

概率图模型推理方法的研究进展

刘建伟 崔立鹏 黎海恩 罗雄麟

(中国石油大学(北京)自动化研究所 北京 102249)

摘 要 近年来概率图模型已成为不确定性推理的研究热点,在人工智能、机器学习与计算机视觉等领域有广阔的应用前景。根据网络结构与查询问题类型的不同,系统地综述了概率图模型的推理算法。首先讨论了贝叶斯网络与马尔可夫网络中解决概率查询问题的精确推理算法与近似推理算法,其中主要介绍精确推理中的 VE 算法、递归约束算法和团树算法,以及近似推理中的变分近似推理和抽样近似推理算法,并给出了解决 MAP 查询问题的常用推理算法;然后分别针对混合网络的连续与混合情况阐述其推理算法,并分析了暂态网络的精确推理、近似推理以及混合情况下的推理;最后指出了概率图模型推理方法未来的研究方向。

关键词 概率图模型, VE 算法, 团树算法, 变分推理, 抽样推理, MAP 推理, 混合网络推理, 暂态网络推理

中图分类号 TP181

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2015.4.001

Research and Development on Inference Technique in Probabilistic Graphical Models

LIU Jian-wei CUI Li-peng LI Hai-en LUO Xiong-lin

(Research Institute of Automation, China University of Petroleum, Beijing 102249, China)

Abstract In recent years, probabilistic graphical models have become the focus of the research in uncertainty inference, because of their bright prospect for the application in artificial intelligence, machine learning, computer vision and so forth. According to different network structures and query questions, the inference algorithms of probabilistic graphical models were summarized in a systematic way. First, exact and approximate inference algorithms for solving the probability queries in Bayesian network and Markov network were discussed, including variable elimination algorithms, conditioning algorithms, clique tree algorithms, variational inference algorithms and sampling algorithms. The common algorithms for solving MAP queries were also introduced. Then the inference algorithms in hybrid networks were described respectively for continuous or hybrid cases. In addition, this work analyzed the exact and approximate inference in temporal networks, and described inference in continuous or hybrid cases for temporal networks. Finally, this work raised some questions that the inference algorithms of probabilistic graphical models are facing with and discussed their development in the future.

Keywords Probabilistic graphical model, Variable elimination, Clique tree, Variational inference, Sampling inference, MAP inference, Hybrid network inference, Temporal network inference

1 引言

概率图模型的提出极大丰富了不确定性推理体系,为许多其他相关领域提供了新技术和新思想。近年来,概率图模型已成功应用在医疗诊断^[1]、故障诊断^[2]、图像处理^[3]、语音识别^[4]与信息解码^[5]等相关领域。概率图模型之所以能在众多不同领域得到广泛应用,正是由于它能在简洁紧凑的表示框架中构建复杂的概率信息,并能够通过推理算法自动获取概率信息中的含义从而做出结论和决策。概率图模型结合了概率论与图论的知识,通过图来表示随机变量间的依赖关系,为多变量统计建模提供有力的框架。由于概率图模型处理不

确定性推理的强大能力,关注其已有研究成果和未来发展势均十分必要。概率图模型的研究包括表示、推理与学习 3 部分,本文针对概率图模型的推理方法进行讨论。

目前,国内外已有多个文献对概率图模型的推理方法进行综述,例如,文献[6]对概率图模型的变分推理进行了简要综述;文献[7]综述了贝叶斯网络中的部分精确推理与近似推理算法,并将其与概率图学习算法作了比较;文献[8]给出了实时贝叶斯网络的推理方法综述;而国内的文献[9]也对变分近似推理算法进行了综述。但是,这些综述文献并没有对概率图模型的推理方法进行全面综述,而且由于概率图模型的推理方法发展非常迅速,因此本文对概率图模型的推理方法

到稿日期:2014-04-29 返修日期:2014-08-23 本文受国家重点基础研究发展计划项目(973 计划)(2012CB720500),国家自然科学基金项目(21006127),中国石油大学(北京)基础学科研究基金项目(JCXK-2011-07)资助。

刘建伟(1966—),男,博士,副研究员,主要研究方向为智能信息处理、复杂系统分析和预测与控制、算法分析与设计,E-mail:liujw@cup.edu.cn;崔立鹏(1990—),男,硕士生,主要研究方向为机器学习等;黎海恩(1988—),女,硕士生,主要研究方向为机器学习等;罗雄麟(1963—),男,博士,教授,主要研究方向为智能控制。

进行更加全面的归纳,并总结了近几年的相关研究成果。

概率图模型的推理就是利用联合概率分布,在已知网络结构和证据的情况下,回答查询问题。网络的结构特征和查询问题的类型是选择和设计推理算法所要考虑的主要因素。与非概率图模型相比,概率图模型主要是利用变量之间的依赖和独立关系,将其用图结构表示,简化推理过程。因此,推理算法通常根据网络结构和查询问题的类型来分类。常见的网络结构有贝叶斯网络(Bayesian Network, BN)、马尔可夫网络(Markov Network, MN)、混合网络(Hybrid Network)和动态贝叶斯网络(Dynamic Bayesian Network, DBN)等。而查询问题一般分3种,包括概率查询、最大后验概率(Max A Posterior, MAP)查询以及边际 MAP 查询。

本文根据网络结构与查询问题类型的不同,对概率图模型的推理方法进行系统的分类。第2节先分别介绍几种常见的概率图模型和查询问题类型,并给出了概率图模型推理方法的分类。第3—8节则分别讨论不同的推理方法,其中第3节先介绍3种精确推理算法:VE算法、递归约束算法和团树算法;第4节和第5节分别讨论近似推理中的变分近似推理算法和抽样近似推理算法;第6节描述回答MAP查询的推理算法;第7节分析混合网络中的推理;第8节阐述暂态模型特别是动态贝叶斯网络的推理算法;最后给出总结和展望。

为了表示的统一性,先约定本文使用的符号。随机变量集为 $\mathbf{X}=\{X_1, \dots, X_n\}$,变量集的取值为 $\mathbf{x}=\{x_1, \dots, x_n\}$ 。一般地,变量为大写体,其取值为小写体。因子 ϕ 的值域即因子函数的取值集合,表示为 $R(\phi)$;因子 ϕ 的定义域即因子函数包含的变量集,表示为 $D(\phi)$ 。归一化分布为 P ,未归一化分布为 $\tilde{P}$ 。

2 概率图模型及其推理问题

概率图模型的推理方法通常根据网络结构和查询问题类型的不同而进行分类。本节先介绍几种常见的概率图模型和查询问题的描述,再根据不同的网络结构和查询问题对推理方法进行系统分类。

2.1 概率图模型

概率图模型^[10,11]的表示由网络结构和参数模型两部分组成。网络结构 $G=(\mathbf{N}, \mathbf{E})$ 由节点集 $\mathbf{N}$ 和边集 $\mathbf{E}$ 构成。 $\mathbf{N}$ 中的节点表示随机变量集 $\mathbf{X}=\{X_1, \dots, X_n\}$ 中的变量。边集 $\mathbf{E}$ 中的边表示两个变量间的概率关系,可以有向的或者无向的。参数模型是网络结构的概率分布模型。整个网络结构的概率分布为联合概率分布,但难以得到联合概率分布的显式表达,因此利用网络结构中变量的独立性,把高维联合概率分布分解为节点上的低维概率分布的乘积。目前,常见的概率图模型表示类型主要有BN、MN、混合网络和DBN等。

2.1.1 贝叶斯网络

BN的网络结构为有向无环图(Directed Acyclic Graph, DAG),节点表示变量,有向边表示变量间的条件依赖关系。BN的参数模型为节点随机变量的条件概率分布(Conditional Probabilistic Distribution, CPD),即已知父节点时该变量的条件概率分布。根据结构中隐含的独立性假设:已知父节点时, X_i 与其非子节点条件独立,那么可以把联合概率分布分解为节点CPD的乘积:

$$P(X_1, \dots, X_n) = \prod_i P(X_i | Pa_{X_i}) \quad (1)$$

其中, Pa_{X_i} 表示变量 X_i 的父节点。

2.1.2 马尔可夫网络

MN的网络结构为无向图,节点表示随机变量,无向边表示变量间的依赖关系。MN的参数模型为因子集,每个因子就是定义在无向图中某个团上的非负函数,因子也称为团的势函数。MN结构中的变量可划分为多个团,那么联合概率分布可以分解为每个团的因子的乘积:

$$P(X_1, \dots, X_n) = \frac{1}{Z} \prod_{k=1}^m \phi_k(\mathbf{C}_k) \quad (2)$$

其中, m 为团数, ϕ_k 为第 k 个团的因子, $\mathbf{C}_k$ 为第 k 个团的随机变量集, $Z = \sum_{X_1, \dots, X_n} \prod_{k=1}^m \phi_k(\mathbf{C}_k)$ 为划分函数。

2.1.3 混合网络

混合网络同时含有离散变量和连续变量。BN和MN都可扩展为混合网络,其网络结构不变,只是参数模型发生变化。由于BN的CPD比MN的因子更易于理解,下面以CPD的形式给出混合网络的参数模型。

当子节点为连续变量而父节点同时具有连续和离散变量时,一般采用条件线性高斯(Conditional Linear Gaussian, CLG)模型来表示其CPD。令 X 为连续变量,并令 $\mathbf{U}=\{U_1, \dots, U_m\}$ 为其离散父节点, $\mathbf{Y}=\{Y_1, \dots, Y_k\}$ 为其连续父节点。对于每个取值 $u \in R(\mathbf{U})$,存在 $k+1$ 个系数 $a_{u,0}, \dots, a_{u,k+1}$ 和一个方差 σ_u^2 使得 X 的条件线性高斯CPD为

$$p(X|u, \mathbf{y}) = N(a_{u,0} + \sum_{i=1}^k a_{u,i} y_i; \sigma_u^2) \quad (3)$$

当子节点为离散变量而父节点为连续变量时,一般使用阈值模型。例如,假设二值离散变量 U 有连续父节点 Y ,其阈值模型为

$$P\{u=1\} = \begin{cases} 0.8, & y \leq 80 \\ 0.05, & \text{others} \end{cases} \quad (4)$$

在阈值模型中,概率的变化是 Y 的一个非连续函数,因此也可以表示为逻辑斯蒂模型或进行多项式展开。

2.1.4 动态贝叶斯网络

DBN是最典型的暂态模型,是BN在时间轴上的扩展。DBN网络结构中的节点呈列状分布,每一列都表示一个时刻上的网络片段。随着时刻的推进,每一列之间存在某些节点的有向联系,从前一时刻的某些节点指向后一时刻的某些节点。从形式上来看,DBN由初始网络和转移网络共同定义。初始网络由初始时刻的节点列以及该列中节点间的有向联系组成,而转移网络描述了两列节点间的有向联系。假设DBN中的转移网络在任意两个时刻都是一致的。因此,只要已知初始网络和转移网络,就可以表示无限时间轨线上的概率分布。

2.2 查询问题

查询问题一般分3种:(1)概率查询,即计算后验概率分布 $P(\mathbf{Y}|\mathbf{E}=\mathbf{e})$,其中 $\mathbf{Y}$ 为查询变量集, $\mathbf{E}$ 为证据变量集;(2)最大后验概率(Max A Posterior, MAP)查询,也称为最有可能解释(Most Probable Explanation, MPE)查询,即求出非证据变量集的最有可能取值。令 $\mathbf{X}$ 为所有变量集合, $\mathbf{E}$ 为证据变量集, $\mathbf{W}=\mathbf{X}-\mathbf{E}$,MAP查询就是在已知证据 $\mathbf{E}=\mathbf{e}$ 时求出 $\mathbf{W}$ 中变量的最有可能取值。MAP查询可形式化表示为: $\text{MAP}(\mathbf{W}|\mathbf{e}) = \arg \max_{\mathbf{w}} P(\mathbf{w}, \mathbf{e})$,其中 $\arg \max_x f(x)$ 表示使 $f(x)$ 最大的 x 的取值;(3)边际MAP查询,其目的是求出非证据变量

集的某个子集的最有可能取值。令 $Y \subseteq W$, 边际 MAP 查询的形式化表示为: $MAP(y|e) = \arg \max_y P(y|e)$ 。从算法复杂性的角度看, MAP 查询是 NP 完全问题, 概率查询是 #P 完全问题, 而边际 MAP 查询是 NP^{PP} 完全问题。

2.3 概率图模型推理方法的分类

概率图模型的推理方法主要分精确推理和近似推理两大类。而根据网络结构和查询问题的不同, 概率图模型的推理方法可进一步细分。下面先根据网络结构的类型把推理方法分为 3 大类: BN 和 MN 的推理、混合网络的推理和 DBN 的推理, 然后再根据查询问题的形式对每一类推理方法继续细分。

2.3.1 BN 和 MN 的推理方法

下面根据 BN 和 MN 的概率查询以及 MAP 查询来对推理算法进行分类, 如图 1 所示。

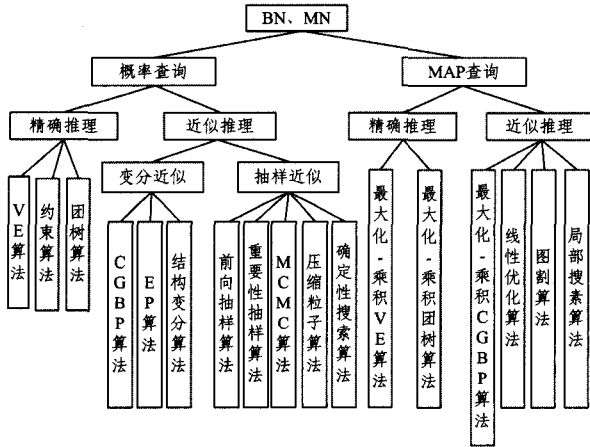


图 1 BN 和 MN 的推理算法分类

精确推理利用概率图模型的网络结构和参数模型, 完全按照基本概率公式来回答查询问题, 从而得到精确的查询结果。但是, 精确推理难以处理大型复杂概率图模型。如图 1 所示, BN 和 MN 中回答概率查询的精确推理算法有 3 种: (1) 消元 (Variable Elimination, VE) 算法^[12], 利用动态规划的思想, 交换乘积和求和运算的顺序, 逐步对非查询变量进行消元; (2) 递归约束 (Recursive Conditioning, RC) 算法^[13], 先对联合概率分布的约束变量进行消元, 然后对剩下的变量使用 VE 算法, 以有效简化求解过程; (3) 团树 (Clique Tree) 算法^[14], 也称为联合树 (Junction Tree) 算法, 利用原模型预构造一个团树结构, 在团之间进行消息传递并校准团树从而得到所求概率。

近似推理在原概率图模型的近似模型上推理, 所得到的结果也只是近似值。虽然近似推理降低了结果的精度, 但是它换取了能够处理大型复杂网络的能力。如图 1 所示, 近似推理主要有两类: 变分 (Variation) 近似推理和抽样 (Sampling) 近似推理。

变分近似推理^[15,16]把概率推理问题看作是变分优化问题, 目标函数为原分布与近似分布的距离度量熵函数, 在约束空间中最小化熵函数得到最优近似分布, 在近似分布上通过推理来回答查询问题。变分近似推理算法分 3 种: (1) 聚类图信任传播 (Cluster Graph Belief Propagation, CGBP) 算法^[17], 也称为环信任传播 (Loopy Belief Propagation) 算法, 把原网络结构近似为聚类图结构, 在聚类图中进行消息传递, 相当于把

目标函数近似为因子能量泛函, 把约束空间放松为伪边际分布空间; (2) 期望传播 (Expectation Propagation, EP) 算法^[18,19], 使用精确的团树结构, 但是在团树中传递近似的信息, 这可视作使用精确的目标函数和近似的伪边际分布空间进行推理; (3) 结构变分 (Structured Variation) 算法^[6,20], 考虑近似分布的结构问题, 使用精确的目标函数和近似的紧约束空间。

