

贝叶斯分层模型变分推理与概率编程方法综述

令晓明, 范少良, 王文强, 顾楠, 张凯越

Overview of variational inference methods and probabilistic programming of Bayesian hierarchical models

LING Xiaoming, FAN Shaoliang, WANG Wenqiang, GU Yannan, and ZHANG Kaiyue

引用本文:

令晓明, 范少良, 王文强, 等. 贝叶斯分层模型变分推理与概率编程方法综述[J]. *铁路计算机应用*, 2022, 31(3): 6-11.

LING Xiaoming, FAN Shaoliang, WANG Wenqiang, et al. Overview of variational inference methods and probabilistic programming of Bayesian hierarchical models[J]. *Railway Computer Application*, 2022, 31(3): 6-11.

在线阅读 View online: <http://tljsjyy.xml-journal.net/2022/13/6>

您可能感兴趣的其他文章

Articles you may be interested in

基于贝叶斯网络的地铁牵引变电所可靠性分析

Reliability analysis of metro traction substation based on Bayesian Network

铁路计算机应用. 2019, 28(8): 68-74

基于EPOLL编程模型的全路车流径路服务的开发与应用

Development and application of whole-railway wagon flow routing services with EPOLL programming model

铁路计算机应用. 2020, 29(10): 13-15

基于粗糙集和模糊推理的高铁扼流适配变压器故障诊断研究

Fault diagnosis of choke adapter transformer based on rough set and fuzzy inference for high-speed railway

铁路计算机应用. 2017, 26(4): 1-5

基于LSTM的云环境异常智能检测方法研究

Anomaly intelligent detection method in cloud environment based on LSTM

铁路计算机应用. 2020, 29(6): 44-48,53

基于非线性规划的铁路应急物资储备点选址模型方案研究

Location model scheme of railway emergency material reserve points based on nonlinear programming

铁路计算机应用. 2019, 28(10): 5-7,15

基于参数优化VMD和SPWVD的轨道波磨辨识方法

Rail corrugation identification method based on parameter optimization VMD and SPWVD

铁路计算机应用. 2020, 29(6): 18-24



关注微信公众号, 获得更多资讯信息

文章编号: 1005-8451 (2022) 03-0006-06

贝叶斯分层模型变分推理与概率编程方法综述

令晓明^{1,2}, 范少良^{1,2}, 王文强^{1,2}, 顾奕楠^{1,2}, 张凯越³

(1. 兰州交通大学 国家绿色镀膜技术与装备工程技术研究中心, 兰州 730070;

2. 兰州交通大学 光电技术与智能控制教育部重点实验室, 兰州 730070;

3. 兰州交通大学 自动化与电气工程学院, 兰州 7300070)

摘要: 在贝叶斯模型中, 往往无法解析计算后验概率, 在实践中依赖于近似推理。变分推理 (VI, Variational Inference) 是重要的确定性近似推理方法, 比马尔科夫链蒙特卡罗 (MCMC, Markov Chain Monte Carlo) 采样具有更高的计算效率, 在大数据时代有着显著优势。文章通过贝叶斯分层模型, 回顾了经典 VI, 分析了随机变分推理 (SVI, Stochastic Variational Inference) 及其在主题模型中的应用, 综述了更稳健的 VI, 概述了概率编程系统的研究进展, 并对 VI 的未来发展趋势做出了展望。

关键词: 贝叶斯; 变分推理; 鲁棒变分推理; 主题模型; 概率编程

中图分类号: U29: TP391 **文献标识码:** A

DOI: 10.3969/j.issn.1005-8451.2022.03.02

Overview of variational inference methods and probabilistic programming of Bayesian hierarchical models

LING Xiaoming^{1,2}, FAN Shaoliang^{1,2}, WANG Wenqiang^{1,2}, GU Yannan^{1,2}, ZHANG Kaiyue³

