

动态贝叶斯网络及其在 自主智能作战中的应用

■ 史建国 高晓光☆著

DONGTAIBEIYESIWANGLUO JIQIZAI ZIZHUZHINENGZUOZHANZHONGDE YINGYONG

兵器工业出版社

动态贝叶斯网络及其在自主 智能作战中的应用

史建国 高晓光 著

兵器工业出版社

内 容 简 介

贝叶斯网络和动态贝叶斯网络是20世纪70年代发展起来的新兴的职能推理技术,和其他推理工具相比较,是它的模型中既有经验知识,又体现观测数据、推理结果,是预测和观测修正的结合,因此可以说是一种更广泛的滤波工具。自从贝叶斯网络出现之后,便在国际上掀起学习和应用的热潮。目前贝叶斯网络被广泛应用在数据挖掘、机器人智能、语音识别、机器推理等诸多领域。

本书首先介绍了贝叶斯网络和动态贝叶斯网络的基本知识、推理方法和应用方法,接着主要介绍在已有的贝叶斯网络和动态贝叶斯网络的基础上,笔者自己的最新研究成果,包括离散模糊动态贝叶斯网络、分级离散动态贝叶斯网络、变结构离散动态贝叶斯网络的基本理论、推理算法以及在目标识别、态势评估和指挥决策中的应用。在本书的最后,以太空作战为背景,介绍了这些新理论的应用。

图书在版编目(CIP)数据

动态贝叶斯网络及其在自主智能作战中的应用/史建国,高晓光著. —北京:兵器工业出版社,2008.12
ISBN 978-7-80172-978-1

I. 动… II. ①史…②高… III. 贝叶斯推断—应用—智能—战争 IV. E81

中国版本图书馆CIP数据核字(2008)第132314号

出版发行:兵器工业出版社
发行电话:010-68962596, 68962591
邮 编:100089
社 址:北京市海淀区车道沟10号
经 销:各地新华书店
印 刷:北京市北中印刷厂
版 次:2008年12月第1版第1次印刷

责任编辑:周宜今
封面设计:李尘工作室
责任校对:郭 芳
责任印制:赵春云
开 本:787×1092 1/16
印 张:8
字 数:202千字
定 价:26.00元

前 言

贝叶斯网络和动态贝叶斯网络是 20 世纪 70 年代发展起来的一门新兴的智能推理工具，是表示因果关系的有向无环图。由于其基本理论来源于贝叶斯条件概率公式和全概率公式，因此得名贝叶斯网络。动态贝叶斯网络是贝叶斯网络随时间的发展。相对而言，动态贝叶斯网络更有利于滤除观测值的噪声，因此特别适合于处理序列数据。

贝叶斯网络、动态贝叶斯网络和其他推理工具相比较，它的模型中既有经验知识，又体现观测数据，将观测数据和经验知识相结合进行精确推理，因此可以说是一种更广泛的滤波工具。自从贝叶斯网络出现之后，便在国际上掀起了学习和应用的热潮。目前贝叶斯网络被广泛应用在数据挖掘、机器人智能、语音识别、机器推理等诸多应用中。

我国关于贝叶斯网络研究较晚，国内也有诸多的学者在进行研究。笔者在总结已有的贝叶斯网络和动态贝叶斯网络的理论和方法的基础上，加入最新研究并公开发表的离散模糊动态贝叶斯网络、分级离散动态贝叶斯网络、变结构离散动态贝叶斯网络、连续型动态贝叶斯网络的基本理论和方法，著成此书，希望能够对贝叶斯网络和动态贝叶斯网络的研究有所裨益。

