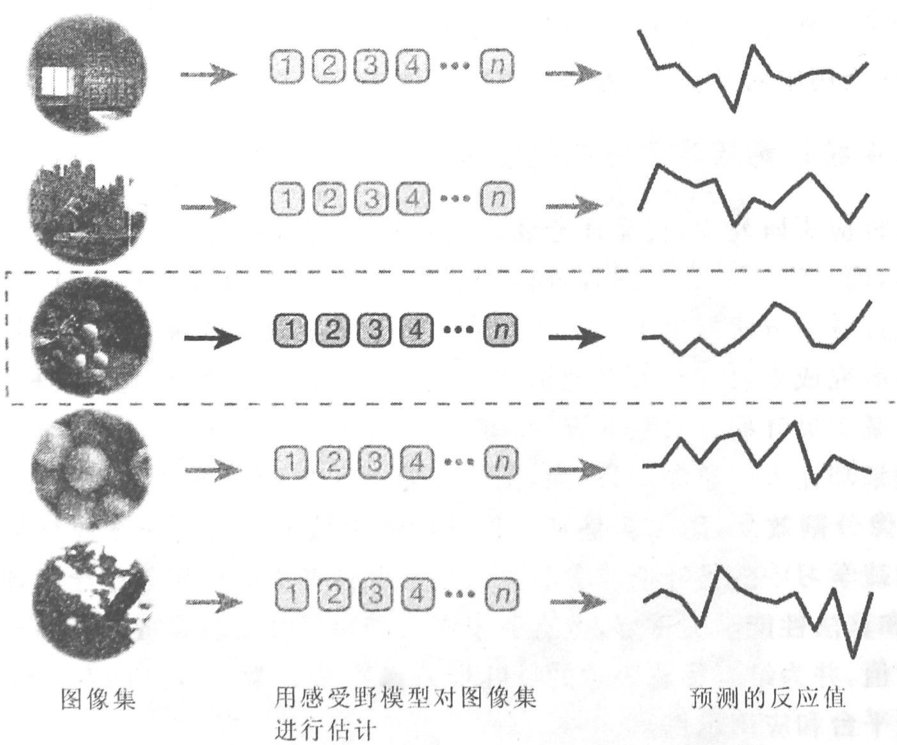


第二步：定义图像

（1）每个图像度量大脑的反应。



（2）用感受野模型为图像集预测大脑反应。



（3）选择与测量的大脑反应值最接近的预测值进行标记。

图 1.1 Kay K. N. 对图像识别的建模过程

1.1.3 贝叶斯非参数方法

贝叶斯学习理论用于统计学习领域是近几年发展起来的最重要的主流研究方向，是目前 JMLR，NIPS，ICML 等机器学习领域国际重要期刊与会议论文的热点讨论内容。贝叶斯学习理论将先验知识与样本信息相结合、依赖关系与概率表示相结合，是不确定知识表示的理想模型，尤其是贝叶斯非参数方法所表现的灵活性引起研究者的广泛关注。然而，贝叶斯非参数方法并不是新的方法，早在 1973 年，Ferguson 就提出了以带有无限维度参数空间的参数模型来表示先验的贝叶斯非参数方法。但由于推理方法不成熟、计算机运算速度慢等原因，贝叶斯非参数方法一直停留于理论研究。近年来，高速计算机的快速发展解决了边缘概率积分的复杂计算问题，同时，MCMC 方法、EM 算法，以及关于边缘概率计算的近似算法如变元推理等计算方法的发展，大大扩展了贝叶斯非参数方法的应用领域。2001 年，Tipping 提出了“稀疏贝叶斯学习(Sparse Bayesian Learning)”的概念，利用层次先验、核函数和贝叶斯推断，给出了稀疏的相关向量机的学习方法，从而为统计稀疏学习方法提供了新的研究思路。此后，基于贝叶斯非参数的统计稀疏学习方法层出不穷，并在文本内容检索、基因数据分析、计算机视觉等领域获得应用。结合贝叶斯非参数方法的不确定性知识表达形式，综合先验知识的增量学习特性和非参数的模型灵活性，研究基于贝叶斯非参数的统计稀疏学习方法的独特性能和技术优势，并在应用中对其模型、方法和算法性能进行全面评估至关重要。

