Statistics and Application 统计学与应用, 2023, 12(2), 283-292
Published Online April 2023 in Hans. https://www.hanspub.org/journal/sa
https://doi.org/10.12677/sa.2023.122030

Hans汉斯

大数据中贝叶斯非参数方法的理论与应用研究

许　蕊，卢志义

天津商业大学理学院，天津

收稿日期：2023年3月10日；录用日期：2023年4月1日；发布日期：2023年4月14日

摘　要

在人工智能高速发展的时代，对机器学习领域的探索占据重要的地位，而机器学习本质上源于对海量数据的分析与学习，这就离不开统计学中模型的建立与推断。贝叶斯方法作为统计学中主要且成熟的建模方法，在充分学习样本信息的前提下引入参数的先验信息，容纳了参数的不确定性，使模型推断更加合理。在贝叶斯框架下的非参数方法进一步扩大了这种不确定性，将参数的先验空间推广到分布空间，用随机过程来进行表示，此时的先验空间是无限维的。贝叶斯非参数建模方法以其巨大的灵活性和稳健性得到了广泛的关注，随着人工智能的迅速发展，研究人员在机器学习领域对贝叶斯非参数方法展开了深入的研究并取得了许多优异的成果。本篇论文探究了贝叶斯非参数的部分基础理论，并对其在大数据背景下的实际应用进行了研究与展望。

关键词

贝叶斯非参数，大数据，机器学习，Dirichlet过程，后验推断

Research on the Theory and Application of Bayesian Nonparametric Methods in Big Data

Rui Xu, Zhiyi Lu

School of Science, Tianjin University of Commerce, Tianjin

Received: Mar. 10th, 2023; accepted: Apr. 1st, 2023; published: Apr. 14th, 2023

Abstract

In the era of rapid development of artificial intelligence, the exploration of the field of machine learning occupies an important position, and machine learning essentially stems from the analysis

文章引用: 许蕊, 卢志义. 大数据中贝叶斯非参数方法的理论与应用研究[J]. 统计学与应用, 2023, 12(2): 283-292.
DOI: 10.12677/sa.2023.122030

and learning of big data, which cannot be separated from the establishment and inference of models in statistics. Bayesian methods, as the main and well-established modelling methods in statistics, introduce a priori information about the parameters with sufficient learning of sample information, accommodating the uncertainty of the parameters and making model inference more reasonable. Nonparametric methods in the Bayesian framework further extend this uncertainty by extending the prior space of parameters to the distribution space, which is represented by a stochastic process, at which point the prior space is infinitely dimensional. Bayesian nonparametric modelling methods have received widespread attention for their great flexibility and robustness, and with the rapid development of artificial intelligence, researchers have conducted in-depth research on Bayesian nonparametric methods in the field of machine learning and achieved many excellent results. This paper explores some of the underlying theory of Bayesian nonparametric and investigates and prospects for its practical application in the context of big data.

Keywords

Bayesian Nonparametric, Big Data, Machine Learning, Dirichlet Process, Posterior Inference

1. 引言

在大数据背景下，机器学习蓬勃发展，但绝大部分的机器学习模型都是“黑箱”模型，无法对其过程和结果进行显式的解释。Judea Pearl 在人工智能领域倡导概率方法并提出了贝叶斯网络[1]，这些开创性工作使得贝叶斯(机器)学习在机器学习领域发挥了重要的不可替代的作用。目前贝叶斯方法在监督学习、迁移学习、强化学习、因子分析等领域取得了众多优秀的成果，应用广泛的模型有：LDA (Latent Dirichlet Allocation) [2]、高斯混合模型[3]、隐马尔可夫模型[4]等。

贝叶斯学习有坚实的概率理论基础和清晰的概率结构，既容纳了不确定性又可以显式地表示，具有内在的可解释性。除此以外，贝叶斯学习将先验信息与观测数据中学习到的信息并行编码，不断更新迭代，实现自适应，且一定程度上可以克服在有限数据量下出现的过拟合问题。在贝叶斯学习中，参数模型被广泛应用并在许多情况下展现出了良好的效果，但参数模型的局限性也是显而易见的，其本身存在的许多限制使得它失去了一定的灵活性，例如参数个数有限、参数先验维度有限等，这些通常依赖于人的经验来设定，但当设置不恰当时就会导致模型出错，从而需要花费更多时间来进行调参，因此参数方法耗时且很难扩展应用到大规模或不熟悉的数据上。而贝叶斯框架下的非参数模型将参数建模包含的不确定性进一步扩大，极大地增加了模型的灵活性，减少了掺杂在模型中的主观性，并且使得模型可以应用于更加广泛的数据，增加了模型的稳健性，这就是统计学中相对来说比较“年轻”的贝叶斯非参数(Bayesian Nonparametric, BNP)。

