

概率图模型的表示理论综述

刘建伟¹,黎海恩²,周佳佳¹,罗雄麟¹

(1. 中国石油大学(北京)自动化系,北京 102249; 2. 新华医药化工设计有限公司,山东淄博 255086)

摘要: 概率图模型结合概率论与图论的知识,利用图结构表示变量的联合概率分布,近年已成为不确定性推理的研究热点.随着概率图模型在实际领域中的应用日益增加,不同的任务和应用环境对概率图模型的表示理论提出了不同的新要求.本文总结出近年来提出的多种概率图模型的表示理论.最后指出概率图模型的进一步研究方向.

关键词: 概率图模型;连续化;非齐次化;贝叶斯逻辑;马尔可夫逻辑;非参数化;矩阵正态图模型;Copula 函数;混合图模型

中图分类号: TP181

文献标识码: A

文章编号: 0372-2112 (2016)05-1219-08

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.3969/j.issn.0372-2112.2016.05.030

A Survey on the Representation Theory of Probabilistic Graphical Models

LIU Jian-wei¹, LI Hai-en², ZHOU Jia-jia¹, LUO Xiong-lin¹

(1. Department of Automation, China University of Petroleum, Beijing 102249, China;

2. Xinhua Pharmaceutical and Chemical Designing Company Limited, Zibo, Shandong 255086, China)

Abstract: Probabilistic Graphical models bring together graph theory and probability theory in a single formalism, so the joint probabilistic distribution of variables can be represented using graph. In recent years, probabilistic graphical models have become the focus of the research in uncertainty inference, because of its bright prospect for the application. In this paper, we conclude the representation of probabilistic graphical models in recent years. Finally, a discussion of the future trend of probabilistic graphical models is given.

Key words: probabilistic graphical models; continuous; inhomogeneous; Bayesian logic; Markov logic; non-parametric; matrix normal graphical model; Couplua function; mixed graphical model

1 引言

近年来,概率图模型(Probabilistic Graphical Models, PGM)已受到众多领域的科学家们的关注. PGM 的表示理论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了有力的表示框架.自 Pearl 建立起贝叶斯网的框架^[1]后,众多学者对此进行研究,不断丰富概率图模型的体系.文献[2,3]详细介绍了概率图模型的表示、推理和学习的相关理论.在 PGM 经典表示理论的基础上,本文主要归纳了近几年来 PGM 的新表示理论.

PGM 的分类可分两种情况:(1)根据边有无方向性分类;(2)根据表示的抽象级别不同分类,如图 1 所示.根据边有无方向性,PGM 可以分为三类:(1)有向图模型^[2],也称为贝叶斯网(Bayesian Network, BN),其网络结构使用有向无环图;(2)无向图模型^[2],也称为马尔

可夫网(Markov Network, MN),其网络结构为无向图;(3)局部有向模型,即同时存在有向边和无向边的模型,包括条件随机场(Conditional Random Field, CRF)^[4]和链图(Chain Graph)^[5].

根据表示的抽象级别不同,PGM 可分两类:(1)基于随机变量的概率图模型,如贝叶斯网、马尔可夫网、条件随机场和链图等;(2)基于模板的概率图模型.这类模型根据应用场景不同又可分为两种:(a)为暂态模型,包括动态贝叶斯网(Dynamic Bayesian Network, DBN)^[6]和状态观测模型,其中状态观测模型又包括线性动态系统(Linear Dynamic System, LDS)和隐马尔可夫模型(Hidden Markov Model, HMM);(b)为对象关系领域的概率图模型^[7],包括盘模型(Plate Model, PM)、概率关系模型(Probabilistic Relational Model, PRM)和关系马尔可夫网(Relational Markov Network, RMN).

所有概率图模型均可使用独立等价手段简化概率图

模型,利用概率图节点的两两独立、局部独立和全局独立关系可以简化联合概率分布的表示,利用网络结构与概率分布间独立等价关系如独立映射、最小 I-映射和完美映射,可以实现不同的概率图同一联合概率分布,从而需要找到最适合学习推理的概率图表示形式. BN、链图、DBN、LDS、HMM、PM 和 PRM 属于有向图模型, MN、CRF 和 RMN 属于无向图模型. 有向图模型的联合概率分布,可利用参数和非参数概率模型对图中节点、边、团、迹等图结构单元参数化得到,联合概率分布可直接分解为每个节点所对应的概率分布的乘积,有利于大规模学习和推理. 有向图模型主要应用于具有因果关系的场景中,如医疗诊断、故障诊断、统计决策和专家系统等方面. 无向图模型的联合概率分布,可利用参数和非参数概率模型对图中的团参数化得到,联合概率分布由图中包含的团势乘积经过归一化得到,给学习和推理带来很大的困难,通常只能通过局部独立假设,通过消息传播,求解到近似联合概率分布. 无向图模型不能表示变量间的因果关系,无向图模型不能对有因果关系的问题进行学习推理,无向图模型主要应用于计算机视觉、自然语言处理、信息解码和语言识别等领域.