本书编写过程中，参考了国内外诸多学者的研究成果，在此表示衷心的感谢，刘振同志编写整理了本书的第 1、2 章有关内容，一并表示感谢。

由于水平有限，书中缺点错误在所难免，恳请读者批评指正！

史建国 高晓光

2008 年 11 月

目 录

第1章 绪论	(1)
1.1 贝叶斯网络的发展历史	(1)
1.2 贝叶斯网络的学习与推理	(3)
1.3 贝叶斯网络的研究与应用现状	(9)
第2章 静态贝叶斯网络	(10)
2.1 静态贝叶斯网络的定义	(10)
2.2 静态贝叶斯网络的推理	(11)
第3章 动态贝叶斯网络与其推理算法	(18)
3.1 动态贝叶斯网络的定义	(18)
3.2 HMM 推理的前向—后向算法	(19)
3.3 DBN 中推理的边界算法	(24)
3.4 动态贝叶斯网络的接口算法	(27)
3.5 离散动态贝叶斯网络的直接计算推理算法	(29)
3.6 连续型动态贝叶斯网络介绍	(33)
第4章 离散模糊动态贝叶斯网络及其应用	(45)
4.1 离散模糊动态贝叶斯网络的概念和推理算法	(45)
4.2 离散模糊动态贝叶斯网络用于空战目标识别	(51)
4.3 离散模糊动态贝叶斯网络用于空战态势评估	(55)
第5章 分级离散动态贝叶斯网络与作战指挥决策	(61)
5.1 思想的提出	(61)
5.2 分级离散动态贝叶斯网络及其推理算法	(62)
5.3 分级离散动态贝叶斯网络用于多无人作战飞机指挥决策	(66)
第6章 变结构离散动态贝叶斯网络及其应用	(70)
6.1 思想的提出	(70)
6.2 变结构离散动态贝叶斯网络的定义	(71)
6.3 变结构离散动态贝叶斯网络的推理	(71)
6.4 变结构离散动态贝叶斯网络的自适应参数产生算法	(74)
6.5 突发固定威胁下的飞行器航路决策	(76)
6.6 突发移动威胁下的飞行器突防航路决策及仿真	(79)
第7章 动态贝叶斯网络太空飞行器的自主智能作战	(87)
7.1 航天器相遇的解命中问题	(87)
7.2 卫星的类型识别问题	(99)

7.3 太空航天器作战的态势评估和威胁计算	(106)
7.4 轨道武器的攻击作战决策	(112)
英语缩略语表	(118)
参考文献	(119)

第1章 绪 论

1.1 贝叶斯网络的发展历史

建立能像人一样思维的计算机系统，是人工智能研究者梦寐以求的事。早期的研究者认为：人类的思维具有普遍的规律，研究和总结这些规律，并建立一个通用、万能的符号逻辑运算体系来模拟这些规律，就能使计算机达到人的智能。但是实际中，人是根据知识行事，而不是简单地依据抽象原则上的推理行事。之后出现的“知识工程”领域，主要就是研究如何更好地表达、学习以及应用知识来解决实际问题。

知识可分为两类：确定知识和不确定知识。确定知识是指人们认识比较清楚，其内部关系把握比较精确的知识。而对于某些知识，由于信息不完备、背景知识不足、信息描述模糊以及信息中含有噪声等原因，人们对其认识存在不确定性，故称之为不确定知识。传统的不确定知识表示模型有模糊逻辑、可信度方法和神经网络等。可信度方法是通过对产生式系统进行扩充而来的。其做法是对每一个产生式规则赋予一个表示其确定程度的值，该值称为该规则的可信度（certain factor），同时对可信度定义了运算规则，在利用产生式规则进行推理的同时，伴随着一个可信度的计算过程，从而使推理结果不再是确定的，而是一个具有可信度的结果。可信度方法结合概率理论知识部分地解决了不确定推理问题，在推广至新领域时必须根据实际情况对其进行修改。与形式逻辑一样，概率论已经形成一套严密的数学体系。人们一直试图用概率来表达不确定知识，但有两个问题曾长期困扰着人工智能研究者：一是对概率的频率性的理解；二是联合概率分布的计算。近年来，概率的另一学派——主观概率学派引起了人们的注意，其中有代表性的是对概率的信度理解。该理论认为，一个事件的概率首先是个人的一种信任度，即人们根据自己现有的知识对事件发生可能性的一种主观估计，称之为先验概率。其次，这种信任度还应该是一种合理的信任度，即这种主观估计应该符合经典概率理论的公理体系。按照这种对概率的主观性理解，一个事件的概率并不需要重复实验来获得，但重复试验的结果会作为新的知识，使人们对事件的概率估计更精确，称之为后验概率。基于这种对概率的主观性理解，著名学者 Pearl 做出了开创性的研究。他认为人在进行推理时，并不以联合概率分布的形式来表示，而是以变量之间的相互关系以及条件独立性来进行。在此分析的基础上，Pearl 提出了一种图形结构来表示这些随机变量之间的条件依赖关系。图形结构的使用使这种关系更直观，并可以利用图形理论来设计更有效的概率计算算法。这种图形结构定义了一种新的知识表达模型——贝叶斯网络，为人工智能领域中不确定知识的处理提供了新的方法。