BNP 的发展历史可以追溯到二十世纪七十年代，统计学家 Thomas S. Ferguson 在 1973 [5]、1974 [6]年撰写了两篇开创性论文，从此开始了 BNP 的统计研究之路。严格来说，“非参数”一词在这里是存在语义错误的，此处的“非参数”并不是没有参数，而是指无穷维参数。非参数建模的目标包括对未知概率测度的泛函进行推断，概率测度本身就是参数，因此这里的参数空间其实是函数空间，是无穷维的。而贝叶斯框架下，非参数则是将参数的先验空间推广到无限维分布空间，本质上可以看作是一个随机过程，其中应用十分广泛的有 Dirichlet 过程[7]、高斯过程[8]、泊松过程[9]、贝塔过程[10] [11]等。随机过

程使得先验不受参数假设的限制，从而可以更加准确地对真实数据建立模型。又因为在贝叶斯框架下，更新的观测值与先验不断迭代，从而实现让数据说话，自适应地进行模型选择，因此贝叶斯非参数方法建模更加灵活、稳健。但极大的灵活性带来的是难以解决的计算问题，其后验分布通常极其复杂，无法直接得到后验结果，这也导致 BNP 在很长一段时间内只能停留在理论研究层面，直到二十世纪九十年代计算机技术和相关算法的突破性发展，才使得 BNP 的后验推断有了大幅的发展。目前最常用到的后验近似推断方法分为两类，一种是基于采样的随机近似推断，如马尔可夫链蒙特卡罗(Markov Chain Monte Carlo, MCMC) [12] [13]，其中应用广泛的方法有 M-H 采样(Metropolis-Hastingsampling)和吉布斯采样(Gibbs Sampling) [14]，在面对十分复杂的后验推断时常常将两种采样方法结合交替进行。另一种为确定性近似推断，如变分推断(Variational Inference, VI) [15]。MCMC 和 VI 一定程度上都可以很好地解决 BNP 的后验推断问题，MCMC 准确性高但收敛速度慢，VI 计算速度快但准确性差一些，根据实际问题选择合适的后验推断方法就可以得到理想的结果。

BNP 在统计学领域发展迅速，但直到 2005 年，BNP 才进入计算机领域，Yee W. Teh 等人[16]将 BNP 应用到了机器学习中，并取得了巨大的成功，自此经过不断发展便产生了贝叶斯非参数(机器)学习(Bayesian Nonparametric Learning, BNL)，并成为了统计学与计算机科学的跨学科课题。随着多核 CPU、GPU 和分布式计算平台等硬件的计算能力的不断增强，通过高效的后验推断算法，可以将 BNL 应用到大数据背景下的众多现实问题中。

BNP 中包含的随机过程数量众多，无法一一叙述，本篇论文基于 Dirichlet 过程相关知识对 BNP 的基础理论展开描述，并对 BNP 在机器学习和实际问题的应用进行探究，最后对 BNP 的研究与发展进行总结与展望。

2. 基于 Dirichlet 过程的 BNP 的基础理论

2.1. BNP 的定义

在统计学中，贝叶斯非参数区别于频率参数、贝叶斯参数和频率非参数，是一个较新且发展迅速的领域。关于 BNP 的定义有很多种描述方法，Peter Müller 等人[17]在书中指出，在非参数模型中进行贝叶斯推断，需要建立具有无限维参数的先验的概率模型，这种先验被称为 BNP 先验，整个推断模型称为 BNP 模型。

BNP 将参数的先验空间推广到无限维，进一步扩大了参数的选择范围，极大地降低了模型构建中的人为主观性。具体表现为在聚类任务中，无需提前设置分组数量；在贝叶斯推断中，无需人为规定先验的具体参数等。此外，BNP 在贝叶斯框架下的自适应特性可以包容更多的不确定性，随着观测数据的增加，模型复杂度也相应增加，在面对复杂的实际问题时模型也更加稳健。近年来许多研究人员对 BNP 的理论方法进行了深入广泛的研究和综述，例如：Xuan, J.等人[18]、Gershman, S. J. and Blei, D. M. [19]、Hjort, N. L.等人[20]、Müller, P. and Mitra, R. [21]等。

在机器学习领域，Orbanz, P. and Teh, Y. W. [22]在机器学习百科全书中给出了一个更适合机器学习领域研究人员的更简单的定义：BNP 是无限维参数空间上的贝叶斯模型，参数空间通常被选为给定学习问题的所有可能解的集合。这个定义更关注 BNP 的无限维特性和学习能力，在无限维参数空间上建立模型意味着参数的先验由概率分布变为随机过程，因此 BNP 对机器学习的主要贡献是为高维空间划分和更复杂的数据结构引入随机过程。从计算机科学的角度来看，BNP 的优势在于其灵活的数据结构和应用。

Xuan, J.等人根据 BNP 的优势对 BNL 作出了定义：BNL 是基于随机过程及其应用，建立和推断特定学习任务的概率模型。他们指出为特定的学习任务构建贝叶斯非参数模型就像使用乐高积木搭建机器人，