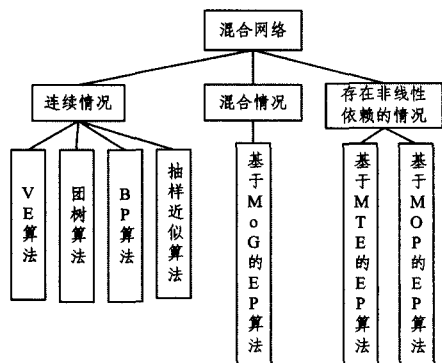
抽样近似推理使用粒子集来近似联合概率分布, 粒子即变量集或变量子集的取值实例, 利用这些粒子集来估计查询问题。根据粒子产生方法的不同, 抽样近似推理分 5 种: (1) 前向抽样 (Forward Sampling) 算法, 只适用于贝叶斯网络, 利用父节点与子节点的顺序关系, 在已知父节点时从子节点的后验分布中抽样; (2) 重要性抽样 (Importance Sampling) 算法^[21], 对变量赋予不同的重要性权值, 在残差网络给出的近似建议分布中使用前向抽样方法产生粒子, 并调整重要性权值来得到无偏估计; (3) 马尔可夫链蒙特卡罗 (Markov Chain Monte Carlo, MCMC) 算法^[22], 预定义平稳分布为真实后验分布的马尔可夫链, 由该链产生粒子; (4) 压缩粒子 (Collapsed Particles) 算法^[23], 通常也称为割集抽样算法或 Rao-Blackwellisation 算法, 压缩粒子为某个变量子集的取值实例, 剩下的变量使用解析表达式表示概率分布, 此过程相当于通过粒子来部分近似推理任务, 而在子网络中进行精确推理; (5) 确定性搜索 (Deterministic Search) 算法^[24], 以确定性非随机规则产生粒子, 使粒子集覆盖概率分布的主要部分, 虽然此算法不能提供无偏估计值, 但是能给出可靠的概率上下限。

前面的推理算法分类主要针对 BN 和 MN 的概率查询问题, 而 MAP 查询的推理算法与此稍有不同, 主要分为 6 种: (1) 最大化-乘积 VE 算法, 把求和运算替换为最大化运算, 并引入回溯过程使得消元完成时能计算出 MAP 取值; (2) 最大化-乘积团树算法, 团树中不再使用和积运算, 而是使用最大化-乘积运算; (3) 最大化-乘积聚类图信任传播 (Max-Product Cluster Graph Belief Propagation, 最大化-乘积 CGBP) 算法^[25,26], 把全局优化问题转化为局部优化问题, 即把求一致联合取值问题转化为求能优化信任的局部取值问题, 使得结果满足强优化性; (4) 线性优化算法^[27], 把推理问题看作线性优化问题, 该算法与凸最大化-乘积 BP 算法关系密切, 通过求解线性优化问题的对偶问题得到凸最大化-乘积 BP 算法的不动点; (5) 图割 (Graph Cut) 算法^[28,29], 其对象为两两关系二值马尔可夫随机场 (Markov Random Field, MRF), 当能量泛函满足子模性质时, 图割方法的运行时间比消息传递方法的更快; (6) 局部搜索算法^[30], 在特定取值空间上搜索 MAP 取值, 能很好地解决边际 MAP 查询问题。

2.3.2 混合网络的推理方法

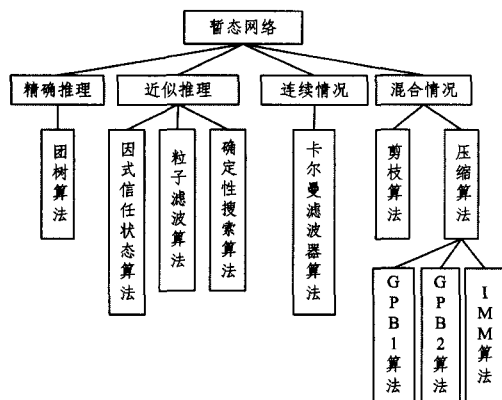
混合网络是同时包含离散变量和连续变量的网络, 由于含有连续变量的因子并没有解析表达式, 这为推理任务带来极大的挑战。如图 2 所示, 当变量全部为连续变量时, 使用线性高斯模型来表示其概率模型, 一般采用基于消息传递的推理算法来回答其查询问题, 如 VE 算法、团树算法和 BP 算法^[31,32]。当存在非高斯分布或者离散变量时, 可以把非高斯分布近似为高斯分布, 采用基于高斯混合分布 (Mixture of Gaussians, MoG) 的 EP 算法^[33]。如果混合网络中的变量存在非线性依赖, 则使用基于截断指数混合 (Mixtures of Trun-

cated Exponentials, MTE)^[34]或基于多项式混合(Mixtures of Polynomials, MOP)^[35]的 EP 算法。另外,混合网络还可以使用非参数推理算法,如重要性抽样、MCMC 和压缩粒子算法等。



2.3.3 暂态模型的推理方法

暂态模型中最重要的是动态贝叶斯网络(Dynamic Bayesian Network, DBN),通过把 DBN 展开并实例化观察变量,就可以在展开的网络上应用前面介绍的适用于 BN 的推理算法,但是展开的网络会相当大而难以处理。如图3所示, DBN 中的精确推理算法主要是团树算法^[36,37],消息从当前时间段传递到下一时间段,消息表示系统状态的信任。DBN 的近似推理算法主要有因式信任状态(Factored Belief State)算法^[38,39]、粒子滤波(Particle Filtering)算法^[40-42]和确定性搜索算法^[43]。因式信任状态算法实际上是 EP 算法的一种,把消息表示为因子乘积并赋予计算数。粒子滤波算法采用重要性抽样的思想,并引入再抽样过程,在样本集中去掉不良样本的同时选择优良样本进行复制。当 DBN 中变量全部为连续变量时,可应用卡尔曼滤波器算法来推理。对于混合 DBN,则需要使用剪枝算法或压缩算法来避免高斯混合项数的无限增大。压缩算法是在时间轴线上利用时间窗口把多个高斯分布压缩为一个高斯分布,如 GPB1 算法、GPB2 算法和交互式多模型(Interacting Multiple Model, IMM)算法^[44],这些方法也是 EP 算法的变型。



3 精确推理

精确推理是在回答概率查询问题时,根据概率图模型的网络结构和参数模型,使用基本概率公式来求解。只要每一步运算都是精确进行的,那么总算法得到的结果为精确值。

当概率图模型的变量增多或结构复杂时,精确推理可能产生指数数量的中间结果,从而占用大量存储空间,因此精确推理难以用于处理大型复杂网络。

BN 和 MN 中的精确推理算法有 3 种:VE 算法、约束算法和团树算法。VE 算法是精确推理中最基本的算法,只返回某个变量子集的边际分布。约束算法对 VE 算法进行了改进,使需要存储的中间结果大幅减少,节省存储空间,但也只返回某个变量子集的边际分布。团树算法是最有效的精确推理算法,能同时返回所有团信任,使得可以计算任意变量子集的边际分布。但是,团树算法的存储空间成本比前两种算法要大,而且因为团树的计算结构是预先固定好的,所以不能灵活利用某些证据和查询结构所带来的计算便利性。

3.1 消元算法

VE 算法^[12]就是在计算边际概率分布时对联合概率分布中某个或某些变量的取值进行求和运算,即把该变量消去。VE 算法的基本思想是:首先把联合概率分布分解成初始因子集的乘积,然后改变因子求和与乘积的运算顺序,以达到减少计算量的目的。初始因子集就是网络的 CPD 集。求某个变量的边际概率分布时,需消去其他变量,该消元过程分两步:第一步,把包含消元变量的所有因子相乘,得到新因子;第二步,在所得的新因子中对消元变量的所有取值情况求和。

概率查询的目的是计算 CPD $P(Y|E=e)$,根据概率公式

$$P(Y|E=e) = \frac{P(Y,e)}{P(e)} \quad (5)$$

可先利用 VE 算法分别求出 $P(Y,e)$ 和 $P(e)$,再代入上式求出 $P(Y|E=e)$ 。计算 $P(Y,e)$ 时,需要在算法执行过程中把证据 $E=e$ 代入到含有该证据变量的因子中,然后再消除 Y 和 E 外的其他变量,此时算法返回的是 $P(Y,e)$ 。再对 $P(Y,e)$ 中 Y 的所有取值进行求和得到 $P(e)$,即 $P(e) = \sum_y P(y,e)$ 。

Nima 等学者^[45,46]对传统的 VE 算法进行改进,提出带有任意约束的上升 VE(Lifted Variable Elimination)算法,该算法充分利用图结构的规律性,可以处理到目前为止只有近似推理才能灵活解决的某些大型网络的推理问题。

3.1.1 消元顺序

当消去多个变量时,消元顺序对算法的计算成本有很大影响。具有较小树宽的消元顺序使得推理过程存储的中间结果更少,极大地节约了计算成本。因此,消元之前必须确定成本最低的最优消元顺序。过去,学者认为消元顺序只影响精确推理算法。但最近研究发现,不少近似推理算法的计算成本和推理效果也受消元顺序的影响,如广义信任传播算法^[47]和最小桶消元算法^[48]等。

然而,确定最优消元顺序是 NP 难的,目前多采用近似方法来解决该问题^[49],常用的两种近似方法为最大基数搜索(Maximum Cardinality Search)^[50]和最小启发式成本贪婪搜索(Minimum Heuristic Cost Greedy Search)^[51]。文献^[52]提出了最小启发式成本贪婪搜索的停止准则,把停止准则形式化表示为回答概率查询所需的总时间,在寻找更好的消元顺序和利用目前找到的最好的消元顺序之间进行权衡,给出了 3 种不同的停止准则,获得了很好的效果。文献^[53]提出的迭代贪婪消元顺序(Iterative Greedy Variable Ordering)算法,可得到具有很小诱导宽度和状态空间大小的消元序列,效果比普通的贪婪搜索算法更好。

3.2 递归约束算法

RC算法^[13]的基本思想是通过约束变量把网络分解为多个小的子网络,分别独立求解子网络并递归使用RC算法。RC算法实际上是对联合概率分布中的约束变量进行求和消元,然后再应用VE算法进行其它变量的消元。算法具体过程为:对约束变量的每个取值都构建一个无向网络,这些网络结构相同而参数不同,然后在每个网络上运行一次VE算法,把每个网络返回的结果组合起来即可得到所求的概率查询。

RC算法能节省大量存储空间,并不需要存储大量的中间结果。但是,RC算法是通过牺牲运行时间来换取存储空间减少,对约束变量的不同取值,所有计算都要重复进行,在重复消元过程中会耗费较多的时间。文献[54]分析了RC算法在全缓存情况下的空间成本,说明了RC算法是BN精确推理中空间成本最低的算法之一。

可以通过改善约束算法来达到降低成本的目的,改善的途径有两种:一是交换约束和消元的顺序,通过适当地交换约束和消元的顺序,可以在约束某个变量之前先消去某些变量来简化网络,从而有效减少后续的计算成本;二是进行网络分解,选择能使网络分解为两个完全独立部分的变量作为约束变量,这样能分别在每个子网络上进行计算,最后把结果相加,有效节省运算时间。文献[55]还提出子缓存(Subcaching)技术,其在保持运行时间不变的情况下能进一步减小缓存大小。

BN的递归约束分解可表示为树的形式,如Darwiche^[56]提出的dtrees和Grant等人^[57]提出的消元树。这种树结构决定了推理过程中运算发生的顺序,影响概率查询的求解。文献[58]提出两种启发式方法来构造平衡消元树,使得构造的消元树更小,从而加快推理过程。而文献[59]考虑消元树的一般结构,利用无标签节点重构消元树,有效减少了推理过程的计算量。

3.3 团树算法

团树算法^[14]使用团树来表示联合概率分布,能便捷有效地实现多个节点的概率查询。团树算法先把贝叶斯网络转化为团树结构,再利用团之间的消息传递机制进行概率推理。

团树为树状的聚类图,每个团的初始团势为与该团有关的因子的乘积,团 C_j 的初始团势表示为

$$\psi_j(C_j) = \prod_{\phi: \alpha(\phi)=j} \phi \quad (6)$$

其中, $\alpha(\phi)=j$ 表示与团 C_j 有关的因子。由 C_i 传递给 C_j 的消息表示为 $\delta_{i \rightarrow j}$ 。每个团上的因子 $\beta_i(C_i)$ 称为团信任。团信任为初始团势与所有邻节点传递过来的消息的乘积,即

$$\beta_i(C_i) = \psi_i \cdot \prod_{k \in N_i^-(j)} \delta_{k \rightarrow i} \quad (7)$$

团信任即为团变量的边际概率分布:

$$\beta_i(C_i) = \sum_{x \sim C_i} \tilde{P}_\Phi(X) \quad (8)$$

两个团之间的边变量集是两个团相交变量集,称为分离集。分离集的因子 $\eta_{i,j}(S_{i,j})$ 称为分离集信任。消息传递完成之后,团树要满足全局一致性,即任意分离集信任满足

$$\eta_{i,j}(S_{i,j}) = \sum_{c_i \sim S_{i,j}} \beta_i(C_i) = \sum_{c_j \sim S_{i,j}} \beta_j(C_j) \quad (9)$$

也就是说,两个邻团关于分离集的边际概率分布是一致的。满足全局一致性的团树,也叫做校准团树,其信任为校准信任。校准团树不仅是一个可以存储概率推理结果的数据结

构,而且还提供了测度 $\tilde{P}_\Phi$ 的另一种参数化表示形式,校准团树算法的收敛点有

$$\tilde{P}_\Phi(X) = \frac{\prod_{i \in V_T} \beta_i(C_i)}{\prod_{(i,j) \in E_T} \eta_{i,j}(S_{i,j})} \quad (10)$$

团树算法收敛后,可以计算原贝叶斯网络中任意变量 X 的边际概率分布。利用包含变量 X 的团信任 $\beta_X(C_X)$,通过计算 $P(X) = \sum_{C_X \sim X} \beta_X(C_X)$ 可以得到 X 的边际概率分布。利用条件概率公式 $P(X=x|E=e) = \frac{P(X=x, E=e)}{P(E=e)}$ 以及相应

变量的边际概率分布,即可得到 $P(X|E)$ 。

3.3.1 消息传递机制

团树的消息传递机制主要有两种:和积消息传递与信任更新消息传递。

和积消息传递机制与VE算法的消元过程类似,从 C_i 传递到 C_j 的消息计算如下:

$$\delta_{i \rightarrow j} = \sum_{c_i \sim S_{i,j}} \psi_i \cdot \prod_{k \in (N_i^-(j))} \delta_{k \rightarrow i} \quad (11)$$

团 C_i 把从除了 C_j 外的其它邻团中传递过来的所有消息与其初始团势相乘,然后对除了分离集变量外的其他变量进行求和,最后把得到的因子作为消息传递给 C_j 。

和积消息传递中的消息 $\delta_{i \rightarrow j}$ 不接受来自团 C_j 的消息,而信任更新消息传递中的消息则包含来自团 C_j 的消息 $\delta_{j \rightarrow i}$,然后再利用除法将其除去。信任更新消息传递中的消息采用和积-除形式,为

$$\delta_{i \rightarrow j} = \frac{\sum_{c_i \sim S_{i,j}} \beta_i}{\delta_{j \rightarrow i}} \quad (12)$$

信任更新消息传递中并不直接计算两个团之间的消息,而是不断更新团信任,从而把消息融入到新的团信任中。

除了和积与信任更新消息传递外,近年来并行化消息传递受到了许多学者的关注,可有效处理高维概率图模型。Kozlov和Singh^[60]最早提出了团树推理的数据并行化实现,但该算法只适用于缓存一致且共享地址空间的物理上分布式计算机推理。文献[61]提出节点级的并行化消息传递机制,以便利用数据的并行化处理模式,但其只适用于带有较大团势的团树。文献[62]提出的一种利用图处理单元并行化的消息传递过程,能有效降低团树结构带来的计算复杂性。而文献[63]提出的分布式的消息传递机制,使得推理计算具有良好的可伸缩性。

除了并行化技术外,对于高维的概率图模型,学者们提出多种其它方法用于降低消息传递的复杂性。文献[64]提出动态量化策略,使得消息传递过程可自适应地增加或删减模型状态,从而实现模型的降维。文献[65]提出基于粒子滤波的随机方法,给出非参数的消息传递过程,其也能够处理复杂高维模型。文献[66]给出了具有低复杂性的随机团树算法,该算法可有效处理高维模型,并能保证算法的收敛性。

3.3.2 团树的构造

团树算法是基于团树结构来进行的,首先要把贝叶斯网络转换为团树结构,才能应用团树算法。团树的构造有两种途径,一是基于消元过程,二是基于带弦图。由于消元过程的中间因子对应于团树中的团,因此可以由消元过程诱导出团树。基于带弦图的团树构造方法是最常用的方法,即从包含团树表示的带弦图中提取团树结构。文献[67]最早提出由带