1.1.4 统计稀疏学习方法的视觉应用

在各种应用研究中，视觉任务面临的往往是有噪声、高维、大批量及多样性的数据样本，而且需要对数据内容进行高层次、结构性的语义分析和自动注解，这对当前的统计学习方法提出了很大的挑战。基于稀疏表达的视觉应用近年来已经取得了一些研究成果，比如：人脸识别、图像超分辨率、图像降噪、背景建模、运动分割等应用。基于贝叶斯非参数的统计稀疏学习方法的视觉应用还处于起步阶段，但应用的效果却让人印象深刻，例如，图 1.2 是基于贝叶斯方法学习的分类字典，图 1.3 是图像分割效果，图 1.4 是基于贝叶斯非参数方法得到的图像插值效果。研究统计稀疏学习中的贝叶斯非参数方法在视觉任务中的应用对于全面评估相应方法、模型和算法性能至关重要，也有助于深入理解贝叶斯非参数统计稀疏学习方法的理论价值，并为结合稀疏表达的贝叶斯非参数统计学习方法的有效性提供了很好的验证平台和应用示例。

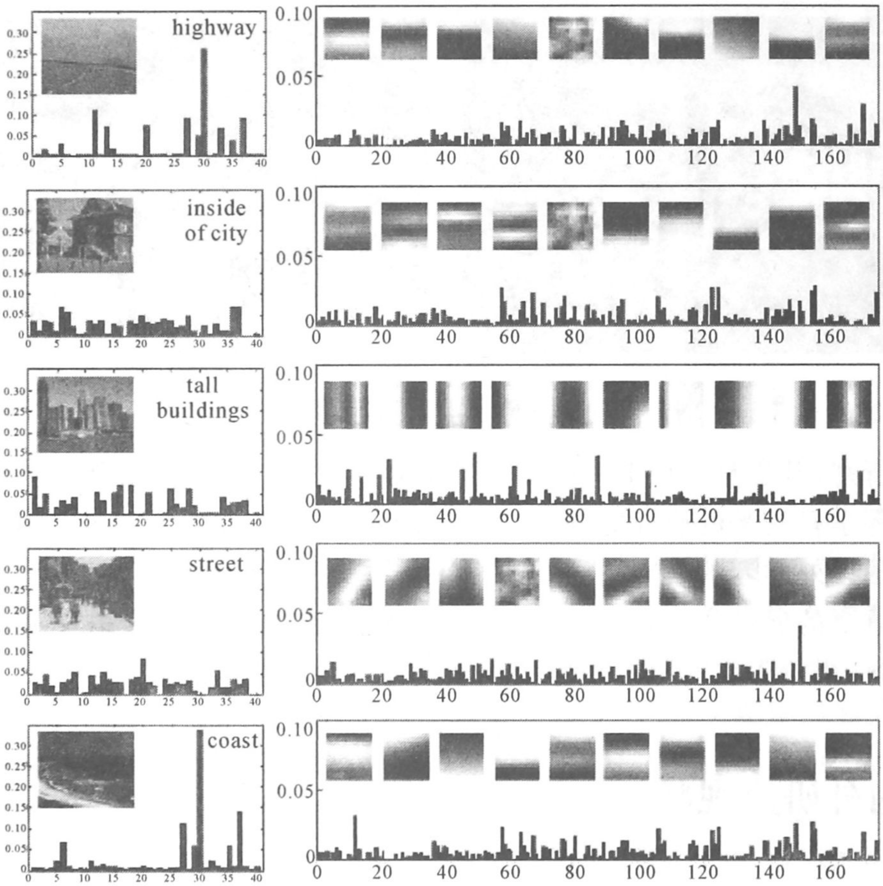


图 1.2 图像分类字典生成效果

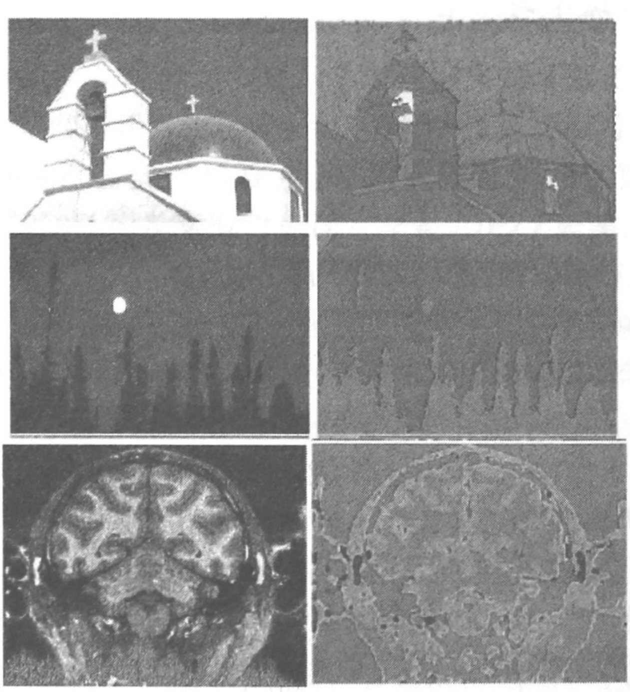


图 1.3 图像分割效果

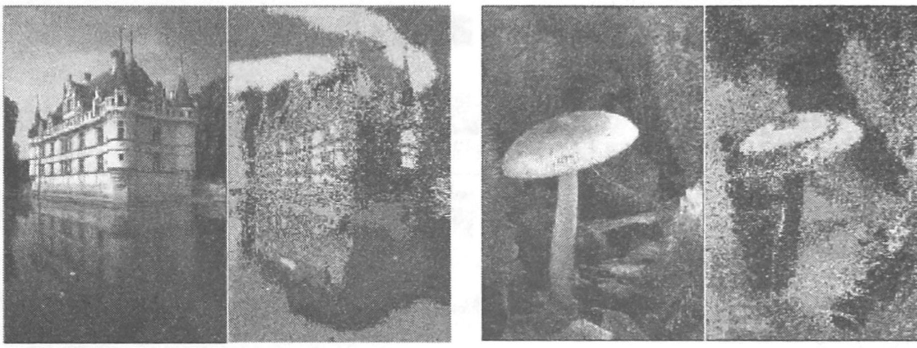


图 1.4 图像插值效果