其中随机过程是不同形状的积木，对随机过程的应用就像将积木组装成复杂的机器人。这个定义生动地展示了 BNL 的核心和建立模型的标准程序。

BNP 中的随机过程种类繁多，但其基础原理与构造应用都有相似之处，下面将以 Dirichlet 过程为例简单介绍其基础理论与相关扩展。

2.2. Dirichlet 过程及其构造方法

2.2.1. Dirichlet 过程定义

Dirichlet 过程(Dirichlet Process, DP)是 BNP 中最早提出也是应用最广泛最受欢迎的一个模型，它由 Ferguson 在 1973 年提出[5]，作为概率测度空间的先验模型，标志着 BNP 研究的开始。关于 DP 的定义不同的学者做出了许多不同的表示形式，在这里我们引用 Müller, P.等人[17]在《贝叶斯非参数数据分析》一书中的表示方法。

假设实数 $M>0$，G_0 是定义在 S 上的一个概率测度，G 是定义在 S 上的随机概率测度，且对每一个集合 B 都有概率 $G(B)$。若对 S 的每一个有限划分 $\{B_1,\cdots,B_k\}$，向量 $\left(G(B_1),\cdots,G(B_k)\right)$ 的联合分布服从参数为 $\left(MG_0(B_1),\cdots,MG_0(B_k)\right)$ 的 Dirichlet 分布，即：

$$\left(G(B_1),\cdots,G(B_k)\right)\sim \mathrm{Dir}\left(MG_0(B_1),\cdots,MG_0(B_k)\right),$$

则称 G 服从一个参数为 (M,G_0) 的 Dirichlet 过程，记作 $\mathrm{DP}(MG_0)$ 或 $\mathrm{DP}(M,G_0)$。其中 M 被称作精度参数或集中参数，G_0 是中心测度，称 $\alpha=MG_0$ 为 DP 的基测度。

对任意的 B 及其补集 B^C，有 $E\left[G(B)\right]=G_0(B)$，$\mathrm{Var}\left[G(B)\right]=G_0(B)\left[1-G_0(B)\right]/(1+M)$，且 $\left(G(B),G\left(B^C\right)\right)\sim \mathrm{Dir}\left\{MG_0(B),M\left[1-G_0(B)\right]\right\}$。以上结果也展示了精度参数 M 对 DP 的影响，当 M 越大时，G 越接近 G_0，当 $M\to\infty$ 时，$G=G_0$。

DP 是共轭的，也就是说 DP 的后验依然是 DP，且后验 DP 的精度参数为 $M+n$，中心测度是 G_0 和经验分布 $\hat{f}_n(\cdot)=\frac{1}{n}\sum_{i=1}^{n}\delta_{y_i}(\cdot)$ 的加权平均，其中 $\delta_x(\cdot)$ 是 x 处的 Dirac 测度。

若 $y_1,\cdots,y_n\,|\,G\overset{iid}{\sim}G, G\sim \mathrm{DP}(MG_0)$，则有，

$$G\,|\,y_1,\cdots,y_n\sim \mathrm{DP}\left(MG_0+\sum_{i=1}^{n}\delta_{y_i}\right).$$

关于 BNP 模型的构造，基于不同的原理有多种不同的构造方法，例如：基于随机过程、Kolmogorov 扩展理论、De Finetti 理论等。下面两节将介绍两种常用的 Dirichlet 过程的构造方法。

2.2.2. Stick-Breaking 构造方法

关于 Stick-Breaking 方法的理论研究有很长的历史，例如：Halmos, P. R. [23], Freedman, D. A. [24], Kingman, J. F. [25], Ishwaran, H. and James, L. F. [26]等。Sethuraman, J. [27]在 1994 年描述了 DP 的一种简洁的构造方式，即 stick-breaking 先验。

假设 $v_h\overset{iid}{\sim}\mathrm{Be}(1,M)$，$w_h=v_h\prod_{\ell<h}(1-v_\ell)$，$m_h\overset{iid}{\sim}G_0$，其中，$\mathrm{Be}(\cdot)$ 表示 Beta 分布，$\{v_h\}$ 和 $\{m_h\}$ 是独立的。那么随机概率测度 $\mathrm{DP}(MG_0)$ 的 stick-breaking 表示为，

$$G(\cdot)=\sum_{h=1}^{\infty}w_h\delta_{m_h}(\cdot).$$

形象化地描述 Stick-Breaking 构造 DP 过程为：假设有一根单位长度为 1 的棍子，首先随机产生一个 $v_1\overset{iid}{\sim}\mathrm{Be}(1,M)$，在棍子的 v_1 比例处截断，截下的部分为 w_1，剩余部分为 $1-v_1$。再随机产生一个

