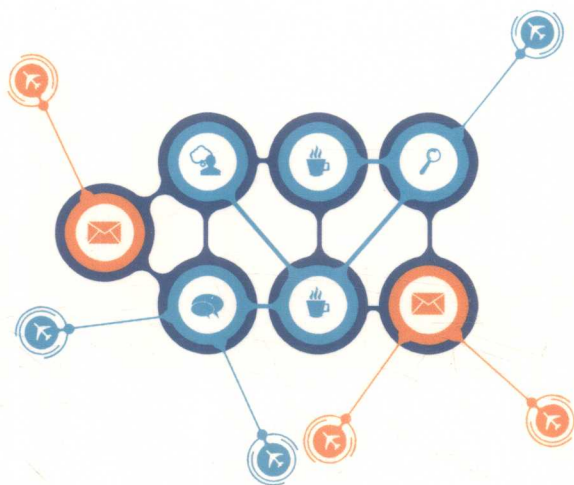
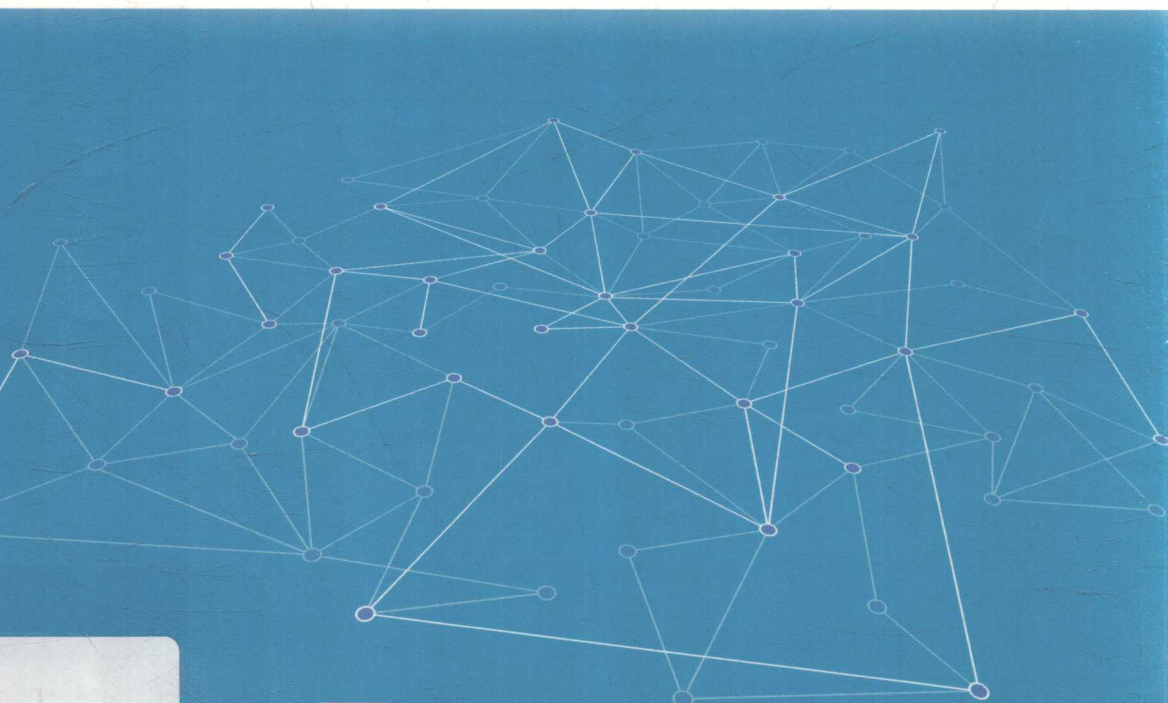




贝叶斯网络学习 及数据分类

李艳颖 著



科学出版社

贝叶斯网络学习及数据分类

李艳颖 著

陕西省教育厅专项科研计划项目 (18JK0045) 资助

宝鸡市技术创新引导计划项目 (2019H-2) 资助

宝鸡文理学院重点学科建设经费资助



科学出版社

北京

内 容 简 介

本书针对大数据网络的结构学习,提出基于条件独立测试的学习贝叶斯网络框架的算法,研究了马尔可夫等价祖先图的共性,提出最大祖先图的最小本质图的概念,还研究了马尔可夫等价的有向最大祖先图的转化问题,并针对数据分类问题,构建一个贝叶斯网络分类器模型,提出的算法精度高,分类效果好。同时,本书还研究了 k -最近邻分类器的 k 值选择问题,建立了快速寻找最优 k 值的搜索策略,并保证模型具有最高留一分类精度。