综上所述,基于贝叶斯非参数的统计稀疏学习方法是人工智能、应用统计学及视觉认知等学科交叉的研究方向,也是当前统计学习领域的最新研究热点之一,已经引起了国内外学者的重点关注和研究兴趣。其研究成果不仅对统计学习的理论研究具有重要的促进作用,而且在大规模数据挖掘、多媒体内容语义分析、视觉行为自动注解、机器人交互等各种应用领域具有巨大的技术潜力。

1.2 国内外研究现状

1.2.1 统计稀疏学习方法

统计稀疏学习方法的研究起源于多个研究领域的成果:①来源于神经生物学对人类视觉皮层的认知机理研究,这为机器学习提供了生物学上的认知模型借鉴。②来源于数学等领域的最新研究进展,主要包括美国 Stanford 大学统计系的 Donoho 和 Candès、美国 UCLA 数学系的陶哲轩及美国 RICE 大学的 Baraniuk 等人在压缩感知理论和稀疏信号编码方面的开创性工作,为稀疏表达提供了基本的理论依据。③来源于凸优化理论方面的研究进展,主要包括 Stanford 大学的 Michael Saunders 等开发的凸优化算法,为稀疏表达的约束优化求解提供了理论支持和可计算方法。④来源于机器学习领域的理论研究进展,主要包括美国 UC Berkeley 大学统计系的 M. I. Jordan 研究组、Stanford 大学统计系的 Trevor Hastie 研究组等,近年来为稀疏表达与统计学习方法的结合提出了很多有效的学习方法与计算模型。⑤计算机视觉领域的研究者,包括法国 INRIA 的 Jean Ponce、美国 Stanford 大学的 F. F Li、UIUC 的 MaYi 等为稀疏表达在计算机视觉任务上的应用做了很多代表性的工作。