贝叶斯（Reverend Thomas Bayes, 1702—1761）学派奠基性的工作是贝叶斯的论文“关于几率性问题求解的评论”。著名的数学家拉普拉斯（Laplace P. S.）用贝叶斯的方法导出

了重要的“相继律”。20 世纪初,意大利的菲纳特 (B. de Finetti) 以及英国的杰弗莱 (Jeffreys H.) 都对贝叶斯学派的理论做出重要的贡献。第二次世界大战后,瓦尔德 (Wald A.) 提出了统计的决策理论,在这一理论中,贝叶斯解占有重要的地位;信息论的发展也对贝叶斯学派做出了新的贡献。1958 年英国最悠久的统计杂志“*Biometrika*”全文重新刊登了贝叶斯的论文,20 世纪 50 年代,以罗宾斯 (Robbins H.) 为代表,提出了经验贝叶斯方法和经典方法相结合,引起统计界的广泛注意,这一方法很快就显示出它的优点,成为很活跃的一个方向。

贝叶斯网络 (Bayesian network) 又称为信念网络 (belief network), 是一种对概率关系的有向图解描述,适用于不确定性和概率性事物,应用于有条件地依赖多种控制因素的决策,在解决许多实际问题中,需要从不完全的、不精确的或不确定的知识和信息中做出推理。而贝叶斯网络是一种概率推理技术,使用概率理论来处理在描述不同知识成分的条件相关而产生的不确定性,提供了一种将知识直觉的图解可视化的方法。贝叶斯网络是一种图形化的模型,能够图形化地表示一组变量间的联合概率分布函数,一个贝叶斯网络包括了一个结构模型和与之相关的一组条件概率分布函数。结构模型是一个有向无环图,其中的节点表示了随机变量,是对于过程、事件、状态等实体的某特性的描述,边则表示变量间的概率依赖关系。图中的每个节点都有一个给定其父节点情况下该节点的条件概率分布函数。这样,一个贝叶斯网络就用图形化的形式表示了如何将与一系列节点相关的条件概率函数组合成为一个整体的联合概率分布函数。因果贝叶斯网络是指具有因果含义的贝叶斯网络,其中每个节点的父节点被解释为该节点相对于模型中其他节点的直接原因。为了与之区别,有时也将没有因果意义的贝叶斯网络称为概率贝叶斯网络。

贝叶斯网络起源于 20 世纪 70 年代末期,是一种有向无环图,有向无环图的概念最早起源于遗传学家 Sewall Wright。他最早提出了一种路径分析 (path analysis) 的方法,这种方法随后被作为因果模型的一种确定的表达方式应用于经济学、社会学以及物理等多个学科中。此时的统计学家将此种网络称之为回归模型 (recursive model)。影响图是有向无环图用于决策分析的另一种表达方式。有向无环图的作用是提供一种高效的概率函数的表达方式,一旦网络模型建立好以后,随后的计算都可以利用概率表示的符号计算得到。

贝叶斯网络是贝叶斯方法与图形理论的有机结合。1986 年 Pearl 首次在专家系统中引进了贝叶斯网络。1988 年 Pearl 明确指出影响图中没有决策节点和结果节点就是贝叶斯网,指出贝叶斯网或许是概率推理中最普及的模型。1989 年 Andreessen 使用贝叶斯网建造专家系统 MUNIN (Muscle and Nerve Inference Network)。Cooper 证明了概率值的传播计算问题是 NP 难的。贝叶斯网络方法由于其理论上的严格性和一致性,以及有效的局部计算机制和直观的图形化知识表达,已经成为人工智能领域的研究热点。国际权威期刊“*Communications of the ACM*”1995 年发表贝叶斯网络研究专辑,在“*Artificial Intelligence*”和“*Machine Learning*”等重要杂志上有很多贝叶斯网络方面的文章,在 UCAI, AAAI, ECAI, ACM, UAI 等重要会议上贝叶斯网络研究也占到了相当大的比重。