本书可作为相关科研工作者学习、研究的参考用书。

图书在版编目(CIP)数据

贝叶斯网络学习及数据分类/李艳颖著. —北京:科学出版社,2018.11
ISBN 978-7-03-058931-6

I. ①贝… II. ①李… III. ①贝叶斯推断-研究 IV. ①O212

中国版本图书馆 CIP 数据核字 (2018) 第 219408 号

责任编辑:宋 丽 王会明 / 责任校对:王万红
责任印制:吕春珉 / 封面设计:东方人华平面设计部

科学出版社出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

北京虎彩文化传播有限公司印刷

科学出版社发行 各地新华书店经销

*

2018年11月第 一 版 开本:B5(720×1000)

2018年11月第一次印刷 印张:8 3/4

字数:180 000

定价:70.00 元

(如有印装质量问题,我社负责调换〈虎彩〉)

销售部电话 010-62136230 编辑部电话 010-62135397-2008

版权所有,侵权必究

举报电话:010-64030229; 010-64034315; 13501151303

前 言

随着信息技术的迅猛发展,如何从中发现并揭示变量间的依赖关系显得尤为重要,贝叶斯网络是将图论与概率论有机结合、应用广泛的一类概率图模型,为数据分类、推理和预测等问题提供了重要的理论基础和有效工具。贝叶斯网络的结构学习是其主要研究内容之一,也是非确定多项式(non-deterministic polynomial, NP)难问题。本书针对该问题,提出了学习高维网络结构的混合算法,探讨了含隐变量和选择变量的贝叶斯网络的本质图及等价转化问题。并针对数据分类问题,构建了一个贝叶斯网络分类器模型,同时研究了 k -最近邻分类器的 k 值选择问题。本书主要内容如下。

首先针对大数据网络的结构学习,提出基于条件独立测试的学习贝叶斯网络框架的算法。高阶条件独立测试具有不可靠性和运算费用高的特点,因此大网络结构学习的主要难点在于如何减少高阶条件独立测试的次数及如何尽可能地降低测试的阶数。本书所提算法采用低阶测试学习目标变量的候选父子节点集,利用候选父子节点集学习目标变量的候选配偶节点集,再结合一个快速的学习父子节点集的算法进一步缩小父子节点集的规模;本书还利用一个补偿机制将那些因过于严格的筛选条件而被删除的父子节点恢复。在分层筛选目标变量父子节点集的过程中,通过不断缩小条件集,有效地提高了高维网络结构学习的精度。

其次研究了马尔可夫等价祖先图的共性,提出了最大祖先图的最小本质图的概念。由于祖先图的空间是随着变量个数的增加而指数级增加的,因此讨论等价类的共性,并用共性刻画等价类,不但可以更加深刻地理解祖先图的本质,而且可大大降低祖先图空间的维数,从而为祖先图结构学习做出有意义的指导。本书基于最小汇连路,给出了学习最小本质图的算法,并将其应用到快速判断祖先图等价问题上,进一步给出了一组定向规则尽可能多的定向本质标记。

然后研究了马尔可夫等价的有向最大祖先图的转化问题,这种等价转化可以使我们从给定的有向最大祖先图出发,通过对其上的一些非紧边做反转或添加箭头或去掉箭头的变化,遍历该最大祖先图所在的等价类。本书着重研究了双向边向有向边转化的条件,再结合已有的研究工作,得到有向边转化为双向边、有向边反转和双向边转化为有向边的充分条件。这些结果使得等价有向最大祖先图之间可以流畅地相互转化,便于提取马尔可夫等价类共性和识别本质标记。

最后针对数据分类问题,本书构建了一个贝叶斯网络分类器模型,并研究了 k -最近邻分类器的 k 值选择问题。贝叶斯网络分类器模型是基于粗糙集理论和进化算法构建的,通过与其他8种常见的分类器相比,实验结果表明该算法设计合理,

且分类效果好。对于 k -最近邻分类器，参数 k 是它唯一的参数并密切影响着算法的效果，因此 k 值的选择问题十分重要。本书采用几乎无偏的留一法来评估 k -最近邻分类器的分类性能，利用留一分类精度和 k 值之间的近似凹的关系，建立了快速寻找最优 k 值的搜索策略。该策略可以快速收敛到最优 k 值，并保证模型有最高的留一分类精度。