(1. National Engineering Research Center for Technology of Environmental Deposition, Lanzhou Jiaotong University, Lanzhou 730070, China; 2. Key Laboratory of Opto-technology and Intelligent Control of Ministry of Education, Lanzhou Jiaotong University, Lanzhou 730070, China; 3. School of Automation and Electrical Engineering of Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract: In Bayesian model, it is often impossible to calculate the posterior analytically, and in practice, it relies on approximate inference. Variational Inference (VI) is an important deterministic approximate inference method, which has a higher computational efficiency than Markov Chain Monte Carlo (MCMC) sampling, and has a distinct advantage in the era of big data. This paper reviewed classical VI, analyzed Stochastic Variational Inference (SVI) and its application in subject model, reviewed more robust VI, summarized the research progress of probabilistic programming system, and prospected the future development trend of VI.

Keywords: Bayes; variational inference(VI); topic model; robust VI; probabilistic programming

贝叶斯分层模型中的推理通过在概率模型中利用贝叶斯定理, 计算给定数据条件下模型的条件概率, 进行模型的估计及对数据的预测, 依赖于推理算法估计后验概率。然而, 后验概率因计算开销巨大难以取得闭合解, 在实践中往往需要近似推理, 主要有两种方法: 马尔科夫链蒙特卡罗 (MCMC, Markov chain Monte Carlo) 和变分推理 (VI, variational inference)。多年来, MCMC 是近似推理的主流范式, 能够实现对目标分布较为准确的近似, 但运算缓慢。

VI 将高维度推理问题转化为优化问题, 实现了更快速的近似计算, 凭借其更快的收敛速度及对海量数据的强适应性, 受到众多研究者的青睐。

1 研究现状

近年来 VI 领域的研究集中于以下 5 个方面。

1.1 提升计算效率

随机变分推理^[1] (SVI, Stochastic Variational Inference) 及其变体^[2] 通过引入随机优化方法, 分别在大型数据集和稀疏情况中取得良好表现。进一步地, Ranganath 等人^[3] 开发了随机推理中的自适应学习率的设置方法; 坍塌变分推理^[4-6] (CVI, Collapsed

收稿日期: 2021-09-10

基金项目: 国家自然科学基金项目 (16663021)

作者简介: 令晓明, 教授; 范少良, 在读硕士研究生。

Variational Inference) 通过积分消除部分变量, 简化了计算。

1.2 重构变量间的依赖关系

文献 [7-9] 使用不同的方式建立了保持一定依赖关系的结构, 可以描述更丰富的变分族并提升 VI 的准确性, 但会引起计算量的显著增加。

1.3 通用变分推理框架与自动化推理

文献 [10] 将原本的变分分布映射为更简单的分布并从中采样; 文献 [11] 对变分目标中难以处理的对象应用蒙特卡罗梯度; 文献 [12] 通过引入随机优化、蒙特卡罗估计和方差减少技术实现了快速收敛。其他具备“黑盒”性的推理模型还包括层次变分模型 (HVM, Hierarchical Variational Models) [8] 和变分高斯过程 (VGP, Variational Gaussian Process) [13] 等。

1.4 提高算法鲁棒性

文献 [14] 通过约束变分参数的后续迭代, 克服了对初始化的敏感性, 且能达到更好的局部最优; 文献 [15] 通过改造近似计算的目标, 在统计和计算特性中取得了较好的平衡。

1.5 概率编程系统

Stan^[16]、PyMC3^[17] 和 Edward^[18] 是 3 款具有代表性且比较完备的概率编程系统, 具有直观、可读、功能强大的语法, 在不同程度上支持 VI 和 MCMC 采样。

大数据时代下, 铁路故障、事故信息数据量庞大, 文本中的非结构化数据蕴藏着大量的信息^[19]。使用变分推理的主题模型在挖掘文本隐藏信息中具有效率上的独特优势^[20-23]。此外, 主题模型还可用于对扣件的识别检测等^[24-27], 使用 VI 可以显著提升识别效率。