统计稀疏学习研究主要包括三个方面的内容:①稀疏建模,即研究如何构造稀

疏形式，建立有效的稀疏模型。②稀疏求解，即针对特定的稀疏模型，研究如何求解。③稀疏应用，稀疏形式的构造主要通过在最小平均误差项的基础上增加 l_0 范数约束来建模；而稀疏模型的求解则通过约束凸优化方法实现；稀疏应用方面的研究主要针对特定的稀疏问题，研究相应的稀疏学习策略。当前主要的稀疏学习方法基本都涉及上述三个方面的研究，代表性问题包括稀疏降维、字典学习、矩阵填充、稀疏高斯图模型、在线稀疏学习、结构稀疏性等。这些稀疏学习问题虽然已经得到了研究者的关注，但相应的理论和方法还有待研究和完善。

1）稀疏建模及求解

稀疏建模的基本思想是正则化方法，即通过构造同时包括“损失项”和“惩罚项”的约束优化函数来实现。损失项通常采用最小均方误差函数，而惩罚项通常采用范数约束。目前常用的稀疏构造形式包括：l_0 范数约束、Lasso 方法和弹性网（Elastic Net）方法等。从理论上说，l_0 范数约束作为惩罚项具有最优的稀疏形式，但 l_0 范数约束对应 NP 难题，通常无法直接求解。Lasso 方法采用 l_1 范数约束代替 l_0 范数约束构造稀疏模型，由于 Lasso 方法用回归模型系数的绝对值函数作为惩罚来压缩模型系数，使得绝对值较小的系数自动为零，从而实现模型参数选择的自然稀疏性。弹性网方法同时采用 l_1 范数约束和 l_2 范数约束，实现对 Lasso 方法的凸松弛，从而得到较“温和”的稀疏模型；当弹性网方法中 l_2 范数惩罚项的系数为零时，其退化为 Lasso 方法。在一些特定稀疏建模中，不仅对模型系数有稀疏性要求，同时还要求为非负。比如图像像素值的生成，可以在 Lasso 方法或者弹性网方法的基础上，增加对模型系数的非负约束，非负约束的模型来源于 Nonnegative Garrote 方法。在上述几种模型的基础上，Yuan 和 Lin 等提出了 Group-Lasso 方法用于处理结构性稀疏建模，考虑结构性稀疏建模的其他工作还包括 Bach，Huang 等提出的方法。

上述几种稀疏模型构成了典型的凸优化问题，因此能够采用相应的凸优化算法求解。需要注意的是，对于 Lasso 方法，由于 l_1 范数构造的约束是不可微的，这为凸优化问题的求解带来了困难。针对此问题，很多研究者提出了有效的求解策略，典型方法包括内点法（Interior Point Algorithms）、最小角度回归算法（Least Angle Regression，LARS）、正交匹配追踪（Orthogonal Matching Pursuit，OMP）、坐标梯度下降（Coordinate Gradient Descent，CGD）、块坐标梯度下降（Block-Coordinate Gradient Descent，BCGD）等。

2）稀疏降维方法

稀疏降维作为一种有效稀疏编码策略，能够对过完备字典（Overcomplete Dictionary）进行稀疏学习并构造精简的压缩表达。用于视觉任务的稀疏降维，其基函数的选择并不需要采用传统的正交基（傅里叶基、小波基等），而是可直接选自

图像、视频样本中的原始信息或特征表达，这在视觉识别、图像分类等应用中能够有效构造基于内容或语义的信息表达。目前提出的稀疏降维方法是通过对传统降维方法——主元分析法(PCA)增加稀疏性约束实现的，主元分析和主坐标分析(PCO)是统计学习方法中两个重要的无监督降维技术，它们互为对偶问题。因为主元分析降维后的主元包含了所有原始变量的线性组合，难以推断主元与原始变量之间的关系，如果主元只与很少的原始变量相关，则在实际应用中可以为主元与原始变量间建立更易解释的关系。稀疏主元分析(sPCA)通过增加对负荷(loadings)的稀疏约束，比如采用Lasso方法或者弹性网方法，构造负荷的自然稀疏性。目前的稀疏降维方法包括两类，一类基于最大主元的协方差性质，比如Jolliffe等提出的SCoTLASS，Shen等提出的sPCA-rSVD等；另一类则是Zou等提出的基于回归问题的sPCA方法。Zass等在考虑稀疏性的同时，结合了非负约束，提出了非负主元分析nsPCA，Jenatton在考虑变量结构性的条件下提出了结构稀疏主元分析(ssPCA)。