弦图构造团树的方法。文献[68]给出简化团树的概念,利用简化团树的性质可在线性时间内由任意带弦图构造消元机制。文献[69]提出枚举带弦图中最大团的最优序列的算法,此方法并不需要构造出一棵团树。文献[70]研究了带弦图的简化团树结构,并基于边收缩和边删减两种运算对简化团树进行分解,然后利用这种分解来归纳带弦图的结构特性。

4 变分近似推理

变分近似推理是近年来概率图模型推理的研究热点,它把查询问题表述为变分优化问题,利用变分优化的灵活框架,能方便地表示目标函数和添加各类约束条件,并且还大量求解优化问题的方法。变分近似推理在目标分布类 Q 中搜索概率分布 P_Φ 的最优近似分布 Q ,回答概率查询时在分布 Q 上进行概率推理,从而简化推理形式。

在变分近似推理中,目标函数为 P_Φ 和 Q 之间的距离,该距离用相关熵函数 $D(Q \| P_\Phi)$ 来度量,优化的目标就是使相关熵函数最小。分布 Q 可由校准信任集来表示,所以变分优化问题的寻优空间为信任集 $Q = \{\beta_i : i \in V_T\} \cup \{\eta_{i,j} : (i,j) \in E_T\}$,且满足全局一致性。为了简化优化问题的求解过程,把 $D(Q \| P_\Phi)$ 化简为

$$D(Q \| P_\Phi) = \ln Z - F[\tilde{P}_\Phi, Q] \quad (13)$$

其中能量泛函 $F[\tilde{P}_\Phi, Q]$ 为

$$F[\tilde{P}_\Phi, Q] = \sum_{\phi \in \Phi} E_Q[\ln \phi(U_\phi)] + H_Q(X) \quad (14)$$

因此,最小化相关熵函数 $D(Q \| P_\Phi)$ 等价于最大化能量泛函 $F[\tilde{P}_\Phi, Q]$ 。由于 $D(Q \| P_\Phi) \geq 0$, 因此 $\ln Z \geq F[\tilde{P}_\Phi, Q]$, 也就是说能量泛函是对数划分函数的下界。在有向图模型中,划分函数是已知证据的概率,是推理过程中最困难的部分。通过最大化能量泛函,可得到划分函数的近似值,从而求出联合概率分布 P_Φ 。

根据优化问题的目标函数、约束空间和寻优空间的近似类型的不同,变分近似推理主要分为 3 种算法: CGBP 算法,把能量泛函近似化,在带环聚类图中进行消息传递; EP 算法,把传递的消息近似化,求解优化问题时,最大化精确能量泛函,但是放松一致性约束条件; 结构变分算法,使用精确能量泛函,但只关注于具有特定简单分解形式的分布 Q 。除了这里介绍的 3 种变分近似推理算法外,还发展出了多种其他的变分近似推理方法,如凸松弛^[71]、对数行列式松弛^[72]、平面图分解^[73]和锥规划^[74]等。

CGBP 算法理论上可应用于任何网络,即使是具有高树宽的网络。目前 CGBP 算法已在多个领域中成功应用,如生物信息建模^[75]、立体匹配^[76]和图像分割^[77]等。EP 算法能够处理指数分布族表示的分布,但要求分布族能有效进行充分统计量数学期望的计算以及矩投影运算,所以一般用于低树宽的带弦图中。目前 EP 算法通常主要应用在混合网络的推理中,并衍生出许多 EP 算法的变型。结构变分算法是 3 种变分近似推理算法中唯一能保证得到对数划分函数的正确下限的算法,并能保证算法收敛,而前两种算法不能保证收敛性。

4.1 CGBP 算法

CGBP 算法^[17]在聚类图上进行消息传递。与团树不同,聚类图可能带环,并且聚类图只是原网络的一个近似结构。聚类图的校准定义比团树的校准定义要弱,因为分离集不再

是相邻聚类簇的交集,两个聚类簇只需满足分离集变量的边际概率分布一致。

CGBP 算法优化问题的目标函数为因子能量泛函

$$\tilde{F}[\tilde{P}_\Phi, Q] = \sum_{i \in V_T} E_{C_i \sim \beta_i} [\ln \psi_i] + \sum_{i \in V_T} H_{\beta_i}(C_i) - \sum_{(i,j) \in E_T} H_{\eta_{i,j}}(S_{i,j}) \quad (15)$$

其中, ψ_i 为团 C_i 的初始团势,而 $E_{C_i \sim \beta_i}[\cdot]$ 表示已知信任 β_i 时 C_i 取值的期望。 $\tilde{F}[\tilde{P}_\Phi, Q]$ 是能量泛函 $F[\tilde{P}_\Phi, Q]$ 的近似,因为 $\tilde{F}[\tilde{P}_\Phi, Q]$ 中的信任只是原边际概率分布的近似。因此, CGBP 算法的优化问题为

$$\begin{aligned} & \text{Find } Q \\ & \max \tilde{F}[\tilde{P}_\Phi, Q] \\ & \text{s. t. } \eta_{i,j}(s_{i,j}) = \sum_{c_i \sim s_{i,j}} \beta_i(c_i), \forall (i,j) \in E_U, \forall s_{i,j} \in R(S_{i,j}) \\ & \sum_{c_i} \beta_i(c_i) = 1, \forall i \in V_U \\ & \beta_i(c_i) \geq 0, \forall i \in V_U, \forall c_i \in R(C_i) \end{aligned}$$

此优化问题包含了两方面的近似: 1) 使用近似能量泛函; 2) 寻优空间为伪边际分布。虽然这里的能量泛函带有近似性,但使其最大化仍然可以得到对数划分函数的较好近似值。可通过拉格朗日乘子法求解该约束优化问题,得到不动点方程

$$\delta_{i \rightarrow j} \propto \sum_{c_i \sim s_{i,j}} \psi_i \left(\prod_{k \in N_i - \{j\}} \delta_{k \rightarrow i} \right) \quad (16)$$

并且有

$$\beta_i \propto \psi_i(c_i) \left(\prod_{j \in N_i} \delta_{j \rightarrow i} \right) \quad (17)$$

$$\eta_{i,j} = \delta_{i \rightarrow j} \cdot \delta_{j \rightarrow i} \quad (18)$$

通过迭代过程可求解所有不动点方程,进而求出最大值 Q 。但是该优化问题在收敛性方面存在问题^[78]: (1) 函数的极大值不一定是不动点方程收敛后得到的不动点。如果函数的极大值为边界点,那它可能不是不动点。(2) 极大值不是信任传播算法的唯一不动点,极小值和鞍点也是不动点^[79]。但 BP 的稳定收敛点通常是函数的局部极大值。(3) 通过信任传播过程不一定能得到极大值。不能保证信任传播的每个消息传递步骤都使能量泛函增大,在实际应用中不一定能得到算法的不动点。文献[80]给出了最小费用流网络中的 CGBP 算法的收敛性与正确性分析,但是,如何在理论上保证任意网络中的 CGBP 算法的收敛性以及解的最优性和唯一性是未解决的问题。

在对一个网络进行 CGBP 推理之前,必须先构造其聚类图。不同的聚类图会导致差异很大的结果,而且具有快速传播过程的聚类图通常产生较差的近似值,所以在构造的时候要同时考虑计算成本和查询结果准确性。目前,聚类图的构建主要有两种方法即转换两两关系 MN 为聚类图或者构造 Bethe 聚类图。很多学者提出了大量方法来改进聚类图的构造过程,以解决 CGBP 算法的收敛性问题,如对能量泛函进行凸化^[81]或者给出更强的收敛性条件^[82,83]等。文献[84]提出为原网络构造区域图,而不是聚类图,能更好地表示变量间的相互作用,提高近似效果。

4.2 EP 算法

EP 算法^[18]在聚类结构中使用近似的消息来简化传播过程,既保持了聚类的正确性又获得了高效的计算速度。当使用指数分布族来描述近似消息期望传播过程^[61]时,所有消息都位于同一个指数分布族中,只需参数来描述这些消息。

因此,传播过程是这些参数的更新过程,期望传播的参数更新法则为

$$\tilde{\sigma}_{i \rightarrow j} \leftarrow (\tilde{\sigma}_{i \rightarrow j} - \tilde{\sigma}_{j \rightarrow i}) \quad (19)$$

即消息 $\tilde{\sigma}_{i \rightarrow j}$ 的参数为近似团边际分布 $\tilde{\sigma}_{i \rightarrow j}$ 和之前的近似消息 $\tilde{\sigma}_{j \rightarrow i}$ 的参数之差。该推理过程的主要问题是计算 $\tilde{\sigma}_{i \rightarrow j}$ 的指数分布族,其计算分两步:首先根据团信任 $\tilde{\beta}_i$ 计算充分统计量 $\tau_{i,j}$ 的数学期望;然后,在 $Q_{i,j}$ 中寻找与统计量数学期望相同分布的参数。

EP 算法的优化问题为

Find Q

$$\max F[\tilde{P}_\Phi, Q]$$

$$\text{s. t. } E_{S_{i,j} \sim \eta_{i,j}}[\tau_{i,j}] = E_{S_{i,j} \sim \beta_j}[\tau_{i,j}], \forall (i,j) \in E_T$$

$$\sum_{c_i} \beta_i(c_i) = 1, \forall i \in V_T$$

$$\sum_{s_{i,j}} \eta_{i,j}[s_{i,j}] = 1, \forall (i,j) \in E_T$$

$$\beta_i(c_i) \geq 0, \forall i \in V_T, c_i \in R(C_i)$$

该优化问题把一致性约束替换为期望一致性约束,使两个相邻聚类簇并不满足全局一致性条件,因此目标函数和寻优空间都是近似的。其不动点方程为

$$\tilde{\sigma}_{i \rightarrow j} = \frac{M - d_{i,j}(\beta_j)}{\delta_{i \rightarrow j}} \quad (20)$$

其中,符号 $M - d_{i,j}$ 表示把 β_i 中与分离集 $S_{i,j}$ 不相同的变量进行边际化并投影至 $Q_{i,j}$ 中。团信任和分离集信任的计算公式如(17)和式(18)所示。

EP 传播算法也存在一些缺陷:EP 消息传递的迭代过程不能保证使目标函数单调增加,算法不一定收敛。另外,即使算法收敛了,由于聚类簇是近似校准的,当使用不同的聚类簇来回答查询问题时,会产生不同的结果。Seeger 等人^[85]把 EP 思想与协方差去耦方法结合起来,提出一种收敛速度明显加快的 EP 算法。

4.3 结构变分近似算法

结构变分近似算法^[20]考虑近似分布族 Q 的分解结构,根据分解结构的不同,可分为平均场近似和结构近似。平均场近似是最基本的结构变分近似,但是它要求 Q 中的所有变量都两两独立,近似效果并不好。结构近似对平均场近似进行改进,使得近似结构更加灵活,也能发现原分布 P_Φ 中的更多依赖性。

4.3.1 平均场近似

平均场近似采用全分解思想,假设 Q 中的所有变量两两独立,分布 Q 的结构表示为独立边际分布的乘积:

$$Q(X) = \prod_i Q(X_i) \quad (21)$$

虽然把 P_Φ 近似为完全因式分解的分布会丢失原分布的大量信息,但是当回答 Q 上的任意查询时,只需把查询问题的变量的相关分布相乘即可,计算十分便利。平均场近似的变分优化问题为:

$$\text{Find } \{Q(X_i)\}$$

$$\max F[\tilde{P}_\Phi, Q]$$

$$\text{s. t. } Q(X) = \prod_i Q(X_i)$$

$$\sum_{x_i} Q(x_i) = 1, \forall i$$

与 CGBP 和 EP 算法不同,平均场近似优化问题^[86]的目标函数是精确的,只是把寻优空间近似为不包含原始分布 P_Φ 信息的分布空间 Q 。求解该优化问题可得不动点方程

$$Q(x_i) = \frac{1}{Z_i} \exp\left\{\sum_{\phi: X_i \in D(\phi)} E_{U_\phi \sim Q}[\ln \phi(U_\phi, x_i)]\right\} \quad (22)$$

即 $Q(X_i)$ 要与它所在的团势的期望值保持一致。计算每个取值 x_i 的指数项并相加,然后归一化赋给 $Q(X_i)$,即可在一步内得到 $Q(X_i)$ 的最优值。平均场近似算法的每次迭代都能对目标分布 P_Φ 产生更好的近似 Q ,收敛性得到保证。但是,平均场近似中要求近似分布 Q 完全分解,当概率图模型的变量的相关性较强时,该算法的推理精度很低。扩展平均场算法^[87]对平均场算法进行改善,即将近似分布分解为不相交聚类簇上的势函数的乘积,保留了原模型的部分结构,能更好地利用目标分布中变量间的依赖性。

4.3.2 结构近似

结构近似算法是平均场近似方法的一般表示形式,近似分布可分解为任意聚类簇上的势函数的乘积:

$$Q(X) = \frac{1}{Z_Q} \prod_{j=1}^J \psi_j \quad (23)$$

其中, ψ_j 是定义域为 C_j 的势函数。根据 Q 的结构,能量泛函可改写为

$$F[\tilde{P}_\Phi, Q] = \sum_{k=1}^K E_Q[\ln \phi_k] - \sum_{j=1}^J E_Q[\ln \psi_j] + \ln Z_Q \quad (24)$$

结构近似推理算法的优化问题为

$$\text{Find } \{\psi_j\}$$

$$\max F[\tilde{P}_\Phi, Q]$$

$$\text{s. t. } Q(X) = \frac{1}{Z_Q} \prod_{j=1}^J \psi_j, \sum_j \psi_j = 1, \forall j$$

该优化问题也使用了精确的目标函数和近似的寻优空间。求解该优化问题可得到不动点方程

$$\psi_j(c_j) \propto \exp\{E_Q[\ln \tilde{P}_\Phi | c_j] - \sum_{k \neq j} E_Q[\ln \psi_k | c_j]\} \quad (25)$$

在实际应用中,可以利用近似分布类 Q 和原始分布 P_Φ 的结构进一步简化不动点方程,使得更新过程更加高效。结构近似算法在一般近似分布集上求解泛函极值问题,提高了推理精度,但增加了计算复杂度;而且近似分布的分解结构的选择是极具挑战性的。显然,依靠专家经验知识的人工结构选择并不适用于复杂网络。文献^[87]根据模型权重来选择分解结构,给出扩展平均场方法的自动结构选择算法。但目前平均场推理方法中的分解结构选择通常只关注原网络的结构特征,而忽略了变量间的相互作用。

还有学者把平均场推理方法与其他相关理论相结合,从原理上对平均场推理算法进行改进,如线性响应理论^[88]和 TAP 方法^[89]。但这些高阶推理方法不能给出对数划分函数的误差下界,无法保证算法误差。

5 抽样近似推理

抽样近似推理过程把联合概率分布近似为所有或部分变量的实例集,这些实例通常称为粒子。抽样近似推理的主要问题是给定网络的联合概率分布 $P(X)$,估计某个事件 $Y=y$ 的概率,其中 $Y \subseteq X, y \in R(Y)$ 。或者估计 P 的密度函数 $f(X)$ 的数学期望,通常选择指示函数 $f(\xi) = 1\{\xi(Y) = y\}$,其中 $\xi(Y)$ 表示 ξ 中在变量 Y 上的取值,通过产生 M 个粒子,估计与每个粒子有关的函数值或其数学期望,然后累加这些结果得到函数数学期望的估计值。

抽样近似推理是唯一可应用于任意概率图模型的推理算法,包括混合网络和暂态网络,并能保证在样本数足够大时得

到正确估计值。虽然抽样近似推理能给出渐进性保证,但其所需样本数很难预测,在实际复杂概率图模型的应用中通常得不到正确估计值。

根据粒子产生过程的不同以及粒子概念定义的不同,抽样近似推理算法包括前向抽样算法、重要性抽样算法、MC-MC 算法、压缩粒子算法和确定性搜索算法。其中前向抽样和重要性抽样算法只适用于贝叶斯网络。前向抽样从原始 BN 的概率分布中抽样,是最基本的抽样算法,但难以得到条件概率分布估计值。重要性抽样过程则从接近于 BN 真实后验分布的近似分布中抽样,可实现条件概率分布的计算。MCMC 算法构造平稳分布为真实后验分布的马尔可夫链,从平稳分布中抽样,样本准确性更好,但当马尔可夫链的混合速度非常慢时,算法运行时间过长,MCMC 算法无法在实际中使用。压缩粒子算法对网络中的部分变量进行抽样,其余变量采用解析表达式表示概率分布,可有效降低抽样算法复杂性,并可在有解析表达式的其余变量的子网络上进行简单的精确推理,估计值准确性得到提高。确定性搜索以非随机规则产生能覆盖概率分布主要部分的粒子,舍弃小概率粒子,可以用较少的粒子数获得比较准确的估计值。