2 贝叶斯分层模型与 VI

2.1 贝叶斯分层模型

贝叶斯推理将模型框定为具有 N 个局部隐变量 $z_{1:N}$ 和观测变量 $x_{1:N}$ 的模型, 另有控制任意局部上下文的全局隐变量 β 。 $p(\cdot)$ 为概率密度函数, 贝叶斯模型中推理的目标是计算后验密度 $p(z, \beta | x) = p(z, \beta, x) / p(x)$, 其中, $p(x)$ 称为证据, 对于许多具有表现力的模型,

$p(x)$ 难以计算, 需要通过其他方法实现对证据的近似。

2.2 VI 框架

VI 将近似推理视为一个优化问题。(1) 引入一个关于隐变量的分布族, 该分布族被构造为各分量独立的平均场形式; (2) 以 KL (Kullback-Leibler) 散度定义变分目标, 通过坐标上升算法迭代得到参数更新方程; (3) 若满足指数族共轭条件, 则能够获得更简单有效的更新形式。但经典框架也暴露出许多问题, 包括参数的更新都依赖于其他参数与观测变量, 无法从无意义的初始化中恢复, 优化计算的过程方差偏大间接影响收敛速度, 易陷入糟糕的局部最优, 依赖于对特定模型的推导等。

2.3 SVI 与主题模型

2.3.1 SVI 概述

SVI 使用自然梯度纠正了经典梯度可能导致错误方向的问题^[28], 并通过以下 3 步有效减少了计算量:

(1) 引入黎曼测度使对数规整器对自然参数的 Hessian 矩阵不再在参数更新方程中出现。

(2) 通过引入随机优化, 将跟踪目标转化为噪声估计^[29]。

(3) 从具有 n 个局部的某局部隐变量 $z_{1:n}$ 中采集单样本 $z_i, i \in \{1, \dots, n\}$, 进而替换 $z_{1:n}$, 导出全局变分参数关于步长的线性更新方式。

2.3.2 主题模型

主题模型是文本挖掘的重要模型, 潜在狄利克雷分配模型 (LDA, Latent Dirichlet Allocation) 是其典型代表。VI 是 LDA 参数估计的重要方法, SVI 可以根据局部的拟合结果得到中间全局变分参数, 并将下一次迭代的全局变分参数设置为当前参数的线性组合, 从而显著提升效率。

随着铁路系统故障文本信息日渐庞大, 基于 LDA 及其变体的 SVI 技术将有助于快速分析故障文本数据, 并发掘热点问题, 对铁路系统故障的预防和分析提供重要依据。对于同样数量庞大的图像信息, 基于主题词包模型将图像量化为图像—单词矩阵用于分类, 度量空间不同造成的距离偏差可通过引入 SVI 的自然度量予以纠正, 从而更准确地实现基于近义词分配的铁路扣件等设备的状态检测。

3 更稳健的变分推理

VI 使用 KL 散度定义变分目标函数，其本质是一个非凸优化问题，包含了许多局部最优解，所采用的迭代算法难以避免陷入糟糕的局部最优解且对初始化参数选取敏感。

(1) 邻近变分推理 (PVI, Proximity Variational Inference) 建立在邻近视角上，认为梯度上升中的每一步是对前一步参数泰勒展开的有约束最小化^[30-31]，强制下一个点必须在以前一个点为球心、半径受控于步长的欧氏球内；(2) 算子变分推理 (OPVI, Operator Variational Inference) 是一种将原变分目标重新设计为算子变分目标的通用框架，基于强化梯度/重参数化梯度计算梯度期望的蒙特卡罗估计，且使用 Rao-Blackwellization 和控制变量技术来减小方差；(3) 变分回火 (VT, Variational Tempering)^[32] 引入了确定性退火，通过在模型中赋予异常值更高的温度，降低了异常值对全局变分参数的影响，提升了局部最优的鲁棒性；(4) 信赖域变分推理 (TRVI, Trust Region Variational Inference)^[33] 使用信赖域步长代替 SVI 中的自然梯度步长，有效改善了 SVI 陷入糟糕局部最优解的情况，也降低了对超参数的敏感性；(5) 总体经验贝叶斯^[34] 明确地将经验总体分布作为贝叶斯分析的一部分，进一步得到碰撞变分推理 (bump-VI)，将优化问题扩展为在寻找最优近似后验分布的同时寻找潜在数据集的索引。