由于作者水平有限，书中不妥之处在所难免，恳请专家和广大读者批评指正。

作 者

2018 年 9 月

符号对照表

符号	符号名称
X	随机变量或图中节点
$\mathcal{X}$	随机变量集
G	有向无环图、祖先图
V	有向无环图节点全集
E	边集
P	概率分布
$I_G(X \perp Y Z)$	用 Z 来 d-分割 X 和 Y
$I_P(X \perp Y Z)$	给定 Z 时 X 和 Y 条件独立
x^T	向量 x 的转置
D	数据集
$\max$	取最大
$\min$	取最小
argmax	最大值点集合
$\forall$	任意的
$\in$	属于
$\subset$	包含于
$\subsetneq$	真包含于

部分缩略语对照表

缩略语	英文全称	中文全称
BN	Bayesian network	贝叶斯网络
DAG	directed acyclic graph	有向无环图
CPT	conditional probability table	条件概率表
AI	artificial intelligence	人工智能
HPC	hybrid parents and children	混合学习父子节点集
CH	cooper-herskovits function	CH 评分函数
BIC	Bayesian information criterion	BIC 评分
AIC	Akaike information criterion	AIC 评分
MDL	minimum description length	最小描述长度
MAP	maximum a posterior hypothesis	最大后验假设
MPE	most probable explanation	最大可能解释
NP-hard	nondeterministic polynomial-hard	NP 难问题
MAG	maximal ancestral graph	最大祖先图
GBNC	general Bayesian network classifier	贝叶斯网络分类器
NBC	naive Bayesian classifier	朴素贝叶斯分类器
EHPC	efficient hybrid parents and children	高效混合学习父子节点集
LOO	leave one out	留一法
LOO-O- k	optimal k based on leave one out	基于留一法的最优 k 值
MEG	minimal essential graph	最小本质图
MCP	minimal collider path	最小汇连路

目 录

第 1 章 绪论	1
1.1 研究背景与意义	1
1.2 贝叶斯网络的发展现状	2
1.2.1 贝叶斯网络学习	2
1.2.2 贝叶斯网络推理	6
1.2.3 含隐变量和选择变量的贝叶斯网络	8
1.3 数据分类问题	9
1.4 本书的主要内容和结构安排	10
第 2 章 贝叶斯网络和祖先图的基础知识	13
2.1 贝叶斯网络基本知识	13
2.1.1 图论知识和概率论知识	13
2.1.2 贝叶斯网络	16
2.2 祖先图基本知识	20
2.2.1 考虑隐变量和选择偏倚的原因	20
2.2.2 祖先图	21
2.2.3 学习隐结构的算法	23
2.3 本章小结	25
第 3 章 基于约束学习贝叶斯网络框架	26
3.1 引言	26
3.2 理论基础	28
3.3 EHPC 算法	28
3.3.1 学习父子节点集的超集	29
3.3.2 学习配偶节点集的超集	32
3.3.3 子程序 Inter-IAPC 和主程序 EHPC	34
3.4 仿真实验	36
3.5 本章小结	41

第 4 章 祖先图的本质图	42
4.1 祖先图的等价类	42
4.2 (最小) 本质图	45
4.3 学习本质图框架	46
4.3.1 算法流程	46
4.3.2 理论证明	49
4.4 实验结果	52
4.5 本章小结	57
第 5 章 等价祖先图的转换	58
5.1 引言	58
5.2 基础知识	59
5.3 转换流程	59
5.4 本章小结	65
第 6 章 数据分类的基础知识	67
6.1 数据分类	67
6.1.1 数据分类的意义	67
6.1.2 常见数据类型	68
6.1.3 数据分类的基本原则	70
6.2 数据分类经典算法	71
6.2.1 决策树分类算法	72
6.2.2 支持向量机分类算法	73
6.2.3 基于关联规则分类算法	75
6.2.4 基于数据库技术的分类算法	76
6.2.5 贝叶斯分类算法	77
6.2.6 数据分类的应用	81
6.3 本章小结	81
第 7 章 贝叶斯网络分类器	82
7.1 粗糙集理论	82
7.1.1 信息系统	83
7.1.2 属性约简	84
7.2 ARE-BNC 算法	84
7.2.1 生成初始种群	85