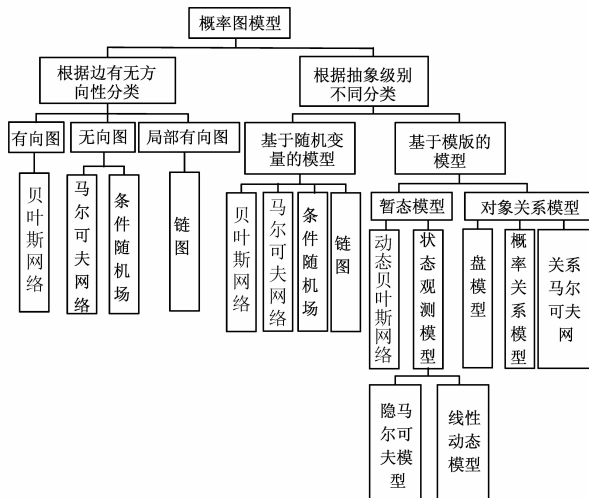


图 1 概率图模型的分类

2 概率图模型的表示理论

随着 PGM 在实际领域中的应用日益增加,不同的任务和应用环境对 PGM 的表示理论提出了不同的新要求. 为了更好地适应新应用场景以及得到能更加准确地反映真实系统的模型,研究人员对 PGM 的表示理论作出了巨大的努力,推动了 PGM 的快速发展.

2.1 概率图模型在暂态场景中的发展

在暂态模型中,为了简化表示问题使模型易于处理,模型需要满足三个假设,包括离散假设、马尔可夫假设和平稳假设. 但是,这些假设条件过于苛刻,通常不能

满足实际场景的情况,这限制了暂态模型的广泛应用. 研究学者对这些假设条件进行放松,推动了暂态模型的进一步发展,特别是动态贝叶斯网表示理论的发展.

2.1.1 连续化

离散假设以固定时间间隔把时间轴离散化,动态系统的模型只能在这些固定的时间点上转移,从上一时间点转移到下一时间点,其推理过程也只能查询这些时间点上的系统状态. 但很多实际场景中不能离散化时间轴,需要在连续的时间轴上建模,如某些变量会随着时间快速变化的系统. 为了解决该问题, Nodelman 等人定义了连续时间贝叶斯网 (Continuous Time Bayesian Network, CTBN)^[8],把马尔可夫过程推广到分解域中. CTBN 编码了一个指数状态空间上的齐次连续马尔可夫过程. 由于 CTBN 状态空间的复杂性,使得其推理和学习过程难以处理,很多学者致力于研究 CTBN 的有效推理和学习算法. 文献[9]提出了 CTBN 的期望传播推理算法. 文献[10]结合期望最大化和结构期望最大化算法,提出一种 CTBN 的结构和参数学习的算法,有效地实现从局部观测数据中学习 CTBN. 文献[11]提出 CTBN 的平均场变分近似推理算法,把 CTBN 的后验概率分布近似为独立单元的乘积,每个单元都是一个非齐次连续马尔可夫过程.

在动态建模过程中,除了把有向图模型 BN 进行连续化作为建模工具外, El-Hay 等人还提出连续时间马尔可夫网 (Continuous Time Markov Network, CTMN)^[12]. CTMN 同样能为连续时间动态系统建模,其暂态过程的平稳分布具有更加简洁的图模型表示.

2.1.2 非齐次化

平稳假设也称为齐次假设,假设了不同时间点上的数据来自于同一个齐次马尔可夫过程. 但实际上,系统变量之间的依赖关系或者其参数模型可能会随着时间的变化而改变,使得数据不再是来自于同一个过程. 为了使 DBN 适用于这类暂态系统的建模,研究人员对平稳假设进行放松,提出非齐次 DBN.

Robinson 和 Hartemink 提出离散非齐次 DBN^[13],不同时间点上的模型具有不同的结构,并且带有对结构差异性进行惩罚的正则化项. Grzegorzcyk 和 Husmeier 提出连续非齐次 DBN^[14],其参数允许随着时间变化,并给出一个共享网络来提供不同时间点上所共享的信息. Lebre^[15]提出另一种连续非齐次 DBN,允许网络结构在不同时间点上变化. 而文献[16]和文献[17]提出的模型虽然基于非贝叶斯方法,但也十分类似于非齐次 DBN,只是其推理基于参数的稀疏 L1 范数正则化回归. 而 Grzegorzcyk 和 Husmeier 提出适用于连续数据的非齐次 DBN^[18]. 为了解决基因调控网络结构随时间变化的问题, Dondelinger 等人提出一种带有贝叶斯正则化

项的非齐次 DBN^[19],可适用于连续型数据,网络结构能够随时间变化,并给出了四种不同的正则化策略. Grzegorzycyk 和 Husmeier 还提出正则化非齐次 DBN^[20],使不同分隔时间段上的信息能够共享,其耦合机制使得不同分隔时间段的参数共享类似的先验概率分布模型.

2.2 概率图模型与逻辑语言的结合

概率图模型可以有效处理不确定性,而逻辑语言可以简洁地表示领域中的大量知识.因此,结合 PGM 与逻辑语言各自的优点一直是人工智能领域的目标,也是多数实际应用的迫切需要.目前,研究人员已开发出多种概率逻辑语言. PRM 结合了有向图模型 BN 与关系逻辑语言的优点,而 RMN 则是基于无向图模型 MN 与关系逻辑语言的表示框架.除了对象关系域模型之外,PGM 与逻辑语言还有多种结合方式.