更稳健的 VI 算法使得计算量明显增加。对鲁棒性的要求过高会直接导致算法更接近于全局寻优，削弱 VI 方法的效率优势。为实现有效的计算，需要避免在错误方向上的步进，保持对新加入数据点的鲁棒性，还要能从浅局部最优解中脱离。满足这些要求的算法在优化过程中能够进一步收集局部的信息，或重新设计推理框架，从而获得全新的特性，有助于形成更准确、统计性质更好的近似结果。

4 概率编程

4.1 概率编程语言概述

概率编程语言 (PPL, Probabilistic Programming Language) 是用于描述带有概率的生成过程的规范化

语言，其目标是用简洁的形式描述一个概率模型，并实现自动的学习与推理，用尽量少的代码实现使用者的想法，显著减轻 VI 研究者编写繁复程序的负担。

(1) Stan 是一种用于指定统计模型的 PPL，定义了以特定数据和常数为条件参数的对数概率函数，能实现变分贝叶斯、期望传播和使用近似积分的边际推断等算法。Stan 比传统声明式图形建模语言更具表现力，可支持使用类型声明变量、局部变量和条件语句。(2) PyMC3 是运行在 Python 语言中的开源 PPL，基于 NUTS 和 HybridMC^[35] 并与众多科学计算库集成，具有强大的表达能力。虽然 PyMC3 中大多数面向用户的功能都是用纯 Python 编写的，但其利用 Theano^[36] 将模型转码为 C 并编译为机器代码，从而提高了性能。(3) Edward 对程序的表达方式具有模型对计算图的表示不变且推理对计算图的表示不变的特点，使得对复杂模型的建模变得简单，且所有的计算都只是象征性地记录在随机变量上，简化了在执行程序之前的确定性和随机性操作^[37-38]，其中，符号表示不需要具体化整个模型，避免了具体化大型模型带来的不合理的内存消耗^[39]，因而，具有极高的执行效率。

4.2 概率编程语言评价

(1) 考虑 VI 的支持性，Edward 和 PyMC3 都支持较多的 VI 算法，相较 Stan 等其他算法匮乏的语言而言是更好的选择；(2) 考虑指定贝叶斯任务的代码量、语言的自然性和概念的可解释性，Edward 和 PyMC3 的使用显著易于 Stan，而 Edward 后续被集成到 TensorFlow 中，易用性得到进一步提升；(3) 在完成贝叶斯推理任务的速度方面，得益于 GPU 加速支持，Edward 和 PyMC3 的运算速度更快，而 PyMC3 并没有在所有的计算中都使用 Theano，导致其效率低于 Edward；(4) 在所采用算法的鲁棒性和 API 的稳定性方面，Stan 较为稳定，它没有大规模使用随机优化技术，且不依赖于 Theano，保证了执行的可靠性。

5 研究展望

VI 是机器学习领域重要的推理方法，该领域进一步的研究主要集中于以下几个方面。

5.1 VI 与 MCMC 结合

近似推理的核心问题是有效性和准确性的矛盾，如何将 MCMC 与 VI 结合一直以来都是近似推理研究的核心命题。目前，仅有少数研究在部分计算环节中做出了尝试，几乎都是基于特定问题做出的小幅改进，还需进一步探索出明确的 VI-MCMC 接口，这一方向的研究对于未来 VI 方法及应用的发展至关重要。

5.2 非共轭推理

VI 方法的发展需要突破共轭限制，有学者提出适用于逻辑回归的变分界^[40]。非共轭模型中的推理可以使用 Laplace 逼近和 Delta 方法^[41]，但仅限于连续变量。Wand^[42] 等人使用辅助变量、求积和混合近似法对指数族之外的分布族进行了探索。但非共轭模型是一个广泛的研究范畴，这些方法尚不足以支撑非共轭模型的推理。基于这一现状，非共轭条件下的推理或逼近方法将作为 VI 研究中自动化推理方向的主要命题长期存在。