5.1 前向抽样算法

前向抽样算法是贝叶斯网络中最基本的抽样算法,从联合概率分布 $P(\mathbf{X})$ 中产生 M 个随机样本 $\xi[1], \dots, \xi[M]$ 。前向抽样的过程很简单,以与 BN 局部路径顺序一致的某个顺序对节点抽样,即抽样某个节点时已知其所有父节点的值,然后从该节点的 CPD 中抽样。

产生粒子集 $D = \{\xi[1], \dots, \xi[M]\}$ 后,可以估计任意函数 f 的期望

$$\hat{E}_D(f) = \frac{1}{M} \sum_{m=1}^M f(\xi[m]) \quad (26)$$

当计算 $P(y)$ 时,该估计值就是事件 y 所出现的粒子数与 M 的比值:

$$\hat{P}_D(y) = \frac{1}{M} \sum_{m=1}^M 1\{y[m]=y\} \quad (27)$$

其中, $y[M]$ 表示 $\xi[m](Y)$, 即粒子 $\xi[m]$ 中 Y 的取值。估计值的准确性依赖于产生的粒子数,可用 Hoeffding 界或 Chernoff 界来确定一定误差界下所需抽样样本个数。

BN 的概率查询主要是计算条件概率,但前向抽样难以处理该任务。目前有两种解决办法:(1)通过拒绝抽样算法^[90]实现从后验概率分布 $P(\chi|e)$ 中产生样本:先从 $P(\mathbf{X})$ 中产生样本 $\mathbf{x}$,然后拒绝与证据 e 不相容的样本,剩下的样本相当于是从后验分布 $P(\mathbf{X}|e)$ 中抽样得到的;(2)分别估计 $P(e)$ 和 $P(y, e)$,然后计算两者的比值,即可得到 $P(\chi|e)$ 的估计值。

5.2 重要性抽样算法

重要性抽样算法^[21]主要估计概率分布 $P(\mathbf{X})$ 的密度函数 $f(\mathbf{X})$ 的数学期望值, $P(\mathbf{X})$ 称为目标函数。当 P 为后验分布时,由于真实后验分布未知而无法产生粒子,因此希望在 P 的某个近似分布 Q 中产生样本, Q 称为建议分布或抽样分布。重要性抽样在建议分布 Q 中采集样本并通过样本的加权平均来近似 $f(\mathbf{X})$ 的数学期望。

重要性抽样分为非归一化重要性抽样和归一化重要性抽样。因为样本从建议分布 Q 中产生,所以需要调整估计值以补偿分布不一致产生的不正确抽样带来的误差。从 Q 中产

生样本集 $D = \{\mathbf{x}[1], \dots, \mathbf{x}[M]\}$ 后,非归一化重要性抽样的估计值为

$$\hat{E}_D[f] = \frac{1}{M} \sum_{m=1}^M f(\mathbf{x}[m]) \frac{P(\mathbf{x}[m])}{Q(\mathbf{x}[m])} \quad (28)$$

其中,因子 $P(\mathbf{x}[m])/Q(\mathbf{x}[m])$ 是 $f(\mathbf{x}[m])$ 项的修正权值,用以补偿把 Q 当作目标分布引起的误差。可以证明,此估计值为无偏估计值,估计值的方差随 Q 与 P 的差值的增大而增加。

令 $\omega(\mathbf{x})$ 表示 $P(\mathbf{x})/Q(\mathbf{x})$,非归一化重要性抽样算法假设 P 已知,然而实际上归一化常量 Z 是未知的, P 也就未知。因此把权值 $\omega(\mathbf{x})$ 修改为

$$\omega(\mathbf{X}) = \frac{\tilde{P}(\mathbf{X})}{Q(\mathbf{X})} \quad (29)$$

所以从 Q 中产生样本集 $D = \{\mathbf{x}[1], \dots, \mathbf{x}[m]\}$ 后,归一化重要性抽样的估计值为

$$\hat{E}_D(f) = \frac{\sum_{m=1}^M f(\mathbf{x}[m]) \omega(\mathbf{x}[m])}{\sum_{m=1}^M \omega(\mathbf{x}[m])} \quad (30)$$

可以证明,该估计值不是无偏的。

BN 中的重要性抽样推理需要先构造其残差网络 $B_{Z=y}$, 即已知观测变量 $Z \in \mathbf{Z}$, 使其 CPD 中的观测变量取预先观测的值。计算非条件概率时,建议分布 Q 可选择为残差网络 $B_{Z=y}$ 的概率分布,那么非条件概率的估计值可以通过非归一化或归一化重要性抽样计算得到。计算条件概率时,可以使用比值似然加权(Likelihood Weighting, LW)算法或归一化似然加权算法。比值 LW 可以估计单个特定事件 y 的条件概率 $P(y|e)$, 对 $P(y, e)/P(e)$ 的分子和分母分别使用未归一化重要性抽样算法,用 $P(y, e)/P(e)$ 来计算条件概率 $P(y|e)$ 的值。而归一化 LW 可以估计某个变量集 Y 的条件概率分布 $P(Y|e)$, 其目标分布为 $P(\mathbf{X}) = P_B(\mathbf{X}|e)$, 但并不直接从 $P(\mathbf{X}) = P_B(\mathbf{X}|e)$ 中抽样,而是对 $P_B(\mathbf{X}, e)$ 使用归一化重要性抽样估计 $\tilde{P}(\mathbf{X}) = P_B(\mathbf{X}, e)$, 从而计算出条件概率分布 $P(\mathbf{X}) = P_B(\mathbf{X}|e)$ 的估计值。

由残差网络得到的建议分布并不是最优建议分布,还可以使用其他的建议分布,如反向重要性抽样算法使用子节点的似然概率来对证据变量的父节点进行抽样;也可以在已知样本组合或某些父节点或子节点的观察值时,对变量值抽样;还可以考虑混合方法,利用某种全局近似推理算法(如变分近似推理)构造出建议分布。

为了解决概率图模型中重要性抽样过程产生大量零权值样本的问题,文献[91]提出抽样搜索方法,把系统回溯搜索与蒙特卡洛抽样结合起来,很好地解决了零权值样本的拒绝问题。而文献[92]提出了 AND/OR 重要性抽样,可有效降低样本均值的方差。目前,重要性抽样在贝叶斯网络中的应用有多种变型,如自重要性抽样(Self Importance Sampling)、启发式重要性抽样、自适应重要性抽样^[93]、动态重要性抽样^[94]和证据预传播重要性抽样算法^[95,96]。

5.3 MCMC 算法

MCMC 算法^[22]的基本思想是根据原概率图模型构造马尔可夫链,通过迭代和混合使马尔可夫链收敛至平稳分布,其平稳分布十分接近于原概率图模型的真实后验分布,然后从平稳分布中抽样计算估计值。

马尔可夫链是通过状态空间 $R(\mathbf{X})$ 和转换模型来定义

的,马尔可夫链的状态节点为变量 $\mathbf{X}$ 的可能取值,转换模型为从状态 $\mathbf{x}$ 到 $\mathbf{x}'$ 的概率 $T(\mathbf{x} \rightarrow \mathbf{x}')$ 。概率图模型的推理中只使用具有唯一平稳分布的马尔可夫链。要保证有唯一平稳分布,马尔可夫链需要满足各态遍历性,这等价于马尔可夫链满足正则性,即存在某个数 k ,使得对所有状态 $\mathbf{x}, \mathbf{x}' \in R(\mathbf{X})$,在 k 步时从 $\mathbf{x}$ 到 $\mathbf{x}'$ 的概率大于 0。而保证正则性的充分非必要条件有两个:(1)在状态图中存在从任意状态到任意其它状态的正概率路径;(2)对每个状态 $\mathbf{x}$,从 $\mathbf{x}$ 到 $\mathbf{x}$ 的自回路转换概率为正。

在进行 MCMC 推理前,必须先根据原概率图模型构造出满足正则性与具有唯一平稳分布的马尔可夫链。吉布斯抽样可构造出在特定假设下的马尔可夫链,但其限制很多。而 Metropolis-Hastings(MH)算法^[97,98]能构造出可逆马尔可夫链,吉布斯抽样只是其一个特例。最近,文献[99]提出的自适应独立 MH 方法可快速可靠地实现马尔可夫链的构造。而文献[100]综述了 MH 方法的计算复杂性的相关研究结果。

采集样本前,需要让马尔可夫链迭代混合收敛至平稳分布 π ,然后再从 π 中采集样本。该过程的主要问题是如何确定混合时间^[101],即采集第一个样本时所经过的步数。确定马尔可夫链混合时间的紧致界的方法有多种;可以基于马尔可夫链中状态之间的转换矩阵的分析,但是需要对指数数量的状态进行详细穷举,计算量大;更常用的方法是使用各类启发式方法估计轨迹的混合程度。

令 $\mathbf{x}^{(0)}, \dots, \mathbf{x}^{(t)}$ 表示混合阶段,混合完成后从平稳分布中采集 M 个样本 $D = \{\mathbf{x}[1], \dots, \mathbf{x}[M]\}$,对 $m = 1, \dots, M$ 有 $\mathbf{x}[m] = \mathbf{x}^{(T+m)}$ 。若 $\mathbf{x}^{(T+1)}$ 从 π 中抽样得到,则 D 中所有样本都是从 π 中抽样得到,概率分布估计值为

$$\hat{E}_D(f) = \frac{1}{M} \sum_{m=1}^M f(\mathbf{x}[m], \mathbf{e}) \quad (31)$$

该期望为 $E_{\pi(\mathbf{x})}[f(\mathbf{X}, \mathbf{e})]$ 的无偏估计值。

但是,由实际概率图模型推导出来的马尔可夫链通常具有多模态的后验分布,各模态之间的混合速度非常慢。为了加快收敛速度,一般根据后验分布的性质,引入适合该状态空间的步长更宽的转换模型,使得链可在模态之间更快速地移动。除了令步长更宽外,还有多种方法可以加快收敛速度,如块吉布斯法(Block Gibbs Approach)和模拟退火法(Simulated Annealing)等。

5.4 压缩粒子算法

压缩粒子^[23]就是对网络变量的某个子集进行局部赋值,而其它变量采用解析表达式表示概率分布。把变量集 $\mathbf{X}$ 分为两个子集: $\mathbf{X}_p$,其取值定义粒子; $\mathbf{X}_d$,其概率分布采用解析表达式表示。那么压缩粒子由实例 $\mathbf{x}_p \in R(\mathbf{X})$ 和分布 $P(\mathbf{X}_d | \mathbf{x}_p, \mathbf{e})$ 的某个表达式组成。后验分布 $P(\mathbf{X}_p, \mathbf{X}_d | \mathbf{e})$ 的密度函数 $f(\xi)$ 的期望估计值为:

$$\begin{aligned} E_{P(\xi|\mathbf{e})}[f(\xi)] &= \sum_{\mathbf{x}_p, \mathbf{x}_d} P(\mathbf{x}_p, \mathbf{x}_d | \mathbf{e}) f(\mathbf{x}_p, \mathbf{x}_d, \mathbf{e}) \\ &= \sum_{\mathbf{x}_p} P(\mathbf{x}_p | \mathbf{e}) \sum_{\mathbf{x}_d} P(\mathbf{x}_d | \mathbf{x}_p, \mathbf{e}) f(\mathbf{x}_p, \mathbf{x}_d, \mathbf{e}) \\ &= \sum_{\mathbf{x}_p} P(\mathbf{x}_p | \mathbf{e}) (E_{P(\mathbf{X}_d | \mathbf{x}_p, \mathbf{e})}[f(\mathbf{x}_p, \mathbf{X}_d, \mathbf{e})]) \end{aligned} \quad (32)$$

可以把分布 $P(\mathbf{X}_d | \mathbf{X}_p, \mathbf{e})$ 显式表达为一个概率图模型,使用混合方法来进行推理,即先从 $\mathbf{X}_p$ 中产生样本 $\mathbf{x}_p$,然后在已知 $\mathbf{x}_p$ 的情况下对 $\mathbf{X}_d$ 进行精确推理。重要性抽样^[102]和

MCMC 算法^[103]都可扩展到压缩粒子的情况。文献[104]提出的一种新的贝叶斯网络压缩抽样算法,能加速收敛过程。

5.5 确定性搜索算法

确定性搜索算法^[24]对高概率发生状态进行显式搜索,以非随机规则产生某个样本集 $D = \{\xi[1], \dots, \xi[M]\}$,然后只利用这些样本来近似联合概率分布 P ,而忽略其它的样本。如果这些样本覆盖联合概率分布的主要部分,则可以得到联合概率分布的合理近似表示, $\hat{P}(\mathbf{y})$ 的估计值为

$$\sum_{m=1}^M 1\{\mathbf{y}[m] = \mathbf{y}\} \hat{P}(\xi[m]) \quad (33)$$

这里通过概率对粒子进行加权。由于采样过程有可能不产生 $\xi(\mathbf{Z}) = \mathbf{y}$ 的某些实例,估计值会有偏的。因此,一般使用粒子为事件 $\mathbf{Z}$ 的未归一化概率来提供上下限估计:

$$\begin{aligned} \sum_{m=1}^M 1\{\mathbf{y}[m] = \mathbf{y}\} \hat{P}(\xi[m]) \\ \leq \hat{P}(\mathbf{Z} = \mathbf{y}) \\ \leq (1 - \sum_{m=1}^M 1\{\mathbf{y}[m] \neq \mathbf{y}\} \hat{P}(\xi[m])) \end{aligned} \quad (34)$$

同样,可以给出概率查询 $P(\mathbf{y} | \mathbf{e}) = \hat{P}(\mathbf{y}, \mathbf{e}) / \hat{P}(\mathbf{e})$ 的上下限,先分别获得分子和分母的上下限,分子的下限和分母的上限共同给出了其比值的下限估计,而分子的上限和分母的下限共同给出了其比值的上限估计。

6 MAP 推理

MAP 查询就是求非证据变量集的最有可能取值。边际 MAP 查询就是求某个变量子集的最有可能取值,而对其它变量进行边际化运算。处理 MAP 查询需要把联合概率分布写为因子的乘积,且需考虑由因子集 Φ 定义分布 $P_\Phi(\mathbf{X})$ 和未归一化测度 $\tilde{P}_\Phi$ 。MAP 查询需要计算

$$\begin{aligned} \xi^{map} &= \arg \max_{\xi} P_\Phi(\xi) = \arg \max_{\xi} \frac{1}{Z} \tilde{P}_\Phi(\xi) \\ &= \arg \max_{\xi} \tilde{P}_\Phi(\xi) \end{aligned} \quad (35)$$

因此,MAP 推理包含最大化运算和乘积运算。由最大化运算得到的边际分布称为最大化-边际分布。

边际 MAP 问题类似于 MAP 任务,需要求出边际概率最大的取值。把 $\mathbf{X}$ 分为两个不相交子集, $\mathbf{X} = \mathbf{Y} \cup \mathbf{W}$, 边际 MAP 查询可表示为

$$\mathbf{y}^{map} = \arg \max_{\mathbf{y}} P_\Phi(\mathbf{y}) = \arg \max_{\mathbf{y}} \sum_{\mathbf{W}} \tilde{P}_\Phi(\mathbf{y}, \mathbf{W}) \quad (36)$$

因此,边际 MAP 问题同时包含乘积、求和与最大化运算,这使得该任务无论在理论上还是在实践中都很困难。精确推理算法难以处理边际 MAP,而 MAP 查询的许多近似方法也不能简单推广到边际 MAP 情况中。边际 MAP 任务的最有效近似方法是对取值 $\mathbf{y}$ 使用启发式搜索,同时在内回路上的 $\mathbf{W}$ 上进行某种(精确或近似)和积推理。