2.2.1 贝叶斯逻辑

Kristian 和 Luc 利用 BN 和有限子句逻辑的优点,提出贝叶斯逻辑程序 (Bayesian Logic Program, BLP),建立起底层网络与随机变量之间的一一映射^[21,22]. Kristian 和 Luc 还把 BLP 扩展到连续变量空间中^[23],并给出了 BLP 的学习方法. Sindhu 和 Raymond 提出贝叶斯溯因逻辑程序 (Bayesian Abductive Logic Program, BALP)^[24],把 BLP 与溯因逻辑联合起来用于溯因推理. BLP 使用逻辑演绎法来构造贝叶斯网,而 BALP 使用逻辑溯因法,更适用于规划识别和行为识别等问题. Laskey 还提出多实体贝叶斯网 (Multi-Entity Bayesian Networks, MEBN)^[25],MEBN 也是一种一阶贝叶斯知识库语言,能表示多种类型的概率分布,并可在多种应用场景中用于领域知识的表示.

2.2.2 马尔可夫逻辑

马尔可夫逻辑网 (Markov Logic Network, MLN) 由 Richardson 和 Domingos 提出^[26],是无向图模型与一阶逻辑结合最典型的情况. MLN 可以简洁地描述巨大的马尔可夫网,并能够灵活地表示丰富的领域知识.由于 MLN 定义在有限域中,不能充分发挥一阶逻辑的优点,因此 Singla 和 Domingos 通过吉布斯测度理论把 MLN 扩展到无限域中,并给出了与 MLN 势函数相一致的测度的存在性与唯一性的充分条件^[27]. Kok 和 Domingos 还把 MLN 扩展到二阶逻辑语言^[28]. Jain 等人提出带有动态参数的自适应 MLN^[29],通过对参数的动态调整,能有效表示变量间的随机作用关系.

溯因法是求解观察现象最优解释的方法,在自然语言处理中十分重要. Blythe 等人在马尔可夫逻辑中实现加权溯因,使用加权一阶规则来表示概率知识^[30].而 Singla 和 Mooney 利用 MLN 进行溯因推理,并提出隐原因模型和溯因模型使得 MLN 规则在规划识别应用中比其它方法更优越^[31].

一阶逻辑概率语言在推理和学习过程中都存在很大的困难,文献[32]提出上升概率推理使 MLN 的推理能有效实现,而文献[33]提出可处理马尔可夫逻辑 (Tractable Markov Logic, TML),当图模型具有高树宽时 TML 仍能有效推理.

2.3 高斯图模型的非参数化

经典概率图模型 BN 和 MN 只能处理离散型数据,因此连续型数据情况下需要对 BN 和 MN 的参数模型进行约束,使参数模型服从高斯分布.但这种高斯图模型的参数假设在实际应用中过于苛刻,不能真实地反应实际场景.一种常用的解决方法是把高斯图模型进行非参数化,如使用核密度估计.如何有效构造这种非参数化高斯图模型,成为研究人员关注的热点.

Meinshausen 和 Buhlmann 提出了一种有效估计高斯图模型的方法,使用套索 (Lasso) 技术保证变量选择过程的正确性^[34]. Liu 等人提出一种非参数算法来估计高斯图模型的结构^[35],该算法称为基于图优化的分类和回归树算法 (Graph-Optimized Classification and Regression Trees, Go-CART).为了得到真实概率分布的一个更好的近似, Lafferty 等人提出两种方法来构建更加灵活的非参数高斯图模型^[36],其中一种方法所构建的图模型可使用任意的网络结构,但参数模型是高斯模型的非参数扩展,而另一种方法使用核密度估计参数模型,并限制网络结构为树状或森林状.

2.4 定性概率图模型

在某些实际应用中,由于数据的缺失可能会导致变量间的关系存在不确定性,此时 BN 中严格的数值量化表示形式并不适用,并可能因为只利用有限的领域知识而导致模型过拟合.因此,研究人员提出多种定性建模与定性推理的方法,试图根据不同的应用场景来定性描述变量间的关系. Wellman 提出的定性概率网 (Qualitative Probabilistic Network, QPN) 是 BN 的一种定性概念形式^[37],QPN 的表达能力低于 BN,但是却简化了表示模型以及加速推理过程,在应用上具有更高的效率.

为了解决由 BN 导出 QPN 的过程中很有可能由于知识的不确定性而产生有向边的不确定性关系, Bolt 等人在 QPN 的框架中融入情景符号 (situational signs) 的概念^[38],基于情景符号获取网络当前状态的影响信息.而 Renooij 等人基于上下文的信息确定定性影响的符号,在推理时预先排除产生歧义的结果^[39].

由于 QPN 缺乏对定性符号权重的描述,使得不能以量化机制来解决推理过程中发生的不一致冲突问题,为此 Liu 和 Wellman 提出增量式方法,从所对应的 BN 中获取数值信息消除冲突^[40]. Renooij 等人通过辨识影响的强弱程度对 QPN 进行扩展,基于“强影响对结

果的作用大于弱影响”的思想来消除推理冲突问题^[41].他们还提出半定性概率网 (Semi-QPN, SQPN) 的概念^[42],每个定性影响的强度使用一个概率区间值描述,从而对定性影响进行量化.文献[43]则基于粗糙集理论,把变量间的依赖度作为 QPN 每一条有向边的定性影响的权重,并使用权重传播的推理方法构造模型.