5.3 与深度学习融合

对于深度学习所表现出的解释性不足和难以包含先验知识等缺陷，有学者使用贝叶斯方法，如变分自编码器（VAE, Variational Auto Encoder）及其变种改善了这些问题。有学者赋予了 VAE 解释性较强的先验模型^[43-44]，在这些模型中合理使用 VI 将带来效率上的飞跃，使得 VAE 具有更好的生成结果。此外，由于 VAE 将数据强行投射到有限维度的指定分布上，导致生成信息质量不高，而 GAN（Generative Adversarial Network）使用神经网络学习该分布，虽然在匹配分布的过程中有可能发生模型坍塌，但其生成的数据质量较高。将 VAE 与 GAN 有机结合是一个极具潜力的研究方向。

5.4 基于分布测度的问题

VI 使用 KL 散度作为分布测度定义变分目标，对 VI 的一系列研究都是从 KL 散度开始的。考虑使用其他散度如 α 散度^[45]、 f 散度^[46]、Stein 散度^[47-48]、JS 散度^[49-50] 等定义的变分目标可以获得更紧致的下界或更快的收敛速度，但基于这些散度的推理方法的研究尚不成熟，其计算方法、统计特性等方面的

研究尚不全面和深入。

5.5 概率编程

尽管概率编程已取得一定的发展，但近年来发展出的众多推理方法在编程实现上仍存在困难。概率编程库中可以使用的 API 接口仍不够丰富，方差减少技术中的控制变量方法仍依赖于模型的特定设计，不利于算法的快速收敛。

5.6 在铁路领域的应用

铁路领域的信息种类繁多、数目庞大且存在大量非结构化数据，而数据间的关联与内在趋势很难被发现，需要具有较强表达能力的模型和快速的计算方法。使用主题模型与 VI 是较为理想的选择，但其准确性和鲁棒性都有待提升。除文本外，VI 在图像和语音方面的研究已有进展，但在铁路领域的应用较少。此外，随着大规模贝叶斯网络被用于解决交通流问题，VI 也更能发挥计算效率上的优势。

6 结束语

近似推理在很大程度上解决了贝叶斯模型中的推理问题，有着广泛的应用。VI 作为近似推理中除 MCMC 之外的另一方法，具有十分重要的意义，应该更深入地研究其理论、方法与实践。目前，针对 VI 方法的研究成果颇丰，在当前的大数据时代，有必要进一步研究和扩展计算更快、更稳定的方法。VI 的实践包括概率编程和应用，已出现一批性能优异且能够与深度学习相结合的概率编程系统，为 VI 在工业中的应用打下了良好的基础。

参考文献

- [1] Hoffman M, Blei D M, Wang C, et al. Stochastic variational inference [J]. *Computer Science*, 2012, 14(1): 1303-1347.
- [2] Mimno D, Hoffman M D, Blei D M. Sparse stochastic inference for latent Dirichlet allocation[C]//*Proceedings of the 29th International Conference on Machine Learning*. Edinburgh, UK, 2012: 1515-1522.
- [3] Ranganath R, Wang C, Blei D M, et al. An adaptive learning rate for stochastic variational inference[C]//*Proceedings of the 30th International Conference on International Conference on Machine Learning*. Atlanta, USA, 2013: 298-306.
- [4] Hensman J, Rattray M, Lawrence N D. Fast variational inference in the conjugate exponential family[C]//*Proceedings*