$v_2 \overset{iid}{\sim} \mathrm{Be}(1, M)$，在剩余部分的 v_2 比例处截断，即截下的部分为 $w_2 = v_2(1-v_1)$，剩余部分为 $(1-v_1)-v_2(1-v_1)=(1-v_2)(1-v_1)$，同样的规律依次进行下去就得到了 DP 的 stick-breaking 构造，这种构造方法也直观地显示出了 DP 的离散的性质。

2.2.3. 中国餐馆过程构造方法

中国餐馆过程(Chinese Restaurant Process, CRP)是 DP 的另一种常见的构造方法，它直观地展示了 DP 的聚类的特性。关于 CRP 详细的理论介绍可以参考 Ishwaran, H. and James, L. F. [28], Teh, Y.等人[16], Pitman, J. [29]的论文。下面将形象化地描述中国餐馆过程。

假设有一家中国餐馆，里面有无数张桌子，每张桌子可以容纳无数个人，每一个顾客按顺序进店，第一个顾客随机选择一张桌子坐下，后面的顾客可以选择已经有人的桌子入座，也可以选择一张新的桌子入座。假设前 n 个顾客已经选座完毕，记作 $X_1, X_2, \cdots, X_n$，所有被选择的不同的桌子记作 $X_1^*, X_2^*, \cdots, X_m^*$，其中每张桌子上已有的人数记作 $n_1, n_2, \cdots, n_m$，那么第 $n+1$ 个顾客将会以 $\frac{n_k}{n+M}$ 的概率坐在第 k 张已经有人的桌子上，以 $\frac{M}{n+M}$ 的概率坐在一张新桌子上，且 $X_{new}^* \sim G_0$。于是可以得到第 n + 1 个顾客选择桌子的预测分布：

$$X_{n+1} \mid X_1, X_2, \cdots, X_n, M, G_0 \sim \frac{1}{M+n}\left(MG_0 + \sum_{k=1}^{K} n_k \delta_{X_k^*}\right) \tag{1}$$

通过以上过程可以直观地看到，每一张桌子就是一类，这种数据的划分符合 DP 的定义。且根据(1)式可以看出当人数不断增多，n_k 不断增大，新的顾客选择已有人的桌子的概率会越高，也就是“多的会更多”，因此能够实现快速聚类。

3. BNP 的后验推断

掌握了 BNP 的基础理论，就可以根据具体的任务和数据构造恰当的 BNP 模型，下一步则需要进行后验推断，重点在于获得模型的后验分布，但 BNP 的后验分布通常极其复杂无法直接计算，因此统计学中解决这个问题主要有两种方法：基于抽样的方法，如马尔科夫链蒙特卡罗方法(MCMC)和基于优化的方法，如变分推断(VI)。而在大数据时代，多数领域的数据量呈指数级增长，一个自然地想法是将以上推断算法扩展为多处理器的并行版本。目前并行 MCMC [30]、并行变分推断[31] [32]等算法都在研究发展过程中。

3.1. MCMC

MCMC 是贝叶斯模型后验推断的最直接的方法，也被广泛地应用于贝叶斯非参数模型，其中最常见的方法是 M-H 采样和 Gibbs 采样，以及复杂情况下两种采样方法的结合。MCMC 的基本思想是根据采样样本的分布情况来近似后验分布，这个过程需要构造马尔科夫链，它的平稳分布即期望中的后验分布。理论上只要采样的样本个数足够多就可以得到后验分布的精确推断，但马尔科夫链达到平稳分布的时间不可控，有时需要耗费极其长的时间，且要进行复杂的平稳检验，因此尤其在面对大数据的情况下，MCMC 若想得到足够多的样本耗时太长，存在效率问题。

面对 BNP 无限维度的性质，其后验推断面临的问题更加严峻，通常采用的解决方法是截断法[33] [34]，可以很好地解决无限维推断问题，但这种方法必然会引入一定的误差。另一种常用的方法是切片采样[35] [36]，Kalli, M.等人[37]描述了 Dirichlet 过程在机器学习中的切片采样过程，Broderick, T.等人[38]在机器学习中应用了 Beta 过程的切片采样方法。

3.2. 变分推断

变分推断是解决贝叶斯后验推断的另一种思路，它使用更简单的参数化的变分分布来近似真实的后验分布，将后验推断问题转化为高维优化问题。借助梯度，该方法可以有效地搜索变分分布的参数空间，以尽可能逼近真实的后验分布。变分推断需要预先设置变分分布并选择优化方法。在大数据等复杂场景下，设置恰当的变分分布可以显著降低附加的逼近误差，选择适当的优化方法可以提高推断效率。

目前在基于大数据的 BNL 中常用的变分推断算法及其变体有：普通变分推断(Ordinary Variational Inference) [39]，通常基于 stick-breaking 表示，用截断法来处理无限维的情况；Collapsed 变分推断(Collapsed Variational Inference) [40]，基于 stick-breaking 表示，将权重边际化以提高推断的准确性和有效性；随机变分推断(Stochastic Variational Inference) [41]，在每一次更新变分参数前，从总体的大数据集中采样一部分数据出来，以提高推断效率；不截断的变分推断(Truncation-Free Variational Inference)，截断法在处理无限维推断问题中有效但无可避免的会引入误差，因此研究无需截断的推断方法是很多研究人员的研究方向，例如变分 DP [42]等；流变分推断(Streaming Variational Inference) [43]，用来处理以数据流形式顺序到达的数据。