7.2.2 进化算法	87
7.3 实验结果	88
7.4 本章小结	92
第 8 章 k -最近邻分类器	93
8.1 k -最近邻分类器的发展	93
8.2 k -最近邻算法基础知识	94
8.2.1 k -最近邻规则	94
8.2.2 k -最近邻算法的基本要素	95
8.2.3 分类精度评估	97
8.2.4 留一法	97
8.3 k -最近邻算法的流程	98
8.3.1 数据预处理和 k 的取值范围	102
8.3.2 选择最优 k 值的近似凹优化框架	103
8.4 数值实验	108
8.4.1 时间及精度对比	109
8.4.2 LOO- $O-k$ 的稳定性	110
8.5 本章小结	114
第 9 章 研究展望	115
参考文献	117

第1章 绪 论

本章首先阐述本书的研究背景和意义,指出在当今大数据时代学习贝叶斯网络的重要性。然后总结贝叶斯网络和数据分类问题的主要研究进展,最后介绍本书的主要研究内容及结构安排。

1.1 研究背景与意义

当今时代,越来越多的科学家相信,不确定性是这个世界的魅力所在,只有不确定性本身才是确定的^[1]。正是在这样的背景下,数据挖掘、机器学习和不确定性人工智能得到了蓬勃发展。近年来涌现了多种处理不确定知识发现的理论和方法。其中具有代表性的有模糊集理论^[2-4]、概率统计理论^[5]、D-S (Dempster/Shafér) 证据理论^[6]和粗糙集理论^[7,8]。随着概率统计学的发展,用图来表示变量间的结构信息越来越受到人们的重视,相关理论也越来越完善。概率图模型将统计模型图形化,从而使传统统计推断方法可以高效地利用变量间的结构信息,这极大地提高了数据分析的质量和效率。因此,图模型目前已成为探讨随机变量间的独立关系和不确定性的一种有力模型,受到许多实际工作者的青睐^[9]。概率图模型是图论和概率论有机结合的模型,从图论角度,它用图形表示随机变量间的结构信息;从概率论角度,它利用图结构和概率规则降低推理的计算复杂度。

贝叶斯网络 (Bayesian network, BN) 是一类应用广泛的概率图模型,也是目前不确定知识表达和推理领域最有效的理论模型之一^[10-13]。贝叶斯网络结构是一个有向无环图 (directed acyclic graph, DAG), 图中节点对应随机变量,有向边表示随机变量间的因果关系或者相关关系。图中每个节点处有一个条件概率表 (conditional probability table, CPT), 定量显示了变量间的关系强度。贝叶斯网络擅长不确定性分析和推断概率性事件。

贝叶斯网络最初主要用于医疗诊断等专家系统,随着其理论的发展,逐渐成为不确定性问题处理的主要方法,在文件分类、基因分析、统计决策、信息检索、学习预测、工业控制、电信网络诊断、工程学、发音特征识别、识别产品故障率、法律、信用评分分析等^[14-18]领域中得到了重要的应用。因此,深入研究贝叶斯网络是十分必要且具有重要价值的。本书在介绍、总结贝叶斯网络的发展现状的基础上,深入探讨、研究贝叶斯网络的学习问题及在数据分类上的应用。

1.2 贝叶斯网络的发展现状

在当今人工智能 (artificial intelligence, AI) 范畴, 美国 UCLA 大学教授朱迪亚·珀尔于 1988 年提出的贝叶斯网络是表达、推理不确定知识的主流方法之一^[19]。时至今日, 贝叶斯网络研究领域主要集中在学习贝叶斯网络、贝叶斯网络推理及含隐变量和选择变量的贝叶斯网络的研究。下面分别介绍这几方面的发展情况。

1.2.1 贝叶斯网络学习

学习参数和结构是贝叶斯网络学习的两个主要分支。学习参数是指在给定贝叶斯网络结构的基础上, 估计各节点处的条件概率。常见的参数学习方法有最大似然估计 (maximum likelihood estimation, MLE) 方法和贝叶斯方法^[20-22]。这两种方法都是基于独立同分布 (independent identify distribution) 前提下的假设。学习结构是指由领域知识丰富的专家构建图结构, 或者在给定观测数据和先验信息的情况下, 利用计算机学习与数据匹配最好的图结构。多数情形下, 学习结构蕴含了学习参数, 因此, 学习贝叶斯网络主要是指学习结构。通过近 30 年诸多研究者的共同努力, 目前根据数据自动学习贝叶斯网络结构已经取得了辉煌的成果。按照数据的完备情形, 我们对学习结构做了如下总结。