- of the 26th International Conference on Neural Information Processing Systems. Lake Tahoe, USA, 2012: 2897-2905.
- [5] King N, Lawrence N. Fast variational inference for gaussian process models through KL-correction[C]//Proceedings of the 17th European Conference on Machine Learning. Berlin, Germany, 2006: 270-281.
- [6] Kurihara K, Welling M, Teh Y W. Collapsed variational Dirichlet process mixture models[C]//Proceedings of the 20th International Joint Conference on Artificial Intelligence. Hyderabad, India, 2007: 2796-2801.
- [7] Hoffman M D, Blei D M. Structured stochastic variational inference [J]. Computer Science, 2014, 14(1): 1303-1347.
- [8] Ranganath R, Tran D, Blei D M. Hierarchical variational models[C]//Proceedings of the 33rd International Conference on International Conference on Machine Learning. New York, USA, 2016: 324-333.
- [9] Tran D, Blei D M, Airoldi E M. Copula variational inference[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Montreal, Canada, 2015: 3564-3572.
- [10] Kingma D P, Welling M. Auto-encoding variational Bayes[C]//Proceedings of the 2nd International Conference on Learning Representations. Banff, Canada, 2014.
- [11] Paisley J, Blei D M, et al. Variational Bayesian inference with stochastic search[C]//Proceedings of the 29th International Conference on International Conference on Machine Learning, Edinburgh, UK, 2012: 1363-1370.
- [12] Ranganath R, Gerrish S, Blei D M. Black box variational inference[C]//Proceedings of the 17th International Conference on Artificial Intelligence and Statistics. Reykjavik, Iceland, 2014: 814-822.
- [13] Tran D, Ranganath R, Blei D M. Variational Gaussian process[C]//Proceedings of 4th International Conference on Learning Representations. San Juan, Puerto Rico, 2015.
- [14] Altosaar J, Ranganath R, Blei D M. Proximity variational inference[C]//Proceedings of the 21st International Conference on Artificial Intelligence and Statistics. Playa Blanca, Lanzarote, Canary, Islands, Spain, 2018: 1961-1969.
- [15] Ranganath R, Altosaar J, Tran D, et al. Operator variational inference[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Barcelona, Spain, 2016, 496-504.
- [16] 袁 婕. 基于机器学习的铁路系统故障文本数据分析挖掘方法研究[D]. 北京: 北京交通大学, 2020.
- [17] 张国琛. 基于词扩展LDA的铁路事故致因分析方法研究[D]. 北京: 北京交通大学, 2019.
- [18] 王 峰. 基于文本挖掘的高铁车载设备故障诊断方法研究[D]. 北京: 北京交通大学, 2016.
- [19] 胡福威. 基于Labeled-LDA的列控车载设备故障特征提取与诊断方法研究[D]. 北京: 北京交通大学, 2018.
- [20] 刘 玄. 基于文本挖掘的高铁经济研究现状及趋势研究[D]. 北京: 北京交通大学, 2019.
- [21] 陈小艳. 融合结构信息的LDA扣件状态识别研究[D]. 成都: 西南交通大学, 2015.
- [22] 杨 飞, 罗建桥, 李柏林. 基于局部特征编码的LDA铁路扣件检测模型 [J]. 传感器与微系统, 2020, 39 (2): 148-152.
- [23] 罗建桥. 面向铁路扣件图像特征的结构化St-LDA语义学习方法研究[D]. 成都: 西南交通大学, 2016.
- [24] 欧 阳, 罗建桥, 李柏林, 等. 嵌入标记信息的铁路扣件状态检测主题模型 [J]. 计算机工程, 2018, 44 (6): 200-206.
- [25] Hoffman M D. Stan: A probabilistic programming language [J]. Journal of Statistical Software, 2017, 76(1): 1-32.
- [26] Salvatier J, Wiecki T V, Fonnesbeck C. Probabilistic programming in Python using PyMC3[J]. PeerJ Computer Science, 2015.
- [27] Tran D, Hoffman M D, Saurous R A, et al. Deep probabilistic programming[C]//Proceedings of the 5th International Conference on Learning Representations. Toulon, France, 2017.
- [28] Amari S I. Natural gradient works efficiently in learning [J]. Neural Computation, 1999, 10(2): 251-276.
- [29] Robbins H, Monro S. A stochastic approximation method [J]. *Annals of Mathematical Statistics*, 1951, 22(3): 400-407.
- [30] Spall J C, Nowak W. Introduction to stochastic search and optimization. estimation, simulation, and control [J]. IEEE Transactions on Neural Networks, 2007, 18(3): 964-965.
- [31] Boyd S, Vandenberghe L. Convex optimization[M]. Cambridge: Cambridge University Press, 2004.
- [32] Mandt S, McInerney J, Abrol F, et al. Variational tempering[C]//Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, Cadiz, Spain, 2016: 704-712.
- [33] Theis L, Hoffman M. A trust-region method for stochastic variational inference with applications to streaming data[C]//Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 2015: 2503-2511.
- [34] Kucukelbir A, Blei D M. Population empirical bayes[C]//Proceedings of the 31st Conference on Uncertainty