贝叶斯网络首先被作为一种计算的框架,随后才被作为一种认知的工具。此时被用于对阅读理解中的分布式信息进行模拟,必须结合语境以及感知到的证据利用双向推理最终形成一个连贯的解释。贝叶斯网络更新的核心是双向推理,这种思想最早来源于 Reverend Bayes 于 1773 年提出的后验概率、先验概率以及似然函数的概念。

贝叶斯网络在认知领域内一开始并未得到充分的重视,但在专家领域却得到了广泛的应用。由于贝叶斯网络的双向推理功能使其填补了专家系统在这方面的空白,也就是在这个时期贝叶斯网络的研究开始繁荣起来。

“贝叶斯网络”这一术语是在1988年由Pearl在论文中提出的,奠定了贝叶斯网络的理论基础,早期的应用主要在专家系统中用于不确定性知识表示和推理。20世纪90年代后期Heckerman把贝叶斯网络用于数据挖掘,由于贝叶斯网络具有独特的不确定表达形式、丰富的概率表达能力、综合先验知识的增量学习特性成为数据挖掘众多方法中研究的热点方法之一。

动态贝叶斯网络(DBN)表示的是制约动态随机过程中轨迹可能发生变化的概率分布,是复杂随机过程的一种压缩表示。动态贝叶斯网络是扩展了的静态贝叶斯网络,可以模拟动态的过程。这里的动态过程不是指贝叶斯网络的结构在改变,而是指贝叶斯网络中的节点的值在动态地变化。因此在空战态势评估以及实时的决策中,以动态贝叶斯网络建立的模型更加与真实情况相符合,得到的效果也更具有鲁棒性。与BN一样,DBN的学习也分为参数学习和结构学习。目前已经有很多DBN参数学习的算法,它们一般都能够处理隐藏变量和缺值数据。而DBN的结构学习,尤其是有隐藏变量的DBN结构学习依然是目前研究的热点之一。1998年,Friedman将SEM算法扩展到DBN的结构学习中,用于从带隐藏变量和缺值的数据中学习DBN的结构。给定一个随机变量集随时间的部分观测,其算法能构造一个与观测拟合得很好的DBN。对动态贝叶斯网络的推理已经有很多成熟的方法,主要采用边界算法以及接口算法进行推理。接口算法是对边界算法的改进,便于计算机编程,有一定的应用范围。

1.2 贝叶斯网络的学习与推理

1.2.1 贝叶斯网络的学习

贝叶斯网络的学习包括参数和结构学习,贝叶斯网络的学习就是找出一种网络,能按某种测度最好与给定的实例数据集吻合。

贝叶斯网络的学习也就是要找出一个能够最真实反映现有研究领域中各数据量相互之间的依赖关系的贝叶斯网模型,即根据数据样本和先验知识找出后验概率最大的贝叶斯网模型。与基于规则等一类知识表示的知识系统相同,贝叶斯网首先必须解决知识获取问题,在这方面,仅仅利用人类专家的知识进行建网是困难的、不准确的,有时甚至是不可能的。从数据中学习建网已引起广大人工智能研究者的浓厚兴趣,成为当今的一个热门研究领域。

一个贝叶斯网络由两部分构成,网络结构 S 以及节点 X 的局部概率 P 。 S 是一个有向图,每个节点代表一个数据变量 X_i , Pa_i 为 S 中节点 X_i 的父节点集合。局部概率 P 中每个元素为数据变量 X_i 的条件概率密度 $P(X_i | Pa_i, \zeta)$, ζ 为观测者的先验知识, $P(X | \zeta) = p(X_1, X_2, \dots, X_n | \zeta) = \prod_{i=1}^n p(X_i | X_1, X_2, \dots, X_{i-1}, \zeta)$ 。对于任一数据变量 X_i ,必可找到一个与 X_i 条件都不独立的最小子集 $\pi_i \in \{X_1, X_2, \dots, X_{i-1}\}$,使得此时 π_i 中的变量就是贝叶斯网络中的 X_i 的