在大数据环境下，变分推断的运行效率要比 MCMC 高，但其中引入的变分分布会造成额外的不必要的逼近误差，因此变分推断在近似精度上会低于 MCMC。

4. BNP 在机器学习中的应用

BNP 在机器学习中的应用被称为 BNL (Bayesian Nonparametric Learning)，它可以有效解决许多机器学习中的任务，并且目前已经有了很多与经典模型相对应的 BNP 的扩展模型。下面简单介绍 BNL 在机器学习中的应用。

- 监督学习：监督学习是机器学习中的一项基本任务，它对数据和协变量之间的关系进行建模以做出预测。以分类为例，响应变量和协变量之间的关系用广义线性模型建模，将 DP [44]作为类别权重的先验。
- 强化学习：强化学习在游戏开发与机器人研究等方面有不可或缺的作用，其行动基于环境的反馈，与环境不断交互、试错的过程使其达到特定的目的。然而有时环境的状态无法被观测到或者只能观测到一部分，提供的信息是不完整的，而 BNP 可以很好地解决此类问题，例如 HDP-HMM (Hierarchical Dirichlet Process-Hidden Markov Model) [45]可以将隐藏状态转换为来自 HDP 的 stick 权重。
- 迁移学习：迁移学习是把已知领域学习到的知识和构建的模型应用到一个新的领域的机器学习方法，用来解决某些领域数据量不足或者减少学习量的问题。在共享因子迁移中，贝叶斯非参数联合因子分析[46]通过分层 Beta 过程先验实现因子共享。共享主题迁移中，分层 DP [47]作为可转移的主题的先验使得它们更加灵活。共享树迁移中，thLDA (The Transfer Hierarchical LDA)将现有树中的知识转移到目标域中，通过 nCRP [48]对目标域中的每个文档进行从根到底节点的树的路径采样。
- 因子分析：因子分析是机器学习中广泛应用的降维方法，但通常对于因子的维度是需要预先设定的，例如主成分分析中，选择的主要成分的个数是要人为给定的，但是当先验知识不足或没有先验知识时就很难做出恰当的指定，BNL 就可以很好地解决这个问题，其中 IBP (Indian Buffe Process)十分受欢迎，例如与主成分分析(PCA)的结合——BNP-PCA [49]有很好的应用效果。
- 因果推断：因果推断研究的是科学地识别变量间的因果关系，其核心是在某一结果发生的条件下对另一个事件的影响，BNP 中的 GP (Gaussian Process) [50]常用来逼近其条件分布，可以取得很好的效果。

5. BNP 在实际问题中的应用

自 BNP 与计算机领域相结合以来发展十分迅速，尤其是在过去的十几年间，是众多研究人员的研究焦点，因此在众多热门的实际问题中，BNP 都得到了很大的应用空间并取得了令人欣慰的成绩。下面将简单介绍个别 BNP 模型在某一实际领域的应用情况。

- 机器人：在机器人的研究中，BNP 可以应用在训练机器人完成特定任务中，例如在机器人导航中，需要机器人根据传感器数据找到通往目标的路线，Jiang, Y.和 Saxena, A. [51], Plagemann, C.等人[52]用 GP 来建模实现回归任务。
- 生物学：Dirichlet 过程及其扩展可以应用在生物信息处理中，例如 Xing, E.P. 等人[53] [54]用分层 Dirichlet 过程对 DNA 排序技术中的单倍体型分期建模，使得长片段测序能力有效提高；在分子生物学中的基因分析中起重要作用的 EST 分析(Expressed Sequence Taganalysis)中，BNP 中的 PYP [55]模型(Pitman-Yor proces)可以有效估计其文库中的基因比例和新基因的数量。
- 计算机视觉：BNP 在计算机视觉领域的应用十分广泛，在图像分割、标注、降噪、插值、运动捕捉等方面都有十分成功的应用。例如 Haines, T. S.和 Xiang, T. [56]提出用 Dirichlet 过程混合模型来进行背景去除；Sudderth, E. B.等人[57]在图像分割与标注中引入了 PYP 模型等。
- 自然语言处理：自然语言处理中的语音识别在近二十年来发展极其迅速，已经在生活中随处可见，其中 BNP 模型也起到了不可忽视的作用，例如 HDP-HMM 模型[58]在 Speaker Diarization 中的应用可以迅速识别出讲话者，即使在十分复杂的环境中，识别准确率也很高。在分词的应用中，CRP [59]的基本原理可以得到充分发挥。