概率查询的推理算法通常都可推广到 MAP 推理情况中。目前,MAP 推理比较常用的算法主要有以下 6 种:最大化-乘积 VE 算法、最大化-乘积团树算法、最大化-乘积 CGBP 算法、线性优化算法、图割算法和局部搜索算法。与概率查询的推理算法类似,最大化-乘积 VE 算法、最大化-乘积团树算法、最大化-乘积 CGBP 算法是基于消息传递的算法。线性优化算法是目前 MAP 推理的热门研究方向,把 MAP 问题表述为线性规划问题,求解其对偶问题即可得近似解。图割算法主要应用在马尔可夫随机场中,在特定情况下,运算速度比其他算法要快。局部搜索算法可以有效地解决边际 MAP 问题。

6.1 最大化-乘积 VE 算法、最大化-乘积团树算法、最大化-乘积 CGBP 算法

最大化-乘积 VE 算法、最大化-乘积团树算法与最大化-乘积 CGBP 算法^[25]分别是前面介绍的 VE 算法、团树算法和 CGBP 算法的简单推广,把求和运算替换为最大化运算,并引入回溯过程来计算 MAP 取值。这 3 种算法大致可分为两个阶段:(1)使用消元过程或者团树和聚类图中的消息传递过程,计算 $\mathbf{X}$ 中所有变量的精确或近似最大化-边际分布集;(2)利用最大化-边际分布集构造所有变量的最有可能联合取值,该过程称为解码。最大化-边际分布可以求解单个变量的局部优化问题组成的全局优化问题,构造联合取值来局部优化每个信任。如果可以构造出这样一个取值,可以保证其强局部最优性甚至全局最优性。

最大化-乘积 VE 算法只返回最后一个消元变量的最大化-边际分布,因此只能确定该变量的取值,然后利用回溯过程来逐一选择其它变量的取值。而最大化-乘积团树算法则返回所有变量的最大化-边际分布,当校准节点信任之间不存在联系时,可以分别局部优化每个变量的取值来求出唯一 MAP 取值。最大化-乘积 CGBP 通常把最大化-乘积校准聚类图转换成离散取值约束满足问题(Constraint Satisfaction Problem, CSP),运行某种 CSP 算法^[105]得到局部最优取值。文献[106]把 MAP 查询转化为成加权 CSP,使得可以采用目前最先进的加权 CSP 算法来求解 MAP 查询问题,并提出了两两关系分解技术来简化网络。

6.2 线性优化算法

基本 MAP 问题可作为整数值变量集上的整数线性规划问题^[27],其中目标函数和约束空间都是线性的。求解整数线性规划是 NP 难的,所以一般采用线性规划松弛,对整数约束条件进行松弛,可在多项式时间内求解,并且存在大量有效的可求解该问题的算法。当线性规划松弛的界为紧时,优化该线性规划(Linear Programming, LP)问题即可回答 MAP 查询。当线性规划松弛的界越紧时,优化结果的精度越高,有很多研究工作致力于收紧线性规划松弛的界^[107,108]。MAP 查询的线性规划的优化算法可以使用优化理论中的大量算法,但目前通常使用能充分利用图结构的方法来优化线性规划的对偶问题,如基于块坐标下降的消息传递算法^[109-111]、基于投影次梯度上升的算法^[112,113]和基于乘子交替方向的算法^[114]。

6.3 图割算法

特定二值 MRF 问题的 MAP 问题可利用简单有效的图割算法^[28,115]来求解最优问题。不管底层图模型的结构复杂性如何,该算法总能保证在多项式时间内返回最优解。图割算法是一种图构造算法,把 MRF 的节点集割成两个不相交的节点子集,分别包含源节点和目标节点,同时使切割成本最小,该过程称为最小割(min-cut)。通过构造满足子模性质的图,即图中的两两关系能量函数 $\epsilon_{i,j}(\cdot, \cdot)$ 满足 $\epsilon_{i,j}(0,0) + \epsilon_{i,j}(1,1) \leq \epsilon_{i,j}(1,0) + \epsilon_{i,j}(0,1)$,然后运行最小割算法,可以使得一个割对应于 $\mathbf{X}$ 的一个取值,而割的成本对应于取值的值,从而得到最小成本取值。文献[116]总结了最小化带有非子模项的二值能量函数时通常使用的图割方法。而文献[117]提出一种适用于高阶 MRF 的图割算法,每一步都同时考虑能量函数中的多个项,在理论和实验中都获得了很好的效果。

当变量为非二值变量时,不能再使用图构造法来求解最优值。一种有效的方法是采用贪婪爬山算法,其每一步都包含简化问题的全局最优解。该算法有 3 种常用的步长寻优过程:alpha 扩充、alpha-beta 交替和大范围移动^[118]。这 3 种步长寻优过程只是搜索空间大小的最优移动不相同,但都可以使用最小割算法作为子程序来求解 MAP 问题。

6.4 局部搜索算法

局部搜索算法^[30]就是在取值空间上搜索变量集的高概率取值。处理这种任务的方法可分两类:(1)系统方法^[119],搜索取值空间来保证未被考虑的取值不是最优的,从而保证最优解。这类方法通常在局部取值空间上搜索,初始化取值为空,然后每次更新一个变量,如分支定界法(branch-and-bound)。(2)非系统方法,在全取值空间上搜索,对取值进行局部修改来提高其变量集的高概率取值分数,不能保证最优性,如局部爬山算法、禁忌搜索(Tabu Search)算法、定向搜索(Beam Search)算法和带有随机重启的贪婪爬山算法等。

局部搜索算法可以很好地解决边际 MAP 问题,但是算法成本很高,利用动态规划可以大大提高算法的性能,如运行一次团树传播来同时计算出搜索中的所有算子的分数,这在团树中相当于没有约束消元顺序的算法。

7 混合网络中的推理

混合网络是同时包含离散变量和连续变量的网络,其推理具有很大的挑战性,如连续变量的因子没有通用的表示形式,必须为每个 CPD 或初始因子选择一个参数分布族。即使为每个初始因子都选取相同的参数分布族,因子乘积或因子边际化后的因子也有可能不属于此参数分布族。另外,边际化步骤需要积分运算而不是求和运算,并不是所有函数都是可积的,即使函数的积分定义良好,但它也可能没有解析表达式的积分,需要使用近似的数值积分法。

下面先讨论变量全部为连续变量的高斯网络的推理,然后分析同时包含离散变量和连续变量的混合网络的推理算法,再介绍处理非线性依赖的方法,最后给出非参数的抽样算法。这些算法都是专门为图模型而设计的方法,除此之外还可使用统计推理中的很多方法,因为连续密度函数的运算是统计推理的主要部分。例如,使用连续密度函数的其它表示形式的消息传递方法或者使用其他的数值积分法。

7.1 高斯网络的推理

高斯网络的所有变量都是连续变量,假设所有局部因子都只编码变量线性依赖性,CPD 表示为线性高斯分布。高斯网络的推理一般采用基于消息传递的方法,如 VE 算法、团树算法或 CGBP 算法。

前面介绍的 VE 算法与团树算法都可应用在线性高斯网络中,只是因子的表示形式以及因子基本运算的实现稍有改变。由于线性高斯 CPD 并不是高斯分布,而是条件分布,因此算法执行过程中的因子采用正规型(Canonical Form)表示,即表示为对数二次型 $\exp(Q(\mathbf{x}))$,其中 $Q(\mathbf{x})$ 为某个二次函数。正规型既能表示高斯分布又能表示线性高斯分布。表示为正规型的所有因子运算都可在多项式时间内完成。

CGBP 算法中高斯分布采用正规型表示,算法过程与前面介绍的 CGBP 算法类似,只是因子乘积与边际化运算采用正规型表示。当算法收敛时,生成的信任编码了联合分布的

正确均值,但估计方差比真实方差低^[120]。文献[121]提出一种松弛高斯 BP 算法,其可有效降低高斯网络 BP 算法的计算复杂性和所需存储空间。

7.2 混合网络的推理

混合网络的推理^[122]主要讨论比较成熟的条件线性高斯(Conditional Linear Gaussian, CLG)网络推理,该网络中不存在父节点为连续变量的离散子节点,所有连续变量的局部概率模型都是条件线性高斯 CPD。CLG 网络的推理主要使用 EP 算法,推理过程中的因子使用正规表(Canonical Table)来表示,融合了连续变量的正规型与离散变量的表格形式。因子的乘积运算为混合正规表,也就是高斯混合分布(Mixture of Gaussians, MoG)^[33]。而因子的连续边际化运算很简单,利用连续变量的正规型很容易实现,但是其离散边际化运算需要定义为弱边际化运算,利用 M-投影把高斯混合分布近似为单个高斯分布。

CLG 网络的推理也可以使用接近精确的团树算法来求解,通过强三角划分构造具有根节点的有根树,在此有根树上运行团树算法。但这种方法在实际中通常不可行,产生的有根树通常太大。文献[123]提出的符号消元(Symbolic Variable Elimination)算法可实现混合网络的精确推理。

7.3 处理非线性依赖

前面把 CPD 表示为线性高斯分布,限制变量之间的依赖关系是线性的。但在实际应用中,变量之间通常存在非线性依赖,例如离散变量具有连续父节点。此时,不能再使用 MoG 分布表示,而是采用截断指数混合(Mixtures of Truncated Exponentials, MTE)分布^[34, 124]。消息传递过程的中间因子需要使用高斯分布族来表示。为了进行消息的传递,需要把聚类簇团树的因子乘积近似为高斯分布,该过程称为线性化运算。然后把生成的高斯分布边际化投影到分离集 $S_{i,j}$ 上,产生近似聚类簇边际分布 $\tilde{\sigma}_{i,j}$,这实际上是 EP 算法的执行过程。

线性化运算主要有两种方式:泰勒级数展开线性化和数值积分矩投影线性化。其中,数值积分矩投影线性化又可以分为高斯求积法和无迹变换法(Unscented Transformation)。高斯求积法是最精确的,但也是成本最高的。在实际应用中,高斯求积法一般应用于精度要求高而维数非常低的情况。这它两种方法的成本取决于需要线性化的函数 f 。当 f 不可微或者没有参数解析表示形式时,不能使用泰勒级数线性化,而应采用数值积分矩投影来进行线性化。

把线性化运算与 EP 算法结合,就可以进行非线性依赖的混合网络的推理。此时,矩投影只能在具有合法分布的中间因子上进行,这极大地约束了聚类图的结构以及消息传递的顺序。

目前,处理含有非线性依赖的混合网络推理的最新方法是多项式混合(Mixtures of Polynomials, MOP)^[35]方法。相比 MTE, MOP 能更好地处理连续变量之间的确定性条件依赖关系^[125]。

7.4 抽样近似算法

如果初始因子不能表示成高斯分布,并且如果这些因子是多峰值的,那么把因子近似为高斯分布的效果很差。使用半参数或非参数方法可以避免这些不适当的参数假设,如近似抽样算法。

混合网络中最常用的近似抽样方法为重要性抽样^[126]和 MCMC 法,这两种方法只需稍作改变就可以应用在混合网络中。当分布的某个局部可以近似为高斯分布时,压缩粒子法会更加有效,可以避免在高维空间中抽样。还可以把消息传递与消息的非参数近似相结合,融合两种方法的优点。

8 暂态模型中的推理

DBN 是暂态模型中最主要的模型,是 BN 在时间轴上的展开。暂态模型的推理任务有 4 种:(1)滤波,也称为追踪,在时刻 t ,已知目前获得的所有观测值 $\mathbf{o}^{(1:t)}$,计算信任状态 $P(\mathbf{X}^{(t)} | \mathbf{o}^{(1:t)})$;(2)预测,在时刻 t ,已知观测值 $\mathbf{o}^{(1:t)}$,预测时刻 $t' > t$ 时变量(或某个子集)的分布;(3)平滑,已知某个更长轨线上的所有证据 $\mathbf{o}^{(1:T)}$,计算 $\mathbf{X}^{(t)}$ 的后验概率 $P(\mathbf{X}^{(t)} | \mathbf{o}^{(1:T)})$;(4)已知证据,求系统的最有可能轨线,即 $\arg \max_{\xi^{(0:T)}} P(\xi^{(0:T)} | \mathbf{o}^{(1:T)})$,该任务是 MAP 问题的一个实例。本节只介绍追踪任务的算法,因为追踪任务的解决方法可以简单地推广到其它推理任务中。

8.1 精确推理

DBN 的精确推理一般先把 DBN 的展开,然后使用团树算法进行推理。由于 DBN 的表示满足马尔可夫独立性假设,即若对所有 $t \geq 0$,有 $(\mathbf{X}^{(t+1)} \perp \mathbf{X}^{(0:t-1)} | \mathbf{X}^{(t)})$,那么概率分布可简化为

$$P(\mathbf{X}^{(0:T)}) = \prod_{t=0}^{T-1} P(\mathbf{X}^{(t+1)} | \mathbf{X}^{(0:t)}) = \prod_{t=0}^{T-1} P(\mathbf{X}^{(t+1)} | \mathbf{X}^{(t)}) \quad (37)$$

因此,可给出一个递归法则,只需保存随时间变化的信任状态,而不用一直保存整个网络或者不断增长的观测序列。这种递归法则可应用在团树算法的向上传递阶段来求解滤波问题,把消息从时间段 0 传递到后续时间段。这些消息表示信任状态,即系统状态的当前信任。

DBN 的消息传播实现如下:构造一个两时间段变量上的模板团树,然后把时刻 t 的消息传递到模板团树中,运行团树算法得到时刻 $t+1$ 的消息,把此消息作为输入消息,重复上述步骤。但是,DBN 存在耦合现象,即信任状态没有条件独立结构,因此无法进行任意形式的因式分解,这使得信任状态的表示可能很复杂。

8.2 近似推理

当把 DBN 看作是 BN 的展开并用 BN 表示 DBN 时,可采用 BN 所适用的任意近似推理算法。但是,暂态背景中所要处理的任务类型通常需要在任意长并且可能是无界长的时间轨线上推理,而且在有限大小 BN 中能获得合理误差的近似推理算法,有可能在任意大的 DBN 中表现不佳。所以,DBN 中的近似推理算法有 3 种专门方法:因式信任状态算法、粒子滤波算法和确定性搜索算法。

因式信任状态算法^[38, 39],把消息表示为因子乘积^[127],利用每个时间段中的团树或者聚类图来进行近似消息传播,把时刻 t 的近似信任状态映射到时刻 $t+1$ 的近似信任状态。在求解滤波任务时,该应用甚至比原来的 EP 算法还要简单,因为不用使用反向消息传递。

粒子滤波算法的问题是^[40, 41, 128]:如果简单地前向传播粒子,使用证据来调整其权值,那么粒子的权值会快速地趋向于 0,使得粒子集与真实状态的相关性很少甚至没有,从而出现粒子贫乏问题,使得估计值的方差随时间 t 的增大而快速增

加。因此,需要引入再抽样步骤^[129,130],以在去掉不良样本的同时对优良样本进行复制。选择新样本的方法,包括自举滤波器和序列重要性抽样^[131],以及使用 MCMC 或者其他方法来构造产生新样本的建议分布^[132]。另外,把抽样与某种全局推理结合起来的混合方案在 DBN 中也非常有效。

确定性搜索算法^[43]寻求能生成高概率取值的粒子集,并且该粒子集能一直追踪最有可能的时间轨线。暂态模型的确定性搜索一般与压缩粒子算法相结合,在某个变量子集上进行搜索,而对其它变量进行边际化。这种方法能给出边际 MAP 问题的近似解。

8.3 混合 DBN 的推理

混合 DBN 的推理类似于普通混合网络的推理情况,因此可把 DBN 的推理算法与混合网络的推理算法相组合,来求解混合 DBN 的推理问题。最常用的组合方式是把粒子滤波方法应用在连续或混合网络中。

在 DBN 高斯网络中,可采用基本高斯 BP 算法来求解滤波问题,这就是卡尔曼滤波器算法。当系统中存在非线性依赖时,可使用混合网络中的方法来推导扩展卡尔曼滤波器和扩展 Unscented 变换滤波器。