2.5 高斯图模型与矩阵型数据的结合

随着信息技术的飞速发展,实际应用中所面临的数据大多数是高维数据.矩阵型变量是处理这些带有重要结构信息的高维数据的一个重要工具,而高斯图模型能有效地为随机变量集之间的条件独立关系建模.因此,高斯图模型与矩阵变量数据的结合有力地推进了高维结构化数据处理的发展,这两者的结合产生的模型称为矩阵正态图模型 (Matrix Normal Graphical Model, MNGM).

但是目前 MNGM 的模型选择和估计方法还不成熟,特别是在高维情况中. Wang 和 West 给出了 MNGM 的贝叶斯分析,提出利用 MCMC 抽样方法来进行推理,并把 MNGM 模型扩展到矩阵时间序列上^[44]. Allen 和 Tibshirani 提出惩罚似然方法来进行 MNGM 的模型估计,在协方差矩阵逆目标函数中引入 L1 范数和 L2 范数惩罚函数^[45]. Yin 和 Li 提出了 L1 范数惩罚似然方法和坐标下降方法来选择和估计 MNGM 的模型^[46].除了使用矩阵范数之外,研究人员还提出使用数组范数来处理高维数据^[47]. Allen 针对数组范数分布提出 L1 罚估计方法构造 MNGM^[48].

2.6 概率图模型与 Copula 函数的结合

PGM 主要应用于多变量场景中,但是在其学习和推理过程中多变量分布函数的估计是极具挑战性的任务. Copula 函数描述的是变量间的相关性,通过利用边缘分布函数之间的 copula 函数 C ,可以把一个多变量分布函数分解为单变量或二变量边缘分布函数的组合. PGM 与 Copula 函数的结合既能利用图结构的独立性简化高维 Copula 函数的表示,又能通过 Copula 函数构造多变量分布函数简化 PGM 的学习和推理过程.因此,近年来,PGM 与 Copula 函数的结合得到了研究人员的关注.

文献[49]和文献[50]最早提出 Copula 函数与图结构独立性结合的思想,但是他们只局限于使用 Pair-Copula 构造 (Pair-Copula Construction, PCC) 概率图模型. Bauer 等人则使用 PCC 来构造 DAG 的非高斯概率分布^[51]. Dobra 和 Lenkoski 提出 Copula 高斯图模型,只对与观测变量有一一对应关系的隐变量作高斯正态分布假设,其图模型选择由半参数高斯 Copula 函数实现^[52]. Elidan 提出 Copula 贝叶斯网 (Copula Bayesian Network, CBN)^[53],对节点的条件概率分布进行基于 Copula 的参

数化,并与隐含独立性关系的图结构结合,实现了高维概率分布建模的灵活性,同时又保留了 Copula 对单变量边缘分布的控制.随后,他还针对非线性 CBN 提出一种快速结构学习方法^[54],并且指出 CBN 在数据缺失的情况下依然具有显著的计算优势^[55]. Elidan 还给出 Copula 网络分类器 (Copula Network Classifier) 用以解决连续解释变量领域的复杂分类问题^[56].

2.7 混合图模型与图模型的混合

混合图模型 (Mixed Graphical Model) 指模型中同时包含离散变量和连续变量.在只有连续变量的情况中,通常假设数据服从高斯分布.而在只有离散变量的情况中,经典的 BN 和两两关系 MN 都可适用.模型同时存在离散和连续变量则为建模问题带来了挑战性. Lauritzen 首先提出使用条件高斯分布为混合数据建模^[57],这一理论成为了混合图模型研究的奠基石.但是,使用条件高斯分布建模时,模型参数的个数随着变量个数的增加而指数增长.文献[58]利用随机森林与稳定性选择理论的结合,解决了混合图模型在高维情况中的估计问题. Lee 和 Hastie 提出一种新两两关系模型 (Pairwise Model),针对离散和连续变量的条件概率分布分别使用多类逻辑斯蒂回归和高斯线性回归,并给出有效的结构学习算法^[59]. Cheng 等人对条件高斯分布进行简化,显著减少模型参数的个数,并对底层网络结构作出稀疏性假设用以处理高维情况的建模^[60].

图模型的混合 (Mixture of Graphical Models) 指图模型中考虑隐变量对观测变量的影响,使得图模型融合了特定上下文依赖性 (Context-Specific Dependencies),观测变量之间的结构关系可随着隐变量的变化而改变.目前,图模型混合的研究主要针对树混合, Kumar 和 Koller 考虑使用基于期望最大化的方法进行树混合的学习^[61],而 Thiesson 等人对该方法进行扩展讨论了 DAG 图模型的混合^[62]. Mossel 和 Roch 提出隐树混合,该模型可认为是乘积分布的分层混合^[63]. Anandkumar 等人提出离散图模型混合的无监督估计算法,可适用于高维情况^[64].

2.8 不同数据类型下的概率图模型

在实际应用领域中常常出现各种不同类型的数数据,最简单的有离散数据和连续数据.数据类型的不同也影响着 PGM 表示理论的发展.