3)稀疏矩阵补充

采用稀疏约束的矩阵补充问题是近年来的研究热点问题，此问题虽然来源于推荐系统研究，但在机器学习和视觉应用中仍然具有广泛的应用价值。矩阵补充通常假设矩阵低秩或近似低秩，并在只有少量观察的情况下，要求恢复矩阵的原始信息，当前有代表性的工作包括Candès等提出的近似最优矩阵补充方法和精确矩阵补充方法、Cai等提出的SVT(Singular value thresholding)算法，以及Raghunandan等提出的从很少观察项中进行矩阵补充的方法。低秩矩阵补充问题主要通过建立最小化目标矩阵的秩并求解相应的约束优化问题。Mazumder提出采用谱正则化算法(Spectral regularization algorithms)求解矩阵补充问题。Cai和Candès等提出采用核范数(Nuclear norm)惩罚项作为秩约束的凸松弛条件求解。但上述两种方法在收敛速度和恢复精度上都无法同时保持高效。目前，矩阵补充问题的研究还在发展，尤其是对于大矩阵求解时的计算效率，以及当观察矩阵附带噪声时的求解等问题都尚待研究。

统计稀疏学习方法近年来在稀疏约束的矩阵分解、在线字典学习、结构稀疏性等方面有一些新的研究思路，尤其是在统计稀疏学习方法的应用方面，很多研究者做了许多探索性的尝试，并取得初步的研究成果，代表性的应用包括人脸识别、图像超分辨率、图像降噪、背景建模、运动和数据分割及图像分类等。2009年Francis Bach和计算机视觉领域的著名学者Jean Ponce等合作，在ICCV国际会议上对稀疏编码和字典学习用于图像分析做了系统介绍，MaYi等从计算机视觉和模式识别的角度也对当前的稀疏表达及其应用做了简介。

目前国内对统计稀疏学习方法的研究也日渐广泛，国内研究者在视觉认知的

编码机制、压缩感知、稀疏表示、字典学习、矩阵分解等的相关理论与应用方面取得了很多重要的研究成果。复旦大学的俞洪波教授等从神经生理学角度对视觉通路的信息编码机制进行研究,南京大学的周志华教授、浙江大学的张志华教授等从统计学习的角度对稀疏学习方法进行了研究。在应用研究方面,杨谦建立了一个基于超定完备基的简单细胞集群稀疏表示的计算模型,实现了自然图像的稀疏编码。中国科学院计算技术研究所李清勇博士设计了面向知觉任务的稀疏编码模型,并扩展了单层的基于 ICA 算法的稀疏编码模型。更多的研究集中于以脊波、曲线波为主线的理论分析及应用、匹配追踪算法在图像处理方面的具体应用、稀疏编码算法在图像处理和图像识别中的应用等。

1.2.2 贝叶斯非参数方法

贝叶斯非参数模型为非参数模型选择和自适应提供了一个贝叶斯框架。然而贝叶斯和非参数方法的结合充满了挑战,因为贝叶斯模型需要明确假设一个在给定参数空间上的概率分布,而非参数模型则根据样本数据改变参数空间的维度。1973 年, Ferguson 在可数样本空间上近似贝叶斯估计的基础上,提出了狄利克雷过程。一方面,既然非参数模型需要一个不受参数数量限制的先验,那么它可以被看作是带有无限维度参数空间的参数模型;另一方面,贝叶斯模型可以通过参数分布来定义,这个参数分布有无限维度的参数空间。这种模型通常称为“贝叶斯非参数模型”。

目前,最常见的贝叶斯非参数模型有高斯过程(Gaussian process)模型和狄利克雷过程(Dirichlet process)模型。高斯过程是传统的多变量高斯分布由向量到函数的自然扩展,其精练的协方差函数结构能极大地降低函数数据分析中的参数估计任务。早在 20 世纪七八十年代,高斯过程就已经以 Kriging 的名义应用于地理统计学领域中,但直到 90 年代中期,经过 Neal, Gibbs 和 MacKay 等人对高斯过程的阐述和发展,高斯过程才受到人们的重视,开始研究应用于机器学习领域,并在各应用领域迅速成为研究的热点。