1. 完备数据下的结构学习

众所周知, 在完备数据下, 学习结构的算法已较为成熟, 可以大致分为两类。

1) 基于约束的方法 (constraint-based)。这种方法是利用统计知识测试随机变量之间的条件独立关系, 根据独立关系集确定结构。最早的基于约束的方法是 SGS 算法 (Spirtes、Glymour 和 Scheines 提出的算法)^[23], 它是 Spirtes 在 1990 年构建的, SGS 算法无须节点拓扑序, 但需要的条件独立测试次数是指数级的。Spirtes 等于次年修改了 SGS 的搜索机制, 得到经典的 PC (Peter-Clark) 算法^[24], PC 算法使用具有 10000 个样本的数据集进行 Alarm 网学习, 产生 3 个缺失边和 2 个冗余边。在稀疏条件下, PC 算法所需计算量较小, 在复杂网络下, PC 算法的学习精度不高。同期 Verma 等提出了相类似的因果归纳 (inductive causation, IC) 算法^[25]。SGS、PC 和 IC 算法是基于约束方法的 3 个代表。Cheng 等于 1999 年构建了著名的三阶段算法^[26]。算法将互信息和条件独立性测试结合起来, 学习过程分为建立结构草图、厚化图形和薄化图形 3 个阶段, 按有节点顺序和无节点顺序两种情况来学习贝叶斯网络结构, 具有很好的性能和精度^[27]。2008 年, Yehezkel 等

提出了递归自主识别 (recursive autonomy identification, RAI) 算法^[28], RAI 利用递归思想和图分解方法对贝叶斯网络结构进行有效的学习。2010 年, Rodrigues 等提出了杂交亲子 (hybrid parents and children, HPC) 算法^[29], HPC 通过学习父子节点集来学习贝叶斯网络结构, 该算法是一个可扩展的处理大型网络的高效算法。针对利用统计学知识估计贝叶斯网络结构的过程中容易出现的第二类错误, Mahdi 等于 2013 年提出了 LMM 算法^[30]。Li 等于 2015 年基于 PC 算法提出了 FEPC 算法^[31,32], 大大减少了条件独立测试的次数和测试条件集的阶数。随着马尔可夫边界 (Markov boundary, MB) 理论知识的不断完善, 以及马尔可夫边界在最优特征子集的选择、高维数据集爆炸问题和分类器学习等方面的广泛应用, 20 世纪 90 年代后期至今又涌现了一批利用马尔可夫边界学习贝叶斯网络结构的算法。KS 算法 (Koller 和 Sahami 提出的算法)^[33]的出现, 使得局部学习马尔可夫边界方法引起了众多学者的注意。Margaritis 等提出的 Grow-Shrink (GS) 算法^[34]是第一个理论上正确的学习马尔可夫边界的算法。GS 算法分为增厚和削薄两个阶段。增厚阶段从一个空马尔可夫边界出发, 以当前马尔可夫边界 (current Markov boundary, CMB) 为条件集, 如果变量和目标变量条件不独立, 则将变量加入当前马尔可夫边界, 直至所有候选变量都测试过。由于这一过程可能有许多误报节点进入当前马尔可夫边界, 所以接下来对其进行进一步的修正。削薄阶段用当前马尔可夫边界的子集做条件集来删除误报节点。此后众多 GS 算法的变体被提出。Tsamardinos 等于 2003 年提出了 GS 算法的变体——马尔可夫边界增量关联 (incremental association Markov blanket, IAMB) 算法^[35]。同样, IAMB 由一个向前阶段和一个向后阶段组成。该算法用条件互信息评价变量之间的强相关特征 (the strongly relevant features)^[36], 减少了 GS 算法中的一些不必要检测, 因此具备处理高维数据 (high-dimensional datasets) 的能力。虽然 IAMB 是正确的且是可扩展的, 却是低效的。Tsamardinos 等在 2003 年提出的 InterIAMBnPC 算法^[37], Yaramakala 等于 2005 年提出的 Fast-IAMB 算法^[38]都是 IAMB 的变体, 它们有效地提高了算法的速度和有效性。Aliferis 等于 2003 年提出的 HITON 算法^[39]比以上几个马尔可夫边界算法增加了一个阶段。HITON 算法先学习目标变量 T 的父子节点集 (parents and children set) $PC(T)$, 然后学习 $PC(T)$ 中变量的父子节点集 $PC(PC(T))$, 这两部分学习到的 $PC(T) \cup PC(PC(T))$ 即为当前得到的马尔可夫边界, 进一步对当前的马尔可夫边界进行修正, 去掉错误学习的配偶节点。最后用一个贪婪向后消除算法进一步修正马尔可夫边界。虽然 HITON 算法大大提高了学习马尔可夫边界算法的学习效率和速度, 但是却被证明是错误的^[40]。此外还有最大-最小马尔可夫毯 (max-min Markov blanket, MMMB) 算法^[41]、基于父子关系的