分类数据是反映事物类别的统计数据,高维相关多变量分类数据的分析是机器学习领域的重要问题. Khan 等人利用隐高斯图模型分析分类数据,并给出一种新的 Stick-Breaking 似然函数,能有效发现数据间的相关性^[65].

多媒体数据同时融合了文本、图像、图形和声音等信息,通常希望使用低维的隐式表达方式来概括这些

高维数据的特征. 为了处理多媒体数据, Xing 等人提出了基于双层无向图模型的 Muti-wing Harmonium 模型, 该模型是有向 Latent Dirichlet Allocation 模型的一种无向化对应实现, 但在多媒体数据推理和学习任务中却更加有效^[66].

文档、互联网、电子表格和数据库中经常以表格的形式编码大量信息, 这些信息的综合或者搜索, 对理解信息的含义并显式表示这些数据非常重要. Mulwad 等人提出一种新方法可在不同领域中解释表格数据的含义并把含义表示为关联数据 (Linked Data), 该方法的核心为图模型和概率推理技术^[67].

3 概率图模型的研究方向

(1) 概率图模型的表示应该更加准确地反映变量集上真实的概率分布和变量间结构关系, 而同时概率图模型表示又要具有能模块化表示的概率分布和简单易懂的网络结构图, 并且能进行有效的推理和学习过程. 下一步开发新 PGM 时应尽量满足上述条件, 或者在构造现有 PGM 时选择满足上述条件的模型. 但是, 构造过程的候选模型结构和模型参数个数通常很多, 最优模型的选择十分重要, 而且模型结构和模型参数是相互依赖的, 借助各种模型选择理论实现同时最优选择模型结构和模型参数是十分重要的研究课题^[68].

(2) BN 和 MN 的分类主要是为了研究和学习的便利. 而实际应用中所使用的模型有可能是 BN 和 MN 的某种形式的结合. 如何结合两种模型的表示能力构造混合概率图模型, 如何判断 BN 和 MN 的结合模型的有效性, 制定出适当的组合原则和评判标准, 是一个值得探讨的问题, 这种混合概率图模型同时也给模型的推理和学习带来新的挑战.

(3) 虽然概率图模型与逻辑语言已成功结合, 并在某些领域得到应用, 但概率逻辑语言仍有待进一步研究. 需要开发更丰富的概率逻辑语言, 考虑概率模型与不同的逻辑语言的结合, 如时态逻辑^[69]和模态逻辑^[70]语言的结合等; 深入研究当前概率逻辑模型的有效推理与学习算法, 使推理与学习任务进一步简化并保证准确性; 提高概率逻辑模型的可扩展性, 从而应用在更多的概率图模型推理学习场景中.

(4) 概率图模型的非参数化具有很强的实际意义, 放松了模型对参数假设的约束, 在某些情况下能更好地反映实际场景. 除了前面介绍过的核密度估计非参数化方法外, 还可考虑通过无限混合模型^[71]或者狄氏过程^[72]来进行概率图模型的非参数化.

(5) 目前的概率图模型只集中于局部表示 (Local Representation) 和局部稀疏表示的研究, 而对分布表示 (Distributed Representation)^[73,74] 和稀疏分布表示

(Sparse Distributed Representation)^[75-77] 这两种重要的表示形式上的概率图模型的研究还很少, 这方面的研究可借鉴深度学习理论.

(6) 概率图模型智能表示随机型的不确定性, 不能表示模糊型的不确定性, 因此应用范围仍受到较大的限制, 特别是在人工智能领域, 如何使概率图模型能够处理模糊信息问题是一个值得关注的方向.

4 结论

本文系统综述了 PGM 的其它表示理论, 包括暂态建模的新发展、PGM 与逻辑语言的结合、PGM 非参数化理论、定性概率图模型、矩阵型变量上的概率图模型、与 Copula 函数的结合、混合图模型以及不同数据类型下的概率图模型. 最后指出了 PGM 的未来研究方向.

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. 虽然概率图模型已经取得了很好的应用效果, 但概率图模型表示理论还应时刻结合人工智能和机器学习领域出现的新理论和新技术, 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] Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference [M]. San Mateo, California: Morgan Kaufmann Publishers, 1988.
- [2] Koller D, Friedman N. Probabilistic Graphical Models: Principles and Techniques [M]. MIT Press, Cambridge, 2009.
- [3] Edwards D. Introduction to Graphical Modeling [M]. Springer, New York, USA, 2000.
- [4] Christopher J P, Jerod J W, Lam C T, Daniel S. On learning conditional random fields for stereo: exploring model structures and approximate inference [J]. International Journal of Computer Vision, 2012, 99(3): 319 - 337.
- [5] Golumbic M C, Maffray F, Morel G. A characterization of chain probe graphs [J]. Annals of Operation Research, 2011, 188(1): 175 - 183.
- [6] Murphy K P. Dynamic bayesian networks: Representation, inference and learning [D]. Berkeley: University of California, Computer Science Division, 2002.
- [7] Getoor L, Taskar B. Introduction to Statistical Relational Learning [M]. MIT Press, Cambridge, USA, 2007.
- [8] Nodelman U, Shelton C R, Koller D. Continuous time Bayesian networks [A]. Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence [C]. Edmonton, Alberta, Canada: UAI, 2002. 378 - 387.
- [9] Nodelman U, Koller D, Shelton C R. Expectation propagation for continuous time bayesian networks [A]. Proceedings of the 21st Conference on Uncertainty in Artificial In-