狄利克雷过程的理论研究在 20 世纪 70 年代是众多研究者关注的热点,研究者们对狄利克雷过程的构造方法、狄利克雷过程的性质、后验计算方法展开了大量的理论研究。但由于其需要大规模的迭代计算,狄利克雷过程的应用一直没有突破性进展。2003 年以来,高速计算机的快速发展解决了边缘概率积分的复杂计算问题,同时,得益于 MCMC 方法、EM 算法、变分方法等的研究,狄利克雷过程的应用研究迅速发展,成为当前机器学习领域的热点。

对贝叶斯非参数方法的研究主要集中在两个方面:模型研究和应用研究。贝叶斯非参数方法以无限维度空间中的随机过程为研究对象,其理论研究包括建模

的方法、模型的性质、模型的推导和演绎等;而应用方面的研究主要针对特定的问题,研究相适应的贝叶斯非参数学习策略。

1)模型研究

贝叶斯非参数方法需要对无限维度空间中的测度进行建模,但在有限维空间中与 Lebesgue 测度相关的密度,如高斯概率密度、狄利克雷分布等,不能直接扩展到无限维空间,需要寻找适合的方法构建贝叶斯非参数模型。Ferguson 通过观察发现,既然非参数模型需要一个不受参数数量限制的先验,那么可以把它看作带有无限维度参数空间的参数模型,贝叶斯模型通过这种参数分布来定义,就得到贝叶斯非参数模型。

对于贝叶斯非参数模型的构造,研究者提出多种方法,主要有基于随机过程的方法、基于 De Finetti 理论的方法、基于 Kolmogorov 扩展理论的方法等。这些方法并不是相互之间完全排他的,例如狄利克雷过程可以通过多种方法构造。

随机过程方法通常适用于生成实数线上或实数区间上的随机概率分布,通过随机过程的非负增量路径来采样,从而描绘累积分布函数。其中,最典型的例子是 Lévy 过程,它是一个递增的过程,其包含的随机变量在一段时间上的概率只与时间段的长度有关。Lévy 过程在贝叶斯非参数中有广泛的应用,其后验在标准化之后得到 Gamma 过程。另外,基于随机过程的贝叶斯非参数模型还有 Griffin 的随机微分方程定义的过程、Küchler 以时间为参数定义具有指数形式的似然函数的随机过程等。

De Finetti 理论阐述了在给定一系列参数的条件下,变量之间是条件独立的。而对于连续的采样空间,这些参数是无限维的。Hewitt 和 Savage 证明了对于一个给定的随机变量序列,它们的混合分布是唯一的。定义一个无限的可交换的随机变量序列,可以通过指定一个生成算法来保证可交换性。例如,Blackwell 和 MacQueen 通过生成模型的混合分布来构建狄利克雷过程,称为无限 Pólya Urn 机制。

Kolmogorov 扩张定理直接从有限维度空间的边缘分布构建无限维度空间中的测度,是 Ferguson 构建狄利克雷过程的理论基础。Ferguson 证明了狄利克雷过程先验能够满足非参数贝叶斯分析的两个基本要求:①在适当的拓扑下,先验分布的支撑要足够大;②给定样本后,后验分布要便于计算。狄利克雷过程的支撑是可测空间上的所有离散概率分布组成的集合,同时,它的后验分布是容易计算的,可表示成先验与经验分布的混合。

狄利克雷过程的性质是狄利克雷分布的性质向无限维空间扩展的结果。Ferguson 在提出狄利克雷过程的同时对狄利克雷过程的诸多性质进行了证明,这奠定了狄利克雷过程的重要理论基础。Orbanz 着重从边缘分布的角度对狄利克

雷过程的性质进行阐述，强调了狄利克雷过程的合并过程。2010 年，Subhashis 对狄利克雷过程的性质进行了比较全面的总结。在阐述和证明狄利克雷过程性质过程中，Dubins 和 Pitman 根据狄利克雷过程和多项式过程的共轭性，提出“中国餐馆过程(Chinese Restaurant Process)”，在不限定类别数量的前提下对数据的聚类特性进行描述。