马尔可夫边界 (parents and children-based Markov boundary, PCMB) 算法^[40]、广度优先搜索马尔可夫毯 (breadth-first search of Markov blanket, BFMB)^[42]、IPC-MB 算法^[43]、MBOR 算法^[44]和 FEIPC-MB 算法^[45]。Geng 等提出了一种分解贝叶斯网络的理论^[46], 在该理论下, 从一个复杂的贝叶斯网络搜索 v-结构 (v-structures) 可以分解为从网络的子图中搜索 v-结构。这样可以降低测试 v-结构时所用条件集的候选集合的维数。Xie 等将该分解理论运用于贝叶斯网络结构学习, 并提出了基于递归分解的贝叶斯网络结构学习的方法^[47,48]。在文献[48]中, 较大的结构学习的问题先分解为在两个较小的节点子集上的学习, 再将较小子结构递归分解成两个更小的子集的问题, 直到子集不能进一步分解。该方法可以提高搜索的效率和降低统计检验的复杂度。

这类算法的一个共同前提假设是数据样本采样是一个忠实于某贝叶斯网络的概率分布。基于约束方法的优点在于可以处理高维数据, 特别是利用马尔可夫边界, 可以局部地学习网络结构, 降低学习的复杂度; 缺点是高维条件独立测试具有不可靠性, 前期的测试错误会影响后续测试的结果。因此提高基于约束方法效率的关键在于减少条件独立测试的次数和条件集的阶数。

2) 基于评分搜索的方法 (based on score-and-search)。此类方法有两个基本要素: 评分准则和搜索机制。评分准则负责评估候选结构和训练数据的匹配程度。常用的评分函数有 CH 评分函数 (cooper-herskovits function)^[49]、贝叶斯信息准则 (Bayesian information criterion, BIC) 评分函数^[50]、最小描述长度 (minimum description length, MDL) 评分函数^[51-54]、Akaike 信息准则 (Akaike information criterion, AIC) 评分函数^[55]、验证数据似然度评分函数——保留验证可能性 (holdout validation likelihood, HVL) 评分函数与交叉验证可能性 (cross validation likelihood, CVL) 评分函数^[56,57]、最优参数对数似然评分函数 (parameter maximized log-likelihood function)^[58]、贝叶斯评分函数 (Bayesian score function)^[58]等。使最优参数对数似然评分函数达到最优的是完全贝叶斯网络, 因此无论真实模型稀疏还是稠密, 用该评分函数都会得到一个完全有向图^[58]。贝叶斯评分函数考虑了人们的先验知识, 在样本数据和先验分布的条件下, 求后验概率值最大的模型结构。CH 评分函数与贝叶斯评分函数类似, 在假设模型的先验分布是均匀分布的情况下, 贝叶斯评分函数和 CH 评分函数是等价的。BIC 评分函数由优参对数似然函数和惩罚函数组成, 前者评价模型结构和数据的拟合程度, 后者防止网络结构过于复杂, 二者的结合可以避免优参对数似然函数学习结果太复杂的缺点, 学者们多采用该方法, 但是 BIC 评分函数对样本数量有所要求。最小描述长度评分函数也是惩罚函数, 其学习目标是寻找模型的描述长度和数据的编码长度总和最