- telligence [C]. Edinburgh, Scotland, UK: UAI, 2005. 431 – 440.
- [10] Nodelman U, Koller D, Shelton C R. Expectation maximization and complex duration distributions for continuous time bayesian networks [A]. Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence [C]. Edinburgh, Scotland, UK: UAI, 2005. 421 – 430.
- [11] Cohn I, El-Hay Tal, Friedman N, Kupferman R. Mean field variational approximation for continuous-time bayesian networks [J]. Journal of Machine Learning Research, 2010, 11 (10): 2745 – 2783.
- [12] El-Hay Tal, Friedman N, Koller D, Kupferman R. Continuous time markov networks [A]. Proceedings of the 22nd Conference on Uncertainty in Artificial Intelligence [C]. Cambridge, MA, USA: UAI, 2006. 155 – 164.
- [13] Robinson J W, Hartemink A J. Non-stationary dynamic Bayesian networks [A]. Proceedings of the 21 Annual Conference on Neural Information Processing Systems [C]. Vancouver, British Columbia, Canada: NIPS, 2008. 21: 1369 – 1376.
- [14] Grzegorzczak M, Husmeier D. Non-stationary continuous dynamic Bayesian networks [A]. Proceedings of the Neural Information Processing Systems 22; 23rd Annual Conference on Neural Information Processing Systems [C]. Vancouver, British Columbia, Canada: NIPS, 2009. 22: 682 – 690.
- [15] Lèbre S. Stochastic process analysis for genomics and dynamic Bayesian networks inference [R]. Ph. D. thesis, Université d'Évry-Val d'Essonne, France, 2007.
- [16] Ahmed A, Xing E P. Recovering time-varying networks of dependencies in social and biological studies [J]. PANS, 2009, 106 (29): 11878 – 11883.
- [17] Kolar M, Song L, Xing E. Sparsistent learning of varying-coefficient models with structural changes [A]. Proceedings of the Neural Information Processing Systems 22; 23rd Annual Conference on Neural Information Processing Systems [C]. Vancouver, British Columbia, Canada: NIPS, 2009. 22: 1006 – 1014.
- [18] Grzegorzczak M, Husmeier D. Non-homogeneous dynamic Bayesian networks for continuous data [J]. Machine Learning, 2011, 83 (3): 355 – 419.
- [19] Dondelinger F, Lèbre S, Husmeier D. Non-homogeneous dynamic Bayesian networks with Bayesian regularization for inferring gene regulatory networks with gradually time-varying structure [J]. Machine Learning, 2013, 90 (2): 191 – 230.
- [20] Grzegorzczak M, Husmeier D. Regularization of non-homogeneous dynamic Bayesian networks with global information-coupling based on hierarchical Bayesian models [J]. Machine Learning, 2013, 91 (1): 105 – 154.
- [21] Kersting K, De Raedt L. Bayesian logic programs [R]. Technical Report 151, University of Freiburg, Institute for Computer Science, April 2001.
- [22] Kersting K, De Raedt L. Bayesian Logic Programming: Theory and Tool [M]. MIT Press, Cambridge, USA, 2005.
- [23] Kersting K, De Raedt L. Adaptive bayesian logic programs [A]. Proceedings of the 11th International Conference Inductive Logic Programming [C]. Strasbourg, France: ILP, 2001. 104 – 117.
- [24] Raghavan S, Mooney R. Bayesian abductive logic programs [A]. Proceedings of the Statistical Relational Artificial Intelligence [C]. Atlanta, Georgia, USA: AAAI workshop, 2010. 629 – 644.
- [25] Laskey K B. MEBN: A language for first-order Bayesian knowledge bases [J]. Artificial Intelligence, 2008, 172 (2 – 3): 140 – 178.
- [26] Richardson M, Domingos P. Markov logic networks [J]. Machine Learning, 2006, 62 (1 – 2): 107 – 136.
- [27] Singla P, Domingos P. Markov Logic in Infinite Domains [A]. Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence [C]. Vancouver, BC, Canada: UAI Press, 2007. 368 – 375.
- [28] Kok S, Domingos P. Statistical predicate invention [A]. Proceedings of the 24th International Conference [C]. Corvallis, Oregon, USA: ICML, 2007. 433 – 440.
- [29] Jain D, Barthels A, Beetz M. Adaptive markov logic networks: Learning statistical relational models with dynamic parameters [A]. Proceedings of the 9th European Conference on Artificial Intelligence [C]. Lisbon, Portugal: ECAI, 2010. 937 – 942.
- [30] Blyth J, Hobbs J R, Domingos P, et al. Implementing weighted abduction in markov logic [A]. Proceedings of the 9th International Conference on Computational Semantics [C]. Stroudsburg, PA, USA: Association for computational linguistics, 2011. 55 – 64.
- [31] Singla P, Mooney R J. Abductive markov logic for plan recognition [A]. Proceedings of the 25th AAAI Conference on Artificial Intelligence [C]. San Francisco, California USA: AAAI, 2011. 1069 – 1075.
- [32] Gogate V, Domingos P. Probabilistic theorem proving [A]. Proceedings of the 27th Conference on Uncertainty in Artificial Intelligence [C]. Barcelona, Spain: UAI, 2011. 256 – 265.
- [33] Domingos P, Webb W A. A tractable first-order probabilistic logic [A]. Proceedings of the 26th AAAI Conference on Artificial Intelligence [C]. Toronto, Ontario, Canada: AAAI, 2012. 1902 – 1909.
- [34] Meinshausen N, Buhlmann P. High-dimensional graphs