当 DBN 系统同时包含离散和连续变量时^[133],为了避免信任状态表示中高斯混合的项数的无限增大,必须控制高斯混合项的个数,解决方法分两类:剪枝算法和压缩算法。剪枝算法通过丢弃某些高斯分布来减少信任状态中混合分布项的个数。压缩算法则把高斯混合划分为若干子集,然后把每个子集中的所有高斯分布压缩成一个高斯分布。常用的压缩算法有 GPB1 算法、GPB2 算法和交互式多模型(Interacting Multiple Model, IMM)算法^[44]。这些压缩方法都是使用时间轨线中的时间窗长来确定压缩哪些高斯分布,可看作是 EP 算法的变型,每种算法使用不同的消息近似策略。

9 概率图模型推理的最新研究进展

9.1 变分推理算法及其变种

针对一般的变分推理算法只实现特定模型上的推理的问题,文献[134]把目标函数的导数作为变分近似的期望,从变分近似抽样得到包含噪声的但却是无偏的梯度值,用梯度值更新模型的参数值;另外,对于每一个样本,噪声梯度的求解需要计算观察样本和抽样变量的联合分布、变分分布和变分分布取对数后的梯度;而且对于任意类型的变分分布,变分分布取对数后的梯度和抽样方法可被多个模型重复使用,从而实现了黑箱变分推理(Black box variational inference)。文献[135]针对相关主题模型(correlated topic model)、贝叶斯逻辑斯蒂回归模型(Bayesian logistic regression model)和层次贝叶斯逻辑斯蒂回归模型(hierarchical Bayesian logistic regression model)等非共轭模型,提出了基于坐标上升优化迭代过程的拉普拉斯变分推理和 delta 变分推理,它们属于平均场变分近似推理算法。文献[136]利用变分近似算法,将图模型的联合概率分布的熵分解为图模型变量子集边际分布上的熵之和,利用熵函数的次模特性,得到最大次模函数的凸变分近似问题,但该方法仅适用于有向图模型且最大次模函数问题硬度很高,而且凸变分近似也不一定紧。

9.2 提升推理算法及其变种

提升推理(lifted inference)算法利用模型关系结构的对

称性解决关系统计模型过于复杂而无法实现推理的问题,但提升推理算法只适用于无条件概率下的关系模型。文献[137,138]指出在一元证据(unary evidence)下提升推理算法的复杂度是多项式时间的(polynomial time),并且提升推理算法可应用于此情形。文献[139]指出在二元证据(binary evidence)下提升推理算法是 #P 困难的(#P-hard),需要通过低布尔秩矩阵(low Boolean rank matrix)对二元证据进行近似,从而保持模型的对称性,使得提升推理算法可高效执行。文献[140]把相似的对象作为一组,即看作同一个对象,构造粗粒度的提升近似网络,从而减少网络构造时间。实验显示该方法只是稍微牺牲了进度却极大地改进了推理效率。

9.3 基于概率程序的推理及其变种

概率程序(probabilistic programs)有命令式(imperative)、函数式(functional)和逻辑式(logic)语言。概率图模型可用概率程序来表示,例如贝叶斯网络能用无循环的概率程序表示,马尔科夫链能用有循环的概率程序表示,利用概率程序可实现有效的推理^[141,142]。文献[143]利用概率程序进行推理,并且设计了有效的 MCMC 核,使得推理速度得到极大提升。另外,随机逻辑程序(stochastic logic programs, SLPs)可看作是非确定性概率程序,文献[144]基于随机逻辑程序提出了一种叫做 ProPPR(Programming with Personalized PageRank)的概率程序,由于其局部查询技术不依赖于问题的大小,通过对相互关联的谓词打分,可对上百万个输入的知识库实现有效推理,实验结果显示其能够学习上百个句子,快于马尔科夫逻辑网(Markov logic networks),而且允许句子出现递归关系。ProbLog 是一种概率逻辑程序(Probabilistic Logic Programming, PLP),文献[145]中提出了一种 MCMC (Markov Chain Monte Carlo, MCMC)算法来对 ProbLog 查询的条件概率进行推理。文献[146]提出了一种自适应的 MC-MC 算法对概率逻辑程序查询的条件概率进行近似推理。文献[147]利用概率图模型的推理技术,通过附加加权布尔公式(weighted Boolean formula)来实现概率逻辑程序推理,并给出了概率逻辑程序、概率图模型和统计关系学习之间的联系。

9.4 关于特殊数据的推理

文献[148]针对噪声聚合数据,诸如列联表数据或计数数据上的集合图模型(collective graphical models),提出期望最大近似边际推理算法。文献[149]考虑具有一致先验(consensus priors)的多个异构信息源的重构问题,为了集成异构的信息源,假定外部信息源具有高度相关性,提出了隐因式分解模型,该模型能够重构作为共同信息源的隐变量。当数据出现非随机缺失时,文献[150]提出了缺失值图(Missingness Graphs),并讨论了关系一致性估计问题,给出了缺失值可重构的条件。众所周知,高斯马尔科夫网和关系马尔科夫网分别适用于连续空间推理和离散标签空间上的推理,文献[151]针对该问题进行了推广,提出了基于 copulas 理论的隐关系模型,该模型能够在实现离散空间推理的同时引入连续空间的关系依赖模型。文献[152]把部分观测马尔科夫决策过程(Partially Observable Markov Decision Processes, POMDP)视为动态贝叶斯网模型上的边际最大后验推理,利用对偶表示得到贝叶斯变分近似上界,设计混合消息传播算法(hybrid message-passing algorithms)来求解与部分观测马尔科夫决策过程等价的动态贝叶斯网模型的近似变分边际最大后验推理问题。

9.5 推理方法在计算机视觉上的应用

现代马尔科夫随机场模型在图像处理中的应用通常考虑存在高阶相互作用的情形,在该情形下图中节点之间具有灵活的连接结构,而且不同的节点变量上具有很大的标签个数。文献[153]讨论了因子阶数达到300、近邻数达27个之多的稠密连接和空间上非一致连接的高阶马尔科夫随机场模型;对于图像分类个数未知的基于划分的超像素模型,文献[153]还通过实验比较了多面体组合优化算法(Polyhedral and Combinatorial Methods)、消息传播算法(Message Passing Methods)和最大流标签轮换算法,实验结果表明,对于小到中等规模的问题,整数规划割平面和分支定界能够提供最优解,速度也较快;而对于大规模问题,最大流标签轮换算法更有效而且精度更高。文献[154]讨论了特殊的应用场景下的概率图模型推理算法的构造,如视频序列,通过连接临近的帧中的图像块构造暂态树结构(temporal tree structure),执行精确的推理,并用环信任传播算法(loopy belief propagation algorithm)对帧标签平滑,用随机决策森林半监督训练估计视频序列像素的类标签。文献[155]利用了基于离散马尔科夫随机场结构特性和部分最优化方法,能够求解具有 2048×2048 网格的广义Potts模型(Potts models),用时少于90秒。

9.6 其它推理方法

文献[156]将随机最优化算法应用于平均场变分推理(mean-field variational inference)、概率主题模型(probabilistic topic models)、隐狄氏分配(latent Dirichlet allocation)和层次狄氏过程主题模型(hierarchical Dirichlet process topic model),可对上百万个文本进行处理。文献[157]把隐狄氏分配表示为马尔科夫随机场中的因子图,用有环信任传播算法(loopy belief propagation algorithm)进行推理,在基于LDA的主题模型的变种如作者-主题模型(Author-Topic Models, ATM)和关系主题模型(Relational Topic Models, RTM)上实现了因子图表示上的推理。文献[158]使用对数线性稀疏化模型实现了带边际分布等式约束的嵌套马尔科夫模型(nested Markov model)上的推理。文献[159]针对多层连续隐变量贝叶斯网,构造了辅助贝叶斯网,通过定义适当的生成函数和辅助隐变量来产生辅助概率密度函数,因为在给定父节点和独立辅助变量集情况下连续隐变量为条件确定的且隐变量具有很大的马尔科夫覆盖(Markov blanket),从而导致非常快的基于梯度的推理。文献[160]针对具有高势的图模型,引入辅助变量得到更紧的线性规划放松(tighter linear program relaxations),从而得到更精确的推理结果。文献[161]对于图模型联合分布的析取范式(Disjunctive Normal Form, DNF)作为查询条件的近似推理问题,给出了近似边际推理(approximate marginal inference)算法,从而推广了仅局限于简单边际查询的经典推理技术。另外,由于现在遇到的问题的规模很大,现有推理算法效率有限,文献[162-164]提出了各种并行推理(parallel inference)语言,其极大地提高了推理效率。

10 概率图模型推理的研究趋势

国内外学者对概率图模型推理做了大量研究工作,概率图模型的推理方法已经取得一定的进展。虽然概率图模型已在多个领域成功应用,但其推理算法仍存在许多问题有待解

决。本节给出概率图模型推理方法的一些未来可能的研究方向。

10.1 变分近似推理优化算法

进一步改善变分近似推理算法,有3个主要方面值得尝试:(1)对于用于变分推理的目标函数,如何找到更紧的目标函数的上界或下界近似能量泛函,需要更精细的研究。在这方面,可以借鉴各种受控不等式理论^[165],也可以分别使用上界近似能量泛函和下界近似能量泛函推理,最后组合两者的结果可能得到更准确的推理结果;(2)关于变分近似推理的约束空间,选择更紧的约束空间的松弛边界,如增加高阶约束^[166],也可提高变分近似精度,但高阶约束必然带来计算复杂度的增加,需要寻求兼顾精度和计算量的最优高阶约束;(3)对于变分近似推理的优化算法,可研究能充分利用图结构内部特性的特定优化算法,使优化算法收敛速度加快。

变分近似推理本质上是在求解非线性泛函方程或非线性泛函不等式,非线性泛函分析理论中与变分法相并行的理论如非扩张映射^[167]、单调算子理论^[168]等也可以用于对能量泛函优化问题的求解。

10.2 变分推理算法收敛性与近似精度的研究

CGBP算法与EP算法在不动点方程迭代的过程中,不能保证每一次迭代都能使目标函数单调增加,而且收敛后得到的不动点的最优性与唯一性也没有理论上的分析证明。可以将有限样本非渐进经验过程理论^[169]和临界点理论^[170]引入概率图模型的推理分析中,用于理论分析变分推理算法解的存在性、唯一性和收敛性。

变分近似推理算法缺乏近似精度的理论研究,这是由于近似精度对网络结构和参数的变化比较敏感。后续研究可针对某些特定网络来详细分析近似精度,如Ising模型、高斯模型、隐马尔科夫模型的近似精度分析等。

10.3 抽样推理的理论分析

归一化重要性抽样获得特定估计值所需的样本数并没有形式化的理论分析,多数情况下都是预置一个似乎足够大的抽样个数,或者利用算法迭代过程中目前产生的粒子的总权值来指导后续粒子的产生,但这些方法在很多情况下不适用,会出现早熟和粒子贫乏问题。最近,非概率图模型的保证错误概率上下界^[171]方面的研究工作非常活跃,建立具有错误概率上下界保证的推理算法理论是一个值得探讨的问题。

另外,MCMC推理方法的应用还有很多方面需要进行详细的理论分析,如建议分布的选择、运行所需的链数、评价混合程度的度量标准、保证样本相互独立的间隔的确定等。

10.4 MAP推理的线性优化算法问题

MAP推理的线性优化算法还存在有待进一步研究的问题:能否给出最大化-乘积CGBP与原始LP的对偶问题之间的联系?能否构建更好的方法来直接求解LP或者其对偶问题?即使推导的再参数化表示没有局部一致的联合取值,能否仍然可以利用各种启发式方法构造出一个取值?对于这些问题的思考,必将搞清MAP推理的线性优化算法机理。

10.5 DBN推理算法在非齐次过程上的扩展

目前DBN的推理算法都是假设不同时间段的近似结构是相同的,所以固定的近似结构可以适用于所有时间段。但是,在实际应用中,有很多动态过程不是齐次的,不同时间段具有不同的随机特性,这就需要根据时间段的不同而动态地

采用不同的近似结构,因此需要开发自适应近似结构选择准则,不同的时间段使用不同的近似结构。该问题还有待解决。

结束语 概率图模型是不确定性推理体系的一种有效工具,正吸引各领域学者投入更多的研究。本文根据不同的网络结构和不同的查询问题类型综述了概率图模型的多种推理算法,梳理了 BN 和 MN 中概率查询的精确推理与近似推理算法,还专门讨论了解决 MAP 查询的多种常用推理算法。BN 和 MN 中的大多数推理算法都可推广到混合网络和暂态网络中,本文也概述了混合网络与暂态网络的推理方法,最后给出了概率图模型推理方法的未来研究趋势。概率图模型是不确定性推理的一种强有力工具,在人工智能和机器学习领域势必日益重要。概率图模型应时刻结合相关领域出现的新理论新技术,不断完善自身体系,以便更好地在实际领域中应用。

参考文献

- [1] Estabragh, Shojaei Z, et al. Bayesian network model for diagnosis of Social Anxiety Disorder[C]// IEEE International Conference on Bioinformatics and Biomedicine Workshops. 2011: 639-640
- [2] Bickson D, Baron D, Ihler A, et al. Fault Identification Via Non-parametric Belief Propagation[J]. IEEE Transactions on Signal Processing, 2011, 59(6): 2602-2613
- [3] Zhang Lei, Ji Qiang. Bayesian Network Model for Automatic and Interactive Image Segmentation[J]. IEEE Transactions on Image Processing, 2011, 20(9): 2582-2593
- [4] Fernandez R, Picard R. Recognizing affect from speech prosody using hierarchical graphical models[J]. Speech Communication, 2011, 53(9/10): 1088-1103
- [5] Badiu M A, Kirkelund G E, Manchón C N, et al. Message-passing algorithms for channel estimation and decoding using approximate inference[C]// IEEE International Symposium on Information Theory. Cambridge, MA, USA: IEEE, 2012: 2376-2380
- [6] Jordan M I, Ghahramani Z, Jaakkola T S, et al. An introduction to variational methods for graphical models[J]. Machine learning, 1999, 37(2): 183-233
- [7] Frey B J, Jojic N. A comparison of algorithms for inference and learning in probabilistic graphical models[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2005, 27(9): 1392-1416
- [8] Guo H, Hsu W. A survey of algorithms for real-time Bayesian network inference[C]// AAAI Workshop on Real-Time Decision Support and Diagnosis Systems. Edmonton, Canada: AAAI Press, 2002: 1-12
- [9] 程强, 陈峰, 董建武, 等. 概率图模型中的变分近似推理方法[J]. 自动化学报, 2012, 38(11): 1721-1734
- [10] Jordan M I. Graphical models[J]. Statistical Science, 2004, 19(1): 140-159
- [11] Koller D, Friedman N. Probabilistic graphical models: Principles and techniques[M]. Cambridge, MA, USA: MIT Press, 2009: 1-283
- [12] Dechter R. Bucket elimination: A unifying framework for reasoning[J]. Artificial Intelligence, 1999, 113(1/2): 41-85
- [13] Darwiche A. Recursive Conditioning: Any-space conditioning algorithm with treewidth-bounded complexity[J]. Artificial Intelligence, 2000, 126(1/2): 5-41
- [14] Lauritzen S L, Spiegelhalter D J. Local computations with probabilities on graphical structures and their application to expert systems[J]. Journal of the Royal Statistical Society, 1988, 50: 157-224
- [15] Jordan M Z, Ghahramani T, et al. Introduction to variational methods for graphical models[J]. Machine Learning, 1999, 37: 183-233
- [16] Wainwright M J, Jordan M I. Graphical Models, Exponential Families, and Variational Inference[J]. Machine Learning, 2008, 1(1/2): 1-305
- [17] Ajroud A, Omri M N, Youssef H, et al. Loopy Belief Propagation in Bayesian Networks: origin and possibilistic perspectives[J]. arXiv preprint arXiv:1206.0976, 2012
- [18] Minka T P. Expectation propagation for approximate Bayesian inference[C]// Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence. Washington, USA: Morgan Kaufmann Publishers, 2001: 362-369
- [19] Opper M, Winther O. Expectation consistent approximate inference[J]. Journal of Machine Learning Research, 2005, 6: 2177-2204
- [20] Geiger D, Meek C, Wexler Y. A variational inference procedure allowing internal structure for overlapping clusters and deterministic constraints[J]. Journal of Artificial Intelligence Research, 2006, 27: 1-23
- [21] Rubinstein R Y, Kroese D P. Simulation and the Monte Carlo method[M]. New Jersey, USA: Wiley Press, 1981
- [22] Andrieu C, De Freitas N, Doucet A, et al. An introduction to MCMC for machine learning[J]. Machine learning, 2003, 50(1/2): 5-43
- [23] Casella G, Robert C P. Rao-Blackwellisation of sampling schemes[J]. Biometrika, 1996, 83(1): 81-94
- [24] Gogate V G. Sampling Algorithms for Probabilistic Graphical Models with Determinism[D]. California: University of California, 2009
- [25] Weiss Y, Freeman W T. On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs[J]. IEEE Transactions on Information Theory, 2001, 47(2): 736-744
- [26] Felzenszwalb P F, Huttenlocher D P. Efficient belief propagation for early vision[J]. International journal of computer vision, 2006, 70(1): 41-54
- [27] Wainwright M J, Jaakkola T S, Willsky A S. MAP estimation via agreement on trees: message-passing and linear programming[J]. IEEE Transactions on Information Theory, 2005, 51(11): 3697-3717
- [28] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(11): 1222-1239
- [29] Sakurikar P, Narayanan P J. Fast graph cuts using shrink-expand reparameterization[C]// Proceedings of the 2012 IEEE Workshop on the Applications of Computer Vision. Breckenridge, CO, USA: IEEE, 2012: 65-71
- [30] Park J D, Darwiche A. Approximating MAP using local search[C]// Proceedings of the Seventeenth conference on Uncertainty