小的网络结构,因此也是一个选择相对简单网络结构的函数,达到算法复杂度和结构精度之间的均衡。AIC 评分函数与 BIC 评分函数类似,也是优参数对数似然函数加一个惩罚函数^[59]。当样本数量趋于无穷时,AIC 评分函数的罚项函数功能消失,退化为最优参数对数似然评分函数。HVL 评分函数、CVL 评分函数也是为了防止模型和数据过度拟合,但其机制和前面的几种加罚项函数的机制不一样。HVL 评分函数、CVL 评分函数将样本数据分为训练集(training sets)和测试集(testing sets)。先用训练集对待选结构进行参数估计得到完整模型,然后用测试集做条件集,计算网络结构的似然度以求得最优结构。当样本量趋于无穷时,HVL 评分函数与 CVL 评分函数等价。

评分搜索方法是从众多的候选结构中发现具有最优评分的结构。最早出现的评分搜索算法是 Chow-Liu 树,并不断有改进的版本^[60-62]。1992 年,Cooper 等建立了著名的 K2 算法^[49],该算法在已知节点顺序的情况下,通过贝叶斯评分准则评估候选结构,再使用爬山搜索策略搜索最优结构。K2 算法兼顾了先验信息的重要性,在贝叶斯网络结构学习的发展史中的地位举足轻重。Bouckaert 于 1994 年提出 K3 算法^[63],他用 MDL 评分准则替换了 K2 算法的评价标准。K3 算法增加了约束网络复杂程度的罚项函数,避免了学到过于复杂的网络。同样,学者们也考虑了无先验信息时的学习。Lam 等分别在 1993 年和 1994 年提出的算法用 MDL 评分准则作为衡量标准,在没有节点拓扑序的前提下,只利用得分函数和搜索策略学习正确结构^[64,65]。

近年来,随着各种智能优化算法的发展,涌现出一批学习贝叶斯网络结构的混合算法。Larranaga 等^[66,67]和 Etxeberria 等^[68]利用遗传算法学习节点拓扑序,再结合 K2 算法搜索最优网络结构,并对学习效果进行了详尽的分析。Campos 等提出了蚁群优化 B (ant colony optimization B, ACO-B) 算法,该算法结合 K2 算法和蚁群算法搜索得分最好结构,具有较高效率^[69]。Pinto 等^[70]、Daly 等^[71]随后也提出了基于蚁群思想学习贝叶斯网络的优化算法。Sebastiani 等^[72]使用界定折叠 (bound and collapse, BC) 算法来学习结构,算法开始于随机选择的一个候选结构,使用 BC 方法给结构进行打分,接着对当前结构的边进行修正以提高结构评分,选择得分最好的作为算法输出。如今模拟退火^[73-79]、分布估计算法^[67,80]、分支定界算法^[81-83]、蜂群算法^[84-87]、进化算法^[88,89]等也都被用到贝叶斯网络的结构学习中。

在忠实数据下,BIC 评分函数评价马尔可夫等价的结构时,评分是相同的。同样,其他几个常见的评分函数如最优参数对数似然评分函数、MDL 评分函数和 AIC 评分函数也如此。因此,我们在学习时只要找到最优网络的一个等价代表即可。基于评分搜索算法的优点是算法思想简单易懂,过程规范,但是也有全局学

习方法的局限性,即具有指数级的学习复杂度。现有的方法大多是在有节点拓扑序下进行的,因此序的选择将影响学习结果。其缺点是容易陷入局部最优,不能处理高维网络。

由于两种方法各有优缺点,因此许多学者提出了将二者融合的算法。基于内容(content-based, CB)算法是一个典型的代表^[73],CB算法首先利用改进后的PC算法学习节点的顺序,再使用改进的K2算法确定最优结构。最大最小爬山(max-min hill climbing, MMHC)算法^[90]也是一个结合基于测试方法和打分方法的混合算法。MMHC首先利用条件独立性测试学习出一个无向的框架图,然后用贪心爬山算法确定边的方向^[90]。2009年,Ji等^[87]结合基于约束方法、蚁群算法和模拟退火机制,构建了一种新的混合算法,取得了较好的效果。2012年Zhu等将分解思想和K2算法相结合学习贝叶斯网络结构,提高了K2算法的效率^[91,92]。