除以上提到的几个领域以外，BNP 在文本挖掘、音乐分析、信号处理等方面也有广泛的应用，并在向其他各个领域持续发展中。

6. 总结与展望

BNP 作为统计学中相对来说比较新的领域发展十分迅速，尤其是大数据时代的到来和人工智能的不断发展，BNP 因其极大的灵活性和稳健性在机器学习领域得到了广泛的研究与应用。本篇论文以 Dirichlet 过程为例，描述了 BNP 的基础理论知识，列举了当下关于 BNP 的后验推断的重要的研究成果，并对 BNP 在机器学习领域和实际问题中的广泛应用做了简单介绍。

虽然 BNP 的发展在这个阶段已经取得了不错的成就，但它依然还处在发展阶段并存在一些亟待解决的问题，例如理论与实际应用之间的不匹配问题，目前关于 BNP 的后验推断大多还是采用截断的方式，根本上还是没有达到无限维的要求，因此未来关于无需截断的后验推断方法是一个研究的重点方向。除此以外，BNP 与深度学习的结合、BNP 在高维数据中的应用、根据复杂的实际情况创造更多新的 BNP 模型等都是未来需要深入探索的方向。由此看来，BNP 在未来依然存在着无限的发展前景。

参考文献

[1] Pearl, J. (1986) Fusion, Propagation, and Structuring in Belief Networks. *Artificial Intelligence*, **29**, 241-288. https://doi.org/10.1016/0004-3702(86)90072-X

[2] Blei, D.M., Ng, A.Y. and Jordan, M.I. (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research*, **3**, 993-1022.

[3] Reynolds, D.A. (2009) Gaussian Mixture Models. In: Li, S.Z. and Jain, A., Eds., *Encyclopedia of Biometrics*, Springer, Berlin, 659-663. https://doi.org/10.1007/978-0-387-73003-5_196

[4] Eddy, S.R. (1996) Hidden Markov Models. *Current Opinion in Structural Biology*, **6**, 361-365. https://doi.org/10.1016/S0959-440X(96)80056-X

[5] Ferguson, T.S. (1973) A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics*, **1**, 209-230. https://doi.org/10.1214/aos/1176342360

[6] Ferguson, T.S. (1974) Prior Distributions on Spaces of Probability Measures. *The Annals of Statistics*, **2**, 615-629. https://doi.org/10.1214/aos/1176342752

[7] Teh, Y.W. (2010) Dirichlet Process. In: Sammut, C. and Webb, G.I., Eds., *Encyclopedia of Machine Learning*, Springer, Berlin, 280-287. https://doi.org/10.1007/978-0-387-30164-8_219

[8] Seeger, M. (2004) Gaussian Processes for Machine Learning. *International Journal of Neural Systems*, **14**, 69-106. https://doi.org/10.1142/S0129065704001899

[9] Kingman, J.F.C. (1992) Poisson Processes. Vol. 3, Clarendon Press, Oxford.

[10] Hjort, N.L. (1990) Nonparametric Bayes Estimators Based on Beta Processes in Models for Life History Data. *The Annals of Statistics*, **18**, 1259-1294. https://doi.org/10.1214/aos/1176347749

[11] Thibaux, R. and Jordan, M.I. (2007) Hierarchical Beta Processes and the Indian Buffet Process. *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, Vol. 2, 564-571.

[12] Geyer, C.J. (1992) Practical Markov Chain Monte Carlo. *Statistical Science*, **7**, 473-483. https://doi.org/10.1214/ss/1177011137

[13] Andrieu, C., De Freitas, N., Doucet, A. and Jordan, M.I. (2003) An Introduction to MCMC for Machine Learning. *Machine Learning*, **50**, 5-43. https://doi.org/10.1023/A:1020281327116

[14] Casella, G. and George, E.I. (1992) Explaining the Gibbs Sampler. *The American Statistician*, **46**, 167-174. https://doi.org/10.1080/00031305.1992.10475878

[15] Blei, D.M., Kucukelbir, A. and McAuliffe, J.D. (2017) Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association*, **112**, 859-877. https://doi.org/10.1080/01621459.2017.1285773

[16] Teh, Y., Jordan, M., Beal, M. and Blei, D. (2004) Sharing Clusters among Related Groups: Hierarchical Dirichlet Processes. *Proceedings of the* 17*th International Conference on Neural Information Processing Systems*, Vancouver, 1 December 2004, 1385-1392.