2)贝叶斯非参数模型的应用研究

近年来，贝叶斯非参数模型被应用于多种问题，例如回归、分类、聚类、隐变量模型，序列模型，图像分割，信号分离和语法归纳等，其中 LDA(Laten dirichlet allocation)模型是最典型也是最成功的贝叶斯非参数模型的应用。其在自然语言和智能信息处理中充分发挥了贝叶斯非参数在无限维度空间建模的优点。模型将主题混合权重视为多维参数的潜在随机变量，推理上采用 Laplace 近似、变分近似、MCMC(Markov chain Monte Carlo)及“期望-扩散(expectation propagation)”等方法获取待估参数值，在自然语言的词性标注、主题分解、信息抽取等方面取得广泛应用。狄利克雷过程在生物信息处理中也获得了令人惊叹的效果，例如在 DNA 排序技术中对单倍体型分期(haplotype phasing)用层次狄利克雷过程建模，从而提高了长片段测序能力。对语音的识别是狄利克雷过程的另一应用领域，以“层次狄利克雷过程-隐马尔可夫模型”建模的“说话人检索(Speaker Diarization)”可以在复杂的环境中快速识别说话者。

Sudderth, Li 和 Paisley 等对狄利克雷过程在计算机视觉中的应用进行了一定的探索。Sudderth 采用 Piteman-Yor 过程实现了对图像的分割和标注，并利用高斯过程对 Piteman-Yor 过程的空间独立性进行描述。Li 利用 LDA 对自然图像进行注解，Paisley 等首次利用“贝努利-贝塔过程”描述字典学习并将其应用在图像降噪、图像插值等方面，尽管效果差强人意，但为图像处理提供了新的方法和解决思路。

国内对于贝叶斯非参数方法的理论研究主要集中在 20 世纪 90 年代中期，主要对狄利克雷过程的性质、Lévy 表示、右中立过程等进行了分析和阐述。然而对贝叶斯非参数模型的应用研究尚处于起步阶段，目前鲜有对贝叶斯非参数模型在应用研究中的综述文献，对贝叶斯非参数中狄利克雷过程、贝塔过程等构造方法的相关研究和应用也亟待发展。目前国内相关研究的论文，有卿湘运等结合狄利克雷过程混合模型和选择特征子集的非参数模型，设计了基于马尔可夫链蒙特卡罗的参数后验推断算法，并将其应用于人脸聚类问题。

第 2 章 贝叶斯非参数模型的构建

自从 Ferguson 在 1973 年提出以带有无限维度参数空间的参数模型来表示先验的方法后，涌现了大量的构建贝叶斯非参数模型的方法。正是基于这些不同的模型构建方法，贝叶斯非参数过程得以广泛地应用在聚类、回归、变量选择等问题中。本章首先介绍贝叶斯非参数的理论基础，在此基础上，分析比较贝叶斯非参数模型的几种构建方法，再针对稀疏表示，对具有稀疏特质的贝叶斯非参数过程的构建和推理方法进行演绎，为全文的研究提供基本的方法。

2.1 符号约定

本书对有关概率模型和随机变量的符号约定如下：

随机变量定义在一个普通、抽象的概率空间 $(\Lambda, \mathcal{A}, P)$ 中，其中 Λ 是一个非空集合，有时称为样本空间，$\mathcal{A}$ 是σ-代数，P 是概率或概率测度。随机变量可以从这个普通的概率空间映射到相应的采样空间。随机变量用 X 表示，采样空间和他们的 σ-代数用相应的随机变量作为索引标注。例如，随机变量 $X:(\Lambda, \mathcal{A}) \rightarrow (\Omega_x, A_x)$，其中 Ω_x 是随机变量X 的采样空间，X 在采样空间中的采样值用小写斜体字母 $\boldsymbol{x}$ 表示。

对于有特定用途的变量，约定用 X 表示观测变量，Θ 表示参数变量，Y 表示超参数。任意的σ- 代数用A，C 等表示，但$\mathcal{B}$表示Borel σ- 代数。随机变量 X 的概率测度 μ 为$\mu = X(P)$。对于同一上下文中的多个随机变量，测度用随机变量作为索引标注，例如 μ_X，μ_Θ。

条件概率记为 $\mu(X \mid \Theta)$，在运算过程中，根据上下文，X 可能被测度集合代替，Θ 可能被σ- 代数代替。概率空间 Λ 中的元素用ω 表示，则把条件概率看作函数的话，可表示为 $\mu(X \mid \Theta)(\omega)$。如果 $\mu(X \mid \Theta)$ 有条件密度，则记为 $p(x \mid \theta)$。字母 s