2. 不完备数据下的结构学习

当数据不完备时,情况相对复杂。实际上,由于观测困难或者人力所不能及等因素,我们掌握的数据常常是缺失性的^[93]。在这种情况下,通常先对数据进行相应的修补,然后用已有的完备数据下的学习方法进行结构学习。因此,设计合理的数据完备化的策略是关键和难点。由Friedman提出的结构最大期望(structural expectation maximization, S-EM)算法是不完备数据下最为经典的算法^[94]。S-EM算法每次迭代都是在结构乘以参数的空间进行搜索,每次迭代,要么找到一个更适合当前结构的参数,要么选择一个新的结构,前者是一个标准的参数的最大期望(expectation maximization, EM)步骤,后者是一个结构的EM步骤。随后Friedman又在文献[95]、[96]对S-EM进行了改进。Borchani等提出的GES-EM(greedy equivalence search EM)算法有效地减少了文献[97]中爬山算法的计算复杂度。

1.2.2 贝叶斯网络推理

贝叶斯网络推理是指在约定条件下,利用网络的有向无环图和联合概率分布,通过概率演算回答查询问题的过程。根据查询问题的不同,推理问题可以概括为后验概率问题、最大后验假设问题及最大可能解释问题^[58]。对一个贝叶斯网络,给定一部分节点变量的取值,求感兴趣的另一部分变量的条件概率分布问题称为后验概率问题。给定证据变量(evidence variables)的取值,求感兴趣变量的后验概率最大的状态组合的问题称为最大后验假设(maximum a posterior hypothesis, MAP)问题,这里感兴趣的变量称为假设变量(hypothesis variables),每个状态组合称为一个假设(hypothesis)。当最大后验假设问题中的假设变量取遍证据变

量外的其余变量时,就是一个最大可能解释 (most probable explanation, MPE) 问题。Cooper 证明了贝叶斯网络推理是 NP 难问题 (nondeterministic polynomial-hard, NP-hard) [98]。根据精度,推理可分为精确推理和近似推理。精确推理方法是指可以精确计算查询问题的算法,一般用来处理稀疏的贝叶斯网络。对稠密且节点数量多的网络,算法的计算量将呈指数级增长。这时为了降低计算量,就只能在保证推理正确的前提下,适当降低推理的精度,近似推理算法就应运而生。精确推理算法主要有消息传递 (message passing) 算法 [99]、条件算法 [100-102]、团树 (clique tree) 传播算法 [103-106]、基于组合优化的推理算法 [107-110]、弧反转 (arc reversal) 算法 [111-113]、微分算法 [114,115]。近似推理算法主要有随机抽样 (stochastic sampling) 算法 [116-119]、基于搜索的近似算法 [120-122]、模型化简 (model simplification) 算法 [123-125]、循环消息传递 (loopy propagation) 算法 [126,127]。

厉海涛等对近年来贝叶斯推理进行了系统的总结,并从网络类型、算法复杂度、精度、算法关键、优点等方面列表对比了常见算法,见表 1.1 [128]。从表 1.1 中可以看出,每种算法都有各自的优势与不足,目前还没有普适方法。因此,在处理实际问题时,针对特定的情况,要选择最优或近似最优的算法。

表 1.1 贝叶斯网络推理算法对比

算法名称	网络类型	算法复杂度	精度	算法关键	优点
消息传递算法	单连通网络	网络节点的多项式	精确解	消息传递方案	计算快速、简单
团树传播算法	单、多连通网络	最大团节点的指数	精确解	寻找最大团节点最小的团树	目前速度最快 尤其适用稀疏网络
条件算法	单、多连通网络	条件节点集的指数	精确解	寻找最小条件节点集	对环路较少的网络,效率很高
基于组合优化的推理算法	单、多连通网络	因式分解中变量的指数	精确解	寻找最优因式分解	过程简单、应用广泛、效率较高
弧反转算法	单、多连通网络	弧反转所涉及节点数的指数	精确解	进行弧方向计算	计算简单,与仿真算法结合更有效
微分算法	单、多连通网络	正比于微分运算的复杂度	精确解	运算电路获取与微分	方法快捷,适于同时解决多个问题
随机抽样算法	单、多连通网络	与证据变量概率成反比	与样本量成正比	抽样算法	效果较好,算法发展完善,应用很广
基于搜索的近似算法	单、多连通网络	取决于网络概率分布的特征	与选择的状态有关	满足精度要求 状态集合求解	多用于实时网络的计算
模型化简算法	单、多连通网络	不同的化简方法有不同的复杂度	与化简方法有关	简化模型精度估计	简单、有效,多用于实时推理
循环消息传递算法	单、多连通网络	网络中环的个数的指数	与迭代次数有关	算法收敛性	收敛情况下近似效果好