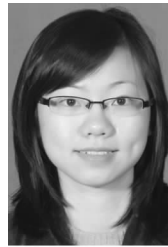
- and variable selection with the lasso[J]. *Annals of Statistics*, 2006, 34(3): 1436 – 1462.
- [35] Liu H, Chen X, Lafferty J, Wasserman L. Graph-valued regression[A]. *Proceedings of the 24th Annual Conference on Neural Information Processing Systems* [C]. Vancouver, British Columbia, Canada: NIPS, 2010. 1423 – 1431.
- [36] Lafferty J, Liu H, Wasserman L. Sparse nonparametric graphical models[J]. *Institute of Mathematical Statistics in Statistical Science*, 2012, 27(4): 519 – 537.
- [37] Wellman M P. Fundamental concepts of qualitative probabilistic networks[J]. *Artificial Intelligence*, 1990, 44(3): 257 – 303.
- [38] Bolt JH, Van der Gaag LC, Renooij S. Introducing situational signs in qualitative probabilistic networks[J]. *International Journal of Approximate Reasoning*, 2005, 38(3): 333 – 354.
- [39] Renooij S, Van der Gaag LC, Parsons S. Context-specific sign-propagation in qualitative probabilistic networks[J]. *Artificial Intelligence*, 2002, 144(1): 207 – 230.
- [40] Liu C L, Wellman M P. Incremental tradeoff resolution in qualitative probabilistic networks[A]. *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence* [C]. Madison, Wisconsin, USA; UAI, 1998. 338 – 345.
- [41] Renooij S, Van der Gaag LC. Enhanced qualitative probabilistic networks for resolving trade-offs[J]. *Artificial Intelligence*, 2008, 172(12 – 13): 1470 – 1494.
- [42] Renooij S, van der Gaag L C. From qualitative to quantitative probabilistic networks[A]. *Proceedings of the 18th Conference in Uncertainty in Artificial Intelligence* [C]. Edmonton, Alberta, Canada; UAI, 2002. 422 – 429.
- [43] Yue K, Liu W Y. Qualitative probabilistic networks with rough-set-based weights[A]. *IEEE International Conference on Machine Learning and Cybernetics* [C]. Kunming, China; IEEE, 2008. 3: 1768 – 1774.
- [44] Wang H, West M. Bayesian analysis of matrix normal graphical models [J]. *Biometrika*, 2009, 96(4): 821 – 834.
- [45] Allen G, Tibshirani R. Transposable regularized covariance models with an application to missing data imputation [J]. *The Annals of Applied Statistics*, 2010, 4(2): 764 – 790.
- [46] Yin J, Li H. Model selection and estimation in the matrix normal graphical model[J]. *Journal of Multivariate Analysis*, 2012, 107(5): 119 – 140.
- [47] Hoff P. Separable covariance arrays via the Tucker product, with applications to multivariate relational data[J]. *Bayesian Analysis*, 2011, 6(2): 179 – 196.
- [48] Allen G. Comment on article by Hoff[J]. *Bayesian Analysis*, 2011, 6(2): 197 – 202.
- [49] Hanea A M, Kurowicka D, Cooke R M. Hybrid method for quantifying and analyzing bayesian belief nets[J]. *Quality and Reliability Engineering*, 2006, 22(6): 709 – 729.
- [50] Kurowicka D, Cooke R. Uncertainty Analysis with High Dimensional Dependence Modelling [M]. John Wiley & Sons, Chichester, 2006.
- [51] Bauer A, Czado C, Klein T. Pair-copula constructions for non-Gaussian DAG models[J]. *Canadian Journal of Statistics*, 2012, 40(1): 86 – 109.
- [52] Dobra A, Lenkoski A. Copula Gaussian graphical models and their application to modeling functional disability data [J]. *The Annals of Applied Statistics*, 2011, 5(2A): 969 – 993.
- [53] Elidan G. Copula bayesian networks[A]. *Processing of the 23; 24rd Annual Conference on Neural Information Processing Systems* [C]. Vancouver, British Columbia, Canada; NIPS, 2010. 559 – 567.
- [54] Elidan G. Lightning-speed structure learning of nonlinear continuous networks[A]. *Proceedings of the 5th International Conference on Artificial Intelligence and Statistics* [C]. La Palma, Canary Islands; AISTATS, 2012. 355 – 363.
- [55] Elidan G. Inference-less density estimation using Copula Bayesian Networks[A]. *Proceedings of the 26th Conference on Uncertainty in Artificial Intelligence* [C]. Catalina Island, CA, USA; UAI, 2010. 151 – 159.
- [56] Elidan G. Copula network classifiers (CNCs) [A]. *Proceedings of the 5th International Conference on Artificial Intelligence and Statistics* [C]. La Palma, Canary Islands; AISTATS, 2012. 346 – 354.
- [57] Lauritzen S. *Graphical Models* [M]. Oxford UK: Oxford University Press. 1996.
- [58] Fellinghauer B, Bühlmann P, Ryffel M, et al. Stable graphical model estimation with Random Forests for discrete, continuous, and mixed variables[J]. *Computational Statistics & Data Analysis*, 2013, 64(8): 132 – 152.
- [59] Lee J D, Hastie T J. Structure learning of mixed graphical models[A]. *Proceedings of the 6th International Conference on Artificial Intelligence and Statistics* [C]. Scottsdale, AZ, USA; AISTATS, 2013, 388 – 396.
- [60] Edwards D, Gabriel C. G. de Abreu. Labouriau R. Selecting high-dimensional mixed graphical models using minimal AIC or BIC forests[J]. *BMC Bioinformatics* 2010, 11(1): 11 – 18.
- [61] Kumar M P, Koller D. Learning a small mixture of trees [A]. *Proceedings of the Neural Information Processing Systems 22; 23rd Annual Conference on Neural Information Processing Systems* [C]. Vancouver, British Columbia, Canada; NIPS, 2009. 22: 1051 – 1059.