- in Artificial Intelligence, Amsterdam, The Netherlands, 2015: 444-453.
- [35] Duane S, Kennedy A D, Pendleton B J, et al. Hybrid Monte Carlo [J]. *Physics Letters*, 1987, 195(2): 216-222.
- [36] Bergstra J, Breuleux O, Bastien F, et al. Theano: A CPU and GPU math expression compiler[C]//Proceedings of the Python for Scientific Computing Conference (SciPy). Austin, USA, 2010.
- [37] Ścibior A, Ghahramani Z, Gordon A D. Practical probabilistic programming with monads[C]//Proceedings of the 2015 Symposium on Haskell. Vancouver, Canada, 2015: 165-176.
- [38] Zinkov R, Shan C C. Composing inference algorithms as program transformations[C]//Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence. Sydney, Australia, 2017: 163-172.
- [39] Tristan J B, Huang D, Tassarotti J, et al. Augur: Data-parallel probabilistic modeling [J]. *Advances in neural information processing systems*, 2014(3): 2600-2608.
- [40] Jaakkola T, Jordan M I. Bayesian parameter estimation via variational methods [J]. *Statistics and Computing*, 2000, 10(1): 25-37.
- [41] Wang C, Blei D M. Variational inference in nonconjugate models [J]. *Journal of Machine Learning Research*, 2013(4): 1005-1031.
- [42] Wand M, Ormerod J, Padoan S, et al. Mean field variational Bayes for elaborate distributions [J]. *Bayesian Analysis*, 2011(6): 847-900.
- [43] Johnson M J, Duvenaud D, Wiltchko A B, et al. Structured VAEs: Composing probabilistic graphical models and variational autoencoders.[EB/OL].[2021-09-10].<https://arxiv.org/abs/1603.06277v5>.
- [44] Deng Z W, Navarathna R, Carr P, et al. Factorized variational autoencoders for modeling audience reactions to movies [C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA:IEEE, 2017: 6014-6023.
- [45] Li Yingzhen, Turner R E. Rényi divergence variational inference[C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Bracelona, Spain, 2016: 1081-1089.
- [46] Wan Neng, Li Dapeng, Hovakimyan N. F-divergence variational inference[C]//Advances in Neural Information Processing Systems 33, 2020: 17370-17379.
- [47] Liu Qiang, Wang Dilin. Stein Variational Gradient Descent: A general purpose Bayesian inference algorithm[C]//Advances in Neural Information Processing Systems 29. Bracelona, Spain, 2016: 2370-2378.
- [48] Liu Yang, Ramachandran P, Liu Qiang, et al. Stein variational policy gradient[C]//Proceedings of the 33rd Conference on Uncertainty in Artificial Intelligence. Sydney, Australia, 2017: 239-247.
- [49] Majtey A P, Lamberti P W, Prato D P. Jensen-Shannon divergence as a measure of distinguishability between mixed quantum states [J]. *Physical Review A*, 2005, 72(5): 052310.
- [50] Weikum G. Foundations of statistical natural language processing [J]. *SIGMOD RECORD*, 2002, 31(3): 37-38.

责任编辑 李依诺