- in artificial intelligence. Washington, USA; Morgan Kaufmann Publishers, 2001; 403-410
- [31] Lerner U, Segal E, Koller D. Exact inference in networks with discrete children of continuous parents[C]// Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence. Washington, USA; Morgan Kaufmann Publishers Inc. , 2001; 319-328
- [32] Madsen A L. Belief update in CLG Bayesian networks with lazy propagation[J]. International Journal of Approximate Reasoning, 2008, 49(2): 503-521
- [33] Lauritzen S L, Jensen F. Stable local computation with conditional Gaussian distributions [J]. Statistics and Computing, 2001, 11(2): 191-203
- [34] Moral S, Rumi R, Salmeron A. Mixtures of truncated exponentials in hybrid Bayesian networks[J]. Lecture notes in computer science, 2001; 156-167
- [35] Shenoy P P, West J C. Inference in hybrid Bayesian networks using mixtures of polynomials[J]. International Journal of Approximate Reasoning, 2011, 52(5): 641-657
- [36] Darwiche A. Constant-space reasoning in dynamic Bayesian networks[J]. International journal of approximate reasoning, 2001, 26(3): 161-178
- [37] Bilmes J A, Bartels C. On triangulating dynamic graphical models[C]// Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence. Edmonton, Alberta, Canada; Morgan Kaufmann Publishers, 2002; 47-56
- [38] Cohn I, El-Hay T, Friedman N, et al. Mean field variational approximation for continuous-time Bayesian networks [J]. The Journal of Machine Learning Research, 2010, 11: 2745-2783
- [39] Nodelman U, Koller D, Shelton C. Expectation propagation for continuous time Bayesian networks [C]// Proceedings of the Twenty-first Conference on Uncertainty in Artificial Intelligence. Edinburgh, Scotland, UK; AUA Press, 2005; 421-430
- [40] Koller D, Lerner U. Sampling in factored dynamic systems[M]. New York: Springer, 2001; 445-464
- [41] Lopes H F, Tsay R S. Particle filters and Bayesian inference in financial econometrics[J]. Journal of Forecasting, 2011, 30(1): 168-209
- [42] Besada-Portas E, Plis S M, Jesus M, et al. Parallel subspace sampling for particle filtering in dynamic Bayesian networks[M]. Springer; Berlin Heidelberg, 2009; 131-146
- [43] Hajishirzi H, Amir E. Reasoning about Deterministic Actions with Probabilistic Prior and Application to Stochastic Filtering [C]// Proceedings of the Twelfth International Conference on the Principles of Knowledge Representation and Reasoning. 2010; 456-464
- [44] Johnston L A, Krishnamurthy V. An improvement to the interacting multiple model (IMM) algorithm[J]. IEEE Transactions on Signal Processing, 2001, 49(12): 2909-2923
- [45] Taghipour N, Fierens D, Davis J, et al. Lifted variable elimination with arbitrary constraints[C]// Proceedings of the fifteenth international conference on Artificial Intelligence and Statistics. La Palma, Canary Islands; JMLR. org, 2012, 22: 1194-1202
- [46] Nima T, Jesse D. Generalized counting for lifted variable elimination[C]// Proceedings of the Second International Workshop on Statistical Relational AI. 2012; 133-159
- [47] Yedidia J S, Freeman W, Weiss Y. Constructing free-energy approximations and generalized belief propagation algorithms[J]. IEEE Transaction on Information Theory, 2005, 51(7): 2282-2312
- [48] Dechter R, Rish I. Mini-buckets: A general scheme for approximating inference[J]. Journal of the ACM, 2002, 50(2): 107-153
- [49] Bodlaender H L. Treewidth: Structure and algorithms [M]// Structural information and communication complexity. Berlin Heidelberg; Springer, 2007; 11-25
- [50] Shoikhet K, Geiger D. A practical algorithm for finding optimal triangulations[C] // Proceedings of the Fourteenth National Conference on Artificial Intelligence and Ninth Conference on Innovative Applications of Artificial Intelligence. Providence, Rhode Island, USA; AAAI Press, 1997; 185-190
- [51] Berry A, Heggernes P, Simonet G. The Minimum Degree Heuristic and the Minimal Triangulation Process[J]. Computer Science, 2003, 2880; 58-70
- [52] Gelfand A, Kask K, Dechter R. Stopping Rules for Randomized Greedy Triangulation Schemes[C]// Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence. 2011; 1043-1048
- [53] Kask K, Gelfand A, Otten L, et al. Pushing the Power of Stochastic Greedy Ordering Schemes for Inference in Graphical Models[C]// Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence. 2011; 54-60
- [54] Allen D, Darwiche A. New advances in inference by recursive conditioning[C]// Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence. 2003; 2-10
- [55] Grant K, Horsch M C. Efficient Caching in Elimination Trees [C]// Proceedings of the Twentieth International Florida Artificial Intelligence Research Society Conference. Key West, Florida, USA; AAI Press, 2007; 98-103
- [56] Darwiche A. Recursive conditioning: any-space conditioning algorithm with treewidth-bounded complexity[C]// Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence. 2000; 5-41
- [57] Grant K, Horsch M. Conditioning graphs: practical structures for inference in bayesian networks[C]// Proceedings of the 18th Australian Joint Conference on Artificial Intelligence. Sydney, Australia; Springer, 2005; 49-59
- [58] Grant K, Horsch M C. Methods for constructing balanced elimination trees and other recursive decompositions[J]. International journal of approximate reasoning, 2009, 50(9): 1416-1424
- [59] Grant K, Scholten K. On the Structure of Elimination Trees for Bayesian Network Inference[J]. Advances in Soft Computing, 2010, 6348: 208-220
- [60] Kozlov A V, Singh J P. A parallel Lauritzen-Spiegelhalter algorithm for probabilistic inference[C]// Proceedings of the 1994 ACM/IEEE conference on Supercomputing. IEEE Computer Society Press, 1994; 320-329
- [61] Xia Y, Prasanna V K. Scalable node-level computation kernels for parallel exact inference[J]. IEEE Transactions on Computers, 2010, 59(1): 103-115
- [62] Lu Zheng, Ole M, Jike Chong. Belief Propagation by Message Passing in Junction Trees; Computing Each Message Faster Using GPU Parallelization[C]// Proceedings of the 27th Confer-

- ence on Uncertainty in Artificial Intelligence, 2011;1-9
- [63] Xia Ying-long, Prasanna V K. Distributed Evidence Propagation in Junction Trees on Clusters[J]. IEEE Transactions on Parallel and Distributed Systems, 2012, 23(7):1169-1177
- [64] Coughlan J, Shen H. Dynamic quantization for belief propagation in sparse spaces[J]. Computer Vision and Image Understanding, 2007, 106(1):47-58
- [65] Sudderth E B, Freeman W T, Ihler A T, et al. Nonparametric belief propagation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Madison, WI, USA: IEEE, 2003;605-612
- [66] Noorshams N, Wainwright M J. Stochastic belief propagation: Low-complexity message-passing with guarantees [C] // The 49th Annual Allerton Conference on Communication, Control, and Computing. Monticello, IL, USA: IEEE, 2011;269-276
- [67] Galinier P, Habib M, Paul C. Chordal Graphs and Their Clique Graphs[C]//Proceedings of the 21st International Workshop on Graph-Theoretic Concepts in Computer Science. Springer-Verlag, 1995;358-371
- [68] Habib M, Limouzy V. On some simplicial elimination schemes for chordal graphs[J]. Electronic Notes in Discrete Mathematics, 2009, 32:125-132
- [69] Matsui Y, Uehara R, Uno T. Enumeration of the perfect sequences of a chordal graph[J]. Theoretical Computer Science, 2010, 411(40):3635-3641
- [70] Habib M, Stacho J. Reduced clique graphs of chordal graphs[J]. European Journal of Combinatorics, 2012, 33(5):712-735
- [71] Wainwright M J, Jaakkola T S, Willsky A S. A new class of upper bounds on the log partition function[J]. IEEE Transactions on Information Theory, 2005, 51(7):2313-2335
- [72] Wainwright M J, Jordan M I. Log-determinant relaxation for approximate inference in discrete Markov random fields[J]. IEEE Transactions on Signal Processing, 2006, 54(6):2099-2109
- [73] Globerson A, Jaakkola T. Approximate inference using planar graph decomposition[C]//Proceedings of Advances in Neural Information Processing Systems. Vancouver, British Columbia, Canada: Curran Associates, 2007;473-480
- [74] Kumar P, Torr P, Zisserman A. Solving Markov random fields using second order cone programming[C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York, USA: IEEE, 2006;1045-1052
- [75] Dura-Bernal S, Wennekers T, Denham S L. Modelling object perception in cortex: Hierarchical Bayesian networks and belief propagation[C]//Proceedings of The 45th Annual Conference on Information Sciences and Systems. Baltimore, USA, 2011;1-6
- [76] Xiang X, Zhang M, Li G, et al. Real-time stereo matching based on fast belief propagation[J]. Machine Vision and Applications, 2012, 23(6):1219-1227
- [77] X Sheng-jun, L Guang-hui, L Xin. Image segmentation via ant colony algorithm and loopy belief propagation algorithm[C]//Proceedings of The International Joint Conference on Neural Networks. Brisbane, Australia, 2012;1-7
- [78] Ihler A T, Iii J, Willsky W A. Loopy belief propagation: Convergence and effects of message errors[J]. Journal of Machine Learning Research, 2006, 6(1):905-936
- [79] Heskes T. On the uniqueness of loopy belief propagation fixed points[J]. Neural Computation, 2004, 16(11):2379-2413
- [80] Gamarnik D, Shah D, Wei Y. Belief propagation for min-cost network flow: Convergence and correctness[J]. Operations research, 2012, 60(2):410-428
- [81] Meshi O, Jaimovich A, Globerson A, et al. Convexifying the Bethe free energy[C]//Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. Montreal, Canada, 2009;402-410
- [82] Ihler A T, Fisher J W, Willsky A S, et al. Loopy belief propagation: convergence and effects of message errors[J]. Journal of Machine Learning Research, 2006, 6(1):905-936
- [83] Meltzer T, Globerson A, Weiss Y. Convergent message passing algorithms: a unifying view[C]//Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence. Montreal, Canada, 2009;393-401
- [84] Zhou Hai-jun, Wang Chuang. Region graph partition function expansion and approximate free energy landscapes: Theory and some numerical Results[J]. Journal of Statistical Physics, 2012, 148(3):513-547
- [85] Seeger M, Nickisch H. Fast convergent algorithms for expectation propagation approximate Bayesian inference[C]//Proceedings of the 14th International Conference on Artificial Intelligence and Statistics. Lauderdale, USA, 2011;15:652-660
- [86] Bouchard-Côté A, Jordan M I. Optimization of structured mean field objectives[C]//Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence. Montreal, QC, Canada: AUAI Press, 2009;67-74
- [87] Xing E P, Jordan M I, Russell S. A generalized mean field algorithm for variational inference in exponential families[C]//Proceedings of the 19th Conference on Uncertainty in Artificial Intelligence. Acapulco, Mexico: Morgan Kaufmann Publishers, 2003;583-591
- [88] Welling M, Teh Y W. Linear Response for Approximate Inference[C]//Proceedings of Advances in Neural Information Processing Systems. Vancouver, British Columbia, Canada: MIT Press, 2003;361-368
- [89] Roudi Y, Hertz J. Mean field theory for nonequilibrium network reconstruction[J]. Physical Review Letters, 2011, 106(4):048702
- [90] Martino L, Míguez J. A generalization of the adaptive rejection sampling algorithm[J]. Statistics and Computing, 2011, 21(4):633-647
- [91] Gogate V, Dechter R. SampleSearch: Importance sampling in presence of determinism[J]. Artificial Intelligence, 2011, 175(2):694-729
- [92] Gogate V, Dechter R. Importance sampling-based estimation over AND/OR search spaces for graphical models[J]. Artificial Intelligence, 2012, 184-185:38-37
- [93] Yu H, van Engelen R. Arc refractor methods for adaptive importance sampling on large Bayesian networks under evidential reasoning[J]. International journal of approximate reasoning, 2010, 51(7):800-819
- [94] Perilla J R, Beckstein O, Denning E J, et al. Computing ensembles of transitions from stable states: Dynamic importance sampling[J]. Journal of computational chemistry, 2011, 32(2):196-209