[17] Müller, P., Quintana, F.A., Jara, A. and Hanson, T. (2015) Bayesian Nonparametric Data Analysis. Vol. 1, Springer, New York. https://doi.org/10.1007/978-3-319-18968-0_1

[18] Xuan, J., Lu, J. and Zhang, G. (2019) A Survey on Bayesian Nonparametric Learning. *ACM Computing Surveys* (*CSUR*), **52**, 1-36. https://doi.org/10.1145/3291044

[19] Gershman, S.J. and Blei, D.M. (2012) A Tutorial on Bayesian Nonparametric Models. *Journal of Mathematical Psychology*, **56**, 1-12. https://doi.org/10.1016/j.jmp.2011.08.004

[20] Hjort, N.L., Holmes, C., Müller, P. and Walker, S.G. (2010) Bayesian Nonparametrics. Vol. 28, Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511802478

[21] Müller, P. and Mitra, R. (2013) Bayesian Nonparametric Inference—Why and How. *Bayesian Analysis*, **8**, 342 p. https://doi.org/10.1214/13-ba811

[22] Orbanz, P. and Teh, Y.W. (2010) Bayesian Nonparametric Models. In: Sammut, C. and Webb, G.I., Eds., *Encyclopedia of Machine Learning*, Springer US, Boston, 81-89. https://doi.org/10.1007/978-0-387-30164-8_66

[23] Halmos, P.R. (1944) Random Alms. *The Annals of Mathematical Statistics*, **15**, 182-189. https://doi.org/10.1214/aoms/1177731283

[24] Freedman, D.A. (1963) On the Asymptotic Behavior of Bayes' Estimates in the Discrete Case. *The Annals of Mathematical Statistics*, **34**, 1386-1403. https://doi.org/10.1214/aoms/1177703871

[25] Kingman, J.F. (1975) Random Discrete Distributions. *Journal of the Royal Statistical Society*: *Series B* (*Methodological*), **37**, 1-15. https://doi.org/10.1111/j.2517-6161.1975.tb01024.x

[26] Ishwaran, H. and James, L.F. (2001) Gibbs Sampling Methods for Stick-Breaking Priors. *Journal of the American Statistical Association*, **96**, 161-173. https://doi.org/10.1198/016214501750332758

[27] Sethuraman, J. (1994) A Constructive Definition of Dirichlet Priors. *Statistica Sinica*, **4**, 639-650.

[28] Ishwaran, H. and James, L.F. (2003) Generalized Weighted Chinese Restaurant Processes for Species Sampling Mixture Models. *Statistica Sinica*, **13**, 1211-1235.

[29] Pitman, J. (2006) Combinatorial Stochastic Processes: Ecole d'Eté de Probabilités de Saint-Flour XXXII-2002. Springer, Berlin.

[30] Smyth, P., Welling, M. and Asuncion, A. (2008) Asynchronous Distributed Learning of Topic Models. *NIPS*'08: *Proceedings of the* 21*st International Conference on Neural Information Processing Systems*, Vancouver, 8-11 December 2008, 81-88.

[31] Campbell, T., Straub, J., Fisher III, J.W. and How, J.P. (2015) Streaming, Distributed Variational Inference for Bayesian Nonparametrics. *Proceedings of the* 28*th International Conference on Neural Information Processing Systems*, Volume 1, 280-288.

[32] Neiswanger, W., Wang, C. and Xing, E. (2015) Embarrassingly Parallel Variational Inference in Nonconjugate Models.

[33] Fox, E.B. (2009) Bayesian Nonparametric Learning of Complex Dynamical Phenomena. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge.

[34] Fox, E., Sudderth, E., Jordan, M. and Willsky, A. (2008) Nonparametric Bayesian Learning of Switching Linear Dynamical Systems. *Proceedings of the* 21*st International Conference on Neural Information Processing Systems*, 8 December 2008, 457-464.

[35] Damlen, P., Wakefield, J. and Walker, S. (1999) Gibbs Sampling for Bayesian Non-Conjugate and Hierarchical Models by Using Auxiliary Variables. *Journal of the Royal Statistical Society*: *Series B* (*Statistical Methodology*), **61**, 331-344. https://doi.org/10.1111/1467-9868.00179

[36] Neal, R.M. (2003) Slice Sampling. *The Annals of Statistics*, **31**, 705-767. https://doi.org/10.1214/aos/1056562461

[37] Kalli, M., Griffin, J.E. and Walker, S.G. (2011) Slice Sampling Mixture Models. *Statistics and Computing*, **21**, 93-105. https://doi.org/10.1007/s11222-009-9150-y

[38] Broderick, T., Mackey, L., Paisley, J. and Jordan, M.I. (2014) Combinatorial Clustering and the Beta Negative Binomial Process. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**, 290-306. https://doi.org/10.1109/TPAMI.2014.2318721

[39] Blei, D.M. and Jordan, M.I. (2006) Variational Inference for Dirichlet Process Mixtures. *Bayesian Analysis*, **1**, 121-143. https://doi.org/10.1214/06-BA104

[40] Kurihara, K., Welling, M. and Teh, Y.W. (2007) Collapsed Variational Dirichlet Process Mixture Models. *Proceedings of the International Joint Conference on Artificial Intelligence*, Vol. 7, 2796-2801.

[41] Bryant, M. and Sudderth, E. (2012) Truly Nonparametric Online Variational Inference for Hierarchical Dirichlet Processes. *Proceedings of the* 25*th International Conference on Neural Information Processing Systems*, Volume 2, 2699-2707.