- [62] Thiesson B, Meek C, Chickering D, Heckerman D. Computationally efficient methods for selecting among mixtures of graphical models[J]. *Bayesian Statistics*, 1999, 6(1):569–576.
- [63] Mossel E, Roch S. Phylogenetic mixtures: concentration of measure in the large-tree limit[J]. *The Annals of Applied Probability*, 2012, 22(6):2429–2459.
- [64] Anandkumar A, Hsu D, Huang F, et al. Learning mixtures of tree graphical models[A]. *Processing of the 25th Annual Conference on Neural Information Processing Systems* [C]. Lake Tahoe, Nevada, USA; NIPS, 2012. 1061–1069.
- [65] Khan M E, Mohamed S, Marlin B M, et al. A stick-breaking likelihood for categorical data analysis with latent Gaussian models[A]. *Proceedings of the 5th International Conference on Artificial Intelligence and Statistics* [C]. La Palma, Canary Islands; AISTATS, 2012. 610–618.
- [66] Xing E P, Yan R, Hauptmann A G. Mining associated text and images with dual-wing harmoniums[A]. *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence* [C]. Edinburgh, Scotland, UK; UAI, 2005. 633–641.
- [67] Mulwad V, Finin T, Joshi A. *Search Computing* [M]. Springer Berlin Heidelberg, 2012, 7538:16–33.
- [68] P M Lukacs, K P Burnham, D R Anderson. Model selection bias and Freedman's paradox[J]. *Annals of the Institute of Statistical Mathematics*, 2010, 62(1):117–125.
- [69] M Reynolds. The complexity of temporal logic over the reals[J]. *Annals of Pure and Applied Logic*, 2010, 161(8):1063–1096.
- [70] Ditmarscha H V, van der Hoek. W, Kooi B. Local properties in modal logic[J]. *Artificial Intelligence*, 2012, 187–188(8):133–155.
- [71] Gershmana S J, Blei D M. A tutorial on Bayesian nonparametric models[J]. *Journal of Mathematical Psychology*, 2012, 56(1):1–12.
- [72] Tayal, Poupart P, Li Y. Hierarchical double dirichlet process mixture of gaussian processes[A]. *Proceedings of the 26th AAAI Conference on Artificial Intelligence* [C]. Toronto, Ontario, Canada; AAAI, 2012. 1126–1133.
- [73] Bengio Y, Ducharme R, Vincent P. A neural probabilistic language model[A]. *Processing of the 14th Annual Conference on Neural Information Processing Systems* [C]. Denver, Co, USA; NIPS, 2001. 933–938.
- [74] Hinton G E. Learning distributed representations of concepts using linear relational embedding[J]. *IEEE Trans. Knowl. Data Eng.*, 2001, 13(2):232–244.
- [75] Bagnell J A, Bradley D M. Differentiable sparse coding[A]. *Proceedings of the Neural Information Processing Systems 22; 23rd Annual Conference on Neural Information Processing Systems* [C]. Vancouver, British Columbia, Canada; NIPS, 2008. 113–120.
- [76] Coates A, NG A Y. The importance of encoding versus training with sparse coding and vector quantization[A]. *Proceedings of the 24; 25th Annual Conference on Neural Information Processing Systems* [C]. Bellevue, Washington, USA; ICML, 2011. 921–928.
- [77] Goodfellow I J, Courville A C, Bengio Y. Large-scale feature learning with spike-and-slab sparse coding[A]. *Proceedings of the 29th International Conference on Machine Learning* [C]. Edinburgh, Scotland, UK; ICML, 2012. 6407–6423.

作者简介



刘建伟 男, 1966 年出生. 博士, 中国石油大学(北京)副研究员, 主要研究方向包括智能信息处理, 机器学习, 算法分析与设计等.
E-mail: liujw@cup.edu.cn



黎海恩 女, 1988 年出生. 硕士, 2014 年毕业于中国石油大学(北京)自动化系. 现在任职山东新华医药化工设计有限公司.
E-mail: lihaien1988@163.com