- [95] Yuan C, Druzdzal M J. An importance sampling algorithm based on evidence pre-propagation[C]//Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence. Edmonton, Alberta, Canada; Morgan Kaufmann Publishers, 2002; 624-631
- [96] Yuan C, Druzdzal M J. Importance sampling in Bayesian networks; An influence-based approximation strategy for importance functions[C]//Proceedings of the 21th Annual Conference on Uncertainty in Artificial Intelligence. Edinburgh, Scotland; AUA Press, 2005; 650-657
- [97] Metropolis N, Rosenbluth A W, Rosenbluth M N, et al. Equation of state calculations by fast computing machines[J]. The journal of chemical physics, 1953, 21(6): 1087-1092
- [98] Hastings W K. Monte Carlo sampling methods using Markov chains and their applications[J]. Biometrika, 1970, 57(1): 97-109
- [99] Giordani P, Kohn R. Adaptive independent Metropolis-Hastings by fast estimation of mixtures of normals[J]. Journal of Computational and Graphical Statistics, 2010, 19(2): 243-259
- [100] Beskos A, Stuart A. Computational complexity of Metropolis-Hastings methods in high dimensions[M]. Berlin Heidelberg: Springer, 2009; 61-71
- [101] Asmussen S, Glynn P W. A new proof of convergence of MC-MC via the ergodic theorem[J]. Statistics & Probability Letters, 2011, 81(10): 1482-1485
- [102] Doucet A, Gordon N J, Krishnamurthy V. Particle filters for state estimation of jump Markov linear systems [J]. IEEE Transactions on Signal Processing, 2001, 49(3): 613-624
- [103] Van Dyk D A, Park T. Partially collapsed Gibbs samplers; Theory and methods[J]. Journal of the American Statistical Association, 2008, 103(482): 790-796
- [104] Bidyuk B, Dechter R. Cutset sampling for Bayesian networks [J]. Journal of Artificial Intelligence Research, 2007, 28(1): 1-48
- [105] Dechter R. Constraint Processing[M]. Morgan Kaufmann Publishers, 2003
- [106] Favier A, De Givry S, Legarra A, et al. Pairwise decomposition for combinatorial optimization in graphical models[J]. Artificial Intelligence, 2011, 22(3): 21-26
- [107] Sontag D, Meltzer T, Globerson A, et al. Tightening LP relaxations for MAP using message passing[C]//Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence. Helsinki, Finland, 2008; 503-510
- [108] Komodakis N, Paragios N. Beyond pairwise energies; Efficient optimization for higher-order MRFs[C]//IEEE Conference on Computer Vision and Pattern Recognition. Miami, FL, USA, 2009; 2985-2992
- [109] Werner T. A linear programming approach to max-sum problem; A review[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(7): 1165-1179
- [110] Globerson A, Jaakkola T. Fixing max-product: Convergent message passing algorithms for MAP LP-relaxations[C]//Proceedings of Advances in Neural Information Processing Systems 21. Vancouver, British Columbia, Canada; Curran Associates, 2007; 553-560
- [111] Zheng Yun, Chen Pei, Cao Jiang-zhong. MAP-MRF inference based on extended junction tree representation [C] // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012; 1696-1703
- [112] Jojic V, Gould S, Koller D. Accelerated dual decomposition for MAP inference [C] // Proceedings of The 27th International Conference on Machine Learning. Haifa, Israel, 2010; 503-510
- [113] Batra D, Gallagher A C, Parikh D, et al. Beyond trees; MRF inference via outer-planar decomposition[C]//IEEE Conference on Computer Vision and Pattern Recognition. San Francisco, CA, USA, 2010; 2496-2503
- [114] Martins A F T, Figueiredo M A T, Aguiar P M Q. An augmented Lagrangian approach to constrained MAP inference[C]//Proceedings of The 28th International Conference on Machine Learning. Bellevue, Washington, USA, 2011; 169-176
- [115] Boykov Y, Veksler O, Zabih R. Fast approximate energy minimization via graph cuts[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2001, 23(11): 1222-1239
- [116] Kolmogorov V, Roth C. Minimizing nonsubmodular functions with graph cuts-a review[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(7): 1274-1279
- [117] Fix A, Gruber A, Boros E, et al. A graph cut algorithm for higher-order Markov random fields [C] // IEEE International Conference on Computer Vision. Barcelona, Spain; IEEE, 2011: 1020-1027
- [118] Kumar M P, Veksler O, Torr P H S. Improved Moves for Truncated Convex Models [J]. Journal of Machine Learning Research, 2011, 12; 31-67
- [119] Park J D, Darwiche A. Solving MAP exactly using systematic search[C]//Proceedings of the Nineteenth Conference on Uncertainty in Artificial Intelligence. Morgan Kaufmann Publishers Inc. , 2002; 459-468
- [120] Malioutov D M, Johnson J K, Willsky A S. Walk-sums and belief propagation in Gaussian graphical models[J]. The Journal of Machine Learning Research, 2006, 7; 2031-2064
- [121] El-Kurdi Y, Giannacopoulos D, Gross W J. Relaxed Gaussian Belief Propagation[C]//IEEE International Symposium on Information Theory Proceedings. IEEE, 2012; 2002-2006
- [122] Langseth H, Nielsen T D, Rumi R, et al. Inference in hybrid Bayesian networks[J]. Reliability Engineering & System Safety, 2009, 94(10): 1499-1509
- [123] Sanner S, Abbasnejad E. Symbolic Variable Elimination for Discrete and Continuous Graphical Models[C]//Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence. 2012; 1954-1960
- [124] Langseth H, Nielsen T D, Salmerón A. Parameter estimation and model selection for mixtures of truncated exponentials[J]. International Journal of Approximate Reasoning, 2010, 51(5): 485-498
- [125] Shenoy P P. A re-definition of mixtures of polynomials for inference in hybrid Bayesian networks [M]. Berlin Heidelberg: Springer, 2011; 98-109
- [126] Fernández A, Rumi R, Salmerón A. Answering queries in hybrid Bayesian networks using importance sampling[J]. Decision Support Systems, 2012, 53(3): 580-590
- [127] Frognier C, Pfeffer A. Discovering weakly-interacting factors in a complex stochastic process [C] // Advances in Neural Information Processing Systems. Vancouver, British Columbia, Canada; Curran Associates, 2007; 481-488

- [128] Johansen A M, Doucet A. A note on auxiliary particle filters [J]. *Statistics & Probability Letters*, 2008, 78(12): 1498-1504
- [129] Del Moral P, Doucet A, Jasra A. On adaptive resampling strategies for sequential Monte Carlo methods[J]. *Bernoulli*, 2012, 18(1): 252-278
- [130] Lopes H F, Polson N G, Carvalho C M. Bayesian statistics with a smile; A resampling-sampling perspective[J]. *Brazilian Journal of Probability and Statistics*, 2012, 26(4): 358-371
- [131] Cappé O, Godsill S J, Moulines E. An overview of existing methods and recent advances in sequential Monte Carlo[J]. *Proceedings of the IEEE*, 2007, 95(5): 899-924
- [132] Li T, Sattar T P, Sun S. Deterministic resampling; unbiased sampling to avoid sample impoverishment in particle filters [J]. *Signal Processing*, 2012, 92(7): 1637-1645
- [133] Gogate V, Dechter R. Approximate Inference Algorithms for Hybrid Bayesian Networks with Discrete Constraints [C] // *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence*, Edinburgh, Scotland; AUSA Press, 2005: 209-216
- [134] Ranganath R, Gerrish S, Blei D. Black Box Variational Inference [C] // *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, Reykjavik, Iceland; JMLR. org, 2014: 814-822
- [135] Wang C, Blei D M. Variational inference in nonconjugate models[J]. *The Journal of Machine Learning Research*, 2013, 14(1): 1005-1031
- [136] Kumar K S, Bach F. Maximizing submodular functions using probabilistic graphical models[J]. *arXiv preprint arXiv: 1309.2593*, 2013
- [137] Van den Broeck G, Davis J. Conditioning in first-order knowledge compilation and lifted probabilistic inference [C] // *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, Toronto, Ontario, Canada; AAAI Press, 2012: 1-7
- [138] Van den Broeck G. Lifted Inference and Learning in Statistical Relational Models [D]. Leuven, Belgium; Catholic University of Leuven, 2013
- [139] Van den Broeck G, Darwiche A. On the complexity and approximation of binary evidence in lifted inference [C] // *Proceedings of the Advances in Neural Information Processing Systems*, Lake Tahoe, Nevada, United States, 2013: 2868-2876
- [140] Singla P, Nath A, Domingos P. Approximate lifted belief propagation [C] // *Proceedings of the AAAI Workshop on Statistical Relational Artificial Intelligence*, Atlanta, USA, 2010: 92-97
- [141] Gordon A D, Henzinger T A, Nori A V, et al. Probabilistic programming [C] // *Proceedings of the International Conference on Software Engineering*, Hyderabad, India; ACM, 2014: 57-70
- [142] Stuhlmüller A, Goodman N D. Reasoning about reasoning by nested conditioning; Modeling theory of mind with probabilistic programs [J]. *Cognitive Systems Research*, 2014, 28: 80-99
- [143] Yang L, Hanrahan P, Goodman N D. Generating Efficient MC-MC Kernels from Probabilistic Programs [C] // *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics*, Reykjavik, Iceland; JMLR. org, 2014: 1068-1076
- [144] Wang W Y, Mazaitis K, Lao N. Efficient Inference and Learning in a Large Knowledge Base; Reasoning with Extracted Information using a Locally Groundable First-Order Probabilistic Logic [J]. *arXiv preprint arXiv: 1404.3301*, 2014
- [145] Moldovan B, Thon I, Davis J, et al. MCMC estimation of conditional probabilities in probabilistic programming languages [C] // *Proceedings of the 12th European conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, Springer-Verlag, 2013: 436-448
- [146] Nampally A, Ramakrishnan C R. Adaptive MCMC-Based Inference in Probabilistic Logic Programs [J]. *arXiv preprint arXiv: 1403.6036*, 2014
- [147] Fierens D, Van den Broeck G, Renkens J, et al. Inference and learning in probabilistic logic programs using weighted boolean formulas [J]. *arxiv preprint arxiv: 1304.6810*, 2013, 2013: 1-44
- [148] Sheldon D, Sun T, Kumar A, et al. Approximate Inference in Collective Graphical Models [C] // *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, GA, USA; JMLR. org, 2013: 1004-1012
- [149] Praveen P, Fröhlich H. Boosting Probabilistic Graphical Model Inference by Incorporating Prior Knowledge from Multiple Sources [J]. *PLoS ONE*, 2012, 8(6): e67410
- [150] Mohan K, Pearl J, Tian J. Graphical models for inference with missing data [C] // *Proceedings of the Advances in Neural Information Processing Systems*, Lake Tahoe, Nevada, United States, 2013: 1277-1285
- [151] Xiang R, Neville J. Collective inference for network data with copula latent markov networks [C] // *Proceedings of the sixth ACM international conference on Web search and data mining*, Rome, Italy; ACM, 2013: 647-656
- [152] Kiselev I, Poupart P. Policy optimization by marginal-map probabilistic inference in generative models [C] // *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*, International Foundation for Autonomous Agents and Multiagent Systems, Paris, France; Springer, 2014: 1611-1612
- [153] Kappes J H, Andres B, Hamprecht F A, et al. A comparative study of modern inference techniques for discrete energy minimization problems [C] // *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Portland, USA: IEEE, 2013: 1328-1335
- [154] Badrinarayanan V, Budvytis I, Cipolla R. Semi-supervised video segmentation using tree structured graphical models [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(11): 2751-2764
- [155] Badrinarayanan V, Budvytis I, Cipolla R. Mixture of Trees Probabilistic Graphical Model for Video Segmentation [J]. *International Journal of Computer Vision*, 2013, 110(1): 1-16
- [156] Hoffman M D, Blei D M, Wang C, et al. Stochastic variational inference [J]. *The Journal of Machine Learning Research*, 2013, 14(1): 1303-1347
- [157] Zeng J, Cheung W K, Liu J. Learning Topic Models by Belief Propagation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013, 35(5): 1121-1134
- [158] Shpitser I, Evans R J, Richardson T S, et al. Sparse Nested Markov Models with Log-linear Parameters [J]. *arXiv preprint arXiv: 1309.6863*, 2013
- [159] Kingma D P. Fast Gradient-Based Inference with Continuous Latent Variable Models in Auxiliary Form [J]. *arXiv preprint arXiv: 1306.0733*, 2013

具体来说,正弦函数与余弦函数自变量的取值范围为 $[0, 2\pi]$,划分的子区间的个数为400,即每个子区间为 $[i(2\pi/400), (i+1)(2\pi/400)]$, $i=0, \dots, 399$ 。测试结果分别如图1(a)和图1(b)所示。指数函数的测试范围为 $[-10, 10]$,分割成200段小区间,误差结果如图1(c)所示,为图示效果,纵坐标为取自然对数之后的值。对数函数的测试范围为 $[0, 5000]$,分割成5000段小区间,测试结果如图1(d)所示。

从测试结果可以看出,尽管对于同一个函数,若自变量取值不同,误差界也是不同的。三角函数是一个周期函数,误差的变也呈现出周期变化的趋势。正弦函数的误差界在0到 $\pi/4$ 区间内逐渐增大,在 $\pi/4$ 到 $3\pi/4$ 区间达到峰值,而在 $3\pi/4$ 到 π 区间内误差界又逐渐减小;指数函数的误差界随着自变量的增大而增大;对数函数误差界的变化呈现跳跃阶梯状,在0到1区间误差界变化平缓,到达临界点1时误差界发生跳变,在临界点3000处再一次发生跳变。

结束语 本文针对GNU下初等函数的实现进行误差分析,得到一个理论误差界,基于此可得到一个理论上的区间赋值。由于理论分析比较精妙,需要较多的数学知识,对于更一般的数值程序,要追踪每一步的误差非常困难。相比之下,区间算术能避免理论分析的复杂性,对函数进行可验证赋值,并给出函数的误差界,主要工作在于把浮点程序改写成区间程序。目前我们已经完成了初等函数的可验证赋值及数值误差界的计算,除本文提到的4个函数,还包括反三角函数、正切函数、双曲函数等。但是,如果待分析和赋值的函数或数值程序的运算(特别是乘除运算)次数较多,就会导致得到的区间宽度太大而难以接受,在这种情况下,需要结合区间的精化等更为精妙的技术。

参 考 文 献

- [1] Wang Wen-hua, GUO Zhi-hua, CAO Huai-xin. An upper bound for the adiabatic approximation error[J]. Science China Physics, Mechanics and Astronomy, 2014, 57(2): 218-224
- [2] Jeannerod C P, Rump S M. Improved error bounds for inner products in floating-point arithmetic[J]. SIAM Journal on Matrix Analysis and Applications, 2013, 34(2): 338-344

(上接第18页)

- [160] Mezuman E, Tarlow D, Globerson A, et al. Tighter linear program relaxations for high order graphical models[J]. arXiv preprint arXiv:1309.6848, 2013
- [161] Dworkin L, Kearns M, Xia L. Efficient Inference for Complex Queries on Complex Distributions[C]// Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics. Reykjavik, Iceland: JMLR. org, 2014: 211-219
- [162] Scutari M. Bayesian Network Constraint - Based Structure Learning Algorithms: Parallel and Optimised Implementations in the bnlearn R Package[J]. arXiv preprint arXiv:1406.7648, 2014
- [163] Tristan J B, Huang D, Tassarotti J. Augur: a Modeling Language for Data-Parallel Probabilistic Inference[J]. arXiv preprint arXiv:1312.3613, 2013
- [164] Kollar T, Tellex S, Walter M R. Approaching the Symbol Grounding Problem with Probabilistic Graphical Models[J]. AI MAGAZINE, 2011, 32(4): 64-76
- [165] 石焕南. 受控理论与解析不等式[M]. 哈尔滨: 哈尔滨工业大学

- [3] 袁梦. 复杂计算的误差定量分析方法及其应用[D]. 杭州: 浙江大学, 2006
- [4] 周腾飞. Bernstein 基多项式函数的高精度计算及其动态误差分析研究[D]. 长沙: 国防科学技术大学, 2011
- [5] Rump S M. Verification methods: Rigorous results using floating-point arithmetic[J]. Acta Numerica, 2010, 19: 287-449
- [6] Muller J M. Elementary functions: Algorithms and Implementation(2nd ed)[M]. Birkhäuser, 2006
- [7] Granlund T. The GNU multiple precision arithmetic library [OL]. <http://www.swox.se/gmp>
- [8] Fousse L, Hanrot G, Lefevre V, et al. MPFR: A multiple-precision binary floating-point library with correct rounding [J]. ACM Transactions on Mathematical Software (TOMS), 2007, 33(2): 1-14
- [9] Backeljauw F, Becuwe S, Cuyt A, et al. Validated evaluation of special mathematical functions [M] // Intelligent Computer Mathematics, 2008: 206-216
- [10] Brisebarre N, De Dinechin F, Jeannerod C P, et al. Handbook of floating-point arithmetic [M]. Springer Science & Business Media, 2009
- [11] Higham N J. Accuracy and stability of numerical algorithms(2nd ed)[M]. Siam, 2011
- [12] Moore R E, Kearfott R B, Cloud M J. Introduction to interval analysis[M]. Siam, 2009
- [13] Stallman R M. Using GCC: The GNU Compiler Collection Reference Manual for GCC 3.3.1[M]. Free Software Foundation, 2003
- [14] Sauer T. Numerical Analysis(2nd ed)[M]. 北京: 机械工业出版社, 2012
- [15] Mathews J H, Fink K D. Numerical methods using MATLAB (第四版)[M]. 周璐, 陈渝, 钱芳, 等. 北京: 电子工业出版社, 2012
- [16] Brönnimann H, Melquiond G, Pion S. The design of the Boost interval arithmetic library[J]. Theoretical Computer Science, 2006, 351(1): 111-118
- [17] 刘保柱, 苏彦华, 张宏林. MATLAB 7.0 从入门到精通[M]. 北京: 人民邮电出版社, 2010

出版社, 2012

- [166] Werner T. High-arity interactions, polyhedral relaxations, and cutting plane algorithm for soft constraint optimisation (MAP-MRF)[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2008(CVPR 2008). IEEE, 2008: 1-8
- [167] Aoyama K, Kohsaka F. Fixed point theorem for a-nonexpansive mappings in banach spaces[J]. Nonlinear Analysis, 2011, 74: 4387-4391
- [168] Kien B T, Wong M M, Wong N C, et al. Solution existence of variational inequalities with pseudo-monotone operators in the sense of Brzbis[J]. Optimal Theory Application, 2009, 140: 249-263
- [169] Genest C, Remillard B. Test of independence and randomness based on the empirical copula process[J]. Test, 2004, 13(2): 335-369
- [170] 张恭庆. 临界点理论及其应用[M]. 上海: 上海科学技术出版社, 1986
- [171] Chen P N, Alajaji F. A Generalized Poor-Verdú Error Bound for Multihypothesis Testing[J]. IEEE Transactions on Information Theory, 2012, 58(1): 311-316