[42] Kurihara, K., Welling, M. and Vlassis, N. (2006) Accelerated Variational Dirichlet Process Mixtures. In: Schölkopf, B., Platt, J. and Hoffman, T., Eds., *Advances in Neural Information Processing Systems*, The MIT Press, Cambridge, 761-768.

[43] Lin, D. (2013) Online Learning of Nonparametric Mixture Models via Sequential Variational Approximation. *Proceedings of the* 26*th International Conference on Neural Information Processing Systems*, Volume 1, 395-403.

[44] Hannah, L.A., Blei, D.M. and Powell, W.B. (2011) Dirichlet Process Mixtures of Generalized Linear Models. *Journal of Machine Learning Research*, **12**, 1923-1953.

[45] Doshi-Velez, F., Pfau, D., Wood, F. and Roy, N. (2013) Bayesian Nonparametric Methods for Partially-Observable Reinforcement Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**, 394-407. https://doi.org/10.1109/TPAMI.2013.191

[46] Gupta, S.K., Phung, D. and Venkatesh, S. (2012) A Bayesian Nonparametric Joint Factor Model for Learning Shared and Individual Subspaces from Multiple Data Sources. *Proceedings of the* 2012 *SIAM International Conference on Data Mining*, Anaheim, 26-28 April 2012, 200-211. https://doi.org/10.1137/1.9781611972825.18

[47] Canini, K.R., Shashkov, M.M. and Griffiths, T.L. (2010) Modeling Transfer Learning in Human Categorization with the Hierarchical Dirichlet Process. *The* 27*th International Conference on Machine Learning* (*ICML* 2010), Haifa, 21-24 June 2010, 151-158.

[48] Kang, J.H., Ma, J. and Liu, Y. (2012) Transfer Topic Modeling with Ease and Scalability. *Proceedings of the* 2012 *SIAM International Conference on Data Mining*, Anaheim, 26-28 April 2012, 564-575. https://doi.org/10.1137/1.9781611972825.49

[49] Elvira, C., Chainais, P. and Dobigeon, N. (2017) Bayesian Nonparametric Principal Component Analysis.

[50] Hill, J.L. (2011) Bayesian Nonparametric Modeling for Causal Inference. *Journal of Computational and Graphical Statistics*, **20**, 217-240. https://doi.org/10.1198/jcgs.2010.08162

[51] Jiang, Y. and Saxena, A. (2013) Infinite Latent Conditional Random Fields for Modeling Environments through Humans. *Robotics*: *Science and Systems*, Berlin, 24-28 June 2013, 1-8. https://doi.org/10.15607/RSS.2013.IX.034

[52] Plagemann, C., Kersting, K., Pfaff, P. and Burgard, W. (2007) Gaussian Beam Processes: A Nonparametric Bayesian Measurement Model for Range Finders. *Robotics*: *Science and Systems* (*RSS*'07), Atlanta, 27-30 June 2007. https://doi.org/10.15607/RSS.2007.III.018

[53] Xing, E.P. and Sohn, K. (2007) Hidden Markov Dirichlet Process: Modeling Genetic Inference in Open Ancestral Space. *Bayesian Analysis*, **2**, 501-527. https://doi.org/10.1214/07-BA220

[54] Xing, E.P., Sohn, K.A., Jordan, M.I. and Teh, Y.W. (2006) Bayesian Multi-Population Haplotype Inference via a Hierarchical Dirichlet Process Mixture. *Proceedings of the* 23*rd International Conference on Machine Learning*, Pittsburgh, 25-29 June 2006, 1049-1056. https://doi.org/10.1145/1143844.1143976

[55] Lijoi, A., Mena, R.H. and Prünster, I. (2007) A Bayesian Nonparametric Method for Prediction in EST Analysis. *BMC Bioinformatics*, **8**, Article No. 339. https://doi.org/10.1186/1471-2105-8-339

[56] Haines, T.S. and Xiang, T. (2013) Background Subtraction with Dirichlet Process Mixture Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **36**, 670-683. https://doi.org/10.1109/TPAMI.2013.239

[57] Sudderth, E.B., Torralba, A., Freeman, W.T. and Willsky, A.S. (2008) Describing Visual Scenes Using Transformed Objects and Parts. *International Journal of Computer Vision*, **77**, 291-330. https://doi.org/10.1007/s11263-007-0069-5

[58] Fox, E.B., Sudderth, E.B., Jordan, M.I. and Willsky, A.S. (2008) An HDP-HMM for Systems with State Persistence. *Proceedings of the* 25*th International Conference on Machine Learning*, Helsinki, 5-9 July 2008, 312-319. https://doi.org/10.1145/1390156.1390196

[59] Goldwater, S., Griffiths, T.L. and Johnson, M. (2009) A Bayesian Framework for Word Segmentation: Exploring the Effects of Context. *Cognition*, **112**, 21-54. https://doi.org/10.1016/j.cognition.2009.03.008