父节点 Pa_i , 故 $P(X|\zeta) = \prod_{i=1}^n p(X_i | Pa_i, \zeta)$ 。

1. 结构学习

近十年来, 贝叶斯网络的学习理论取得了较大的进展。目前贝叶斯网络结构学习的方法通常分为两大类。

① 基于搜索与评分的方法。运用评分函数对网络模型进行评价。通常是给定一个初始结构或空结构, 逐步增加或删减连接边, 改进网络模型, 从而搜索和选择出一个与样本数据拟合得最好的结构。根据不同的评分准则, 学习算法可分为基于贝叶斯方法的算法、基于最大熵的算法和基于最小描述长度的算法。

② 基于依赖关系分析的方法, 节点之间依赖关系的判断通过条件独立性 (conditional independence) 测试来实现。

对网络结构的学习已经有很多的方法, 除了局部 BD (Bayes - Dirichlet) 度量以外, 主要还有最小描述长度以及贪婪算法。最小描述长度 (MDL) 原理是 Rissanen 在研究通用编码时提出的。其基本原理是: 对于一组给定的实例数据 C , 如果要对其进行保存, 为了节省存储空间, 一般采用某种模型对其进行编码压缩, 然后再保存压缩后的数据。同时, 为了以后正确恢复这些实例数据, 将所用的模型也保存起来。所以需要保存的数据长度等于这些实例数据进行编码压缩后的长度加上保存模型所需的数据长度, 将该数据长度称为总描述长度。最小描述长度原理就是要求选择总描述长度最小的模型。

具有缺失数据的贝叶斯网结构学习比较困难, 现有的研究主要是基于打分的方法。数据的缺失导致两方面问题的出现, 一方面, 打分函数不再具有可分解形式, 不能进行局部搜索; 另一方面, 一些充分统计因子不存在, 无法直接进行结构打分。围绕这两个问题相继出现了一些解决的方法。

(1) 早期工作

Heckerman 等人给出了解决后一个问题的一些方法。这些方法对选择的贝叶斯网结构首先基于梯度的优化 (gradient-based optimization) 或期望最大化 EM (expectation maximization) 算法进行最大后验参数估计, 然后使用拉普拉斯近似 (Laplace approximation) 或贝叶斯信息标准 (Bayesian information criterion) 等大样本近似方法进行近似结构打分, 由于搜索空间大以及存在近似打分的误差, 使学习效率较低, 并且结果也不够可靠。

(2) 将期望最大化 (EM) 思想引入到缺失数据情况下的结构学习

1997 年, Friedman 借鉴参数学习的 EM 算法, 提出模型选择期望最大化 MS - EM (model selection EM) 的结构学习方法。该算法是将不完备数据下的结构学习问题转化为较容易解决的完备数据下的结构学习问题, 在 EM 过程的内部利用启发式搜索算法搜索最佳的网络结构。算法执行过程中始终有一个候选的网络结构, 在每次循环中, 通过计算评价网络结构需要的期望统计因子, 以试图发现某种评分标准下更好的网络结构。因为算法对于网络结构的搜索是在完备数据集条件下进行的, 所以能够利用完备数据条件下评分函数的可分解性简化搜索过程。进一步, 注意到对给定的网络结构寻找使打分最大的参数, 比寻找一个更好的网络模型在计算上要容易得多。因而可以交替进行对一个固定模型的参数的迭代优化和对不同网络模型的迭代搜索, 该算法称为交替模型选择一期望最大算法 AMS - EM (alternative model selection EM)。在 MS - EM、AMS - EM 算法中, 使用的是 BIC、MDL 等贝叶斯评分标准

的近似方法。因为理论和实践都表明,使用精确的贝叶斯打分往往是对领域模型一般性质的更好刻画。此外,使用贝叶斯评分标准,还可以在学习过程中利用先验知识。1998年,Friedman引入一种使用贝叶斯打分学习概率模型的新方法,贝叶斯结构期望最大化算法(Bayesian structure EM algorithm),简称为Bayesian-SEM算法。像MS-EM、AMS-EM算法一样,Bayesian-SEM算法也是把不完全数据转化为完全数据,在当前对于完备数据的最佳猜测的基础上,寻找贝叶斯分值最高的网络结构。与MS-EM、AMS-EM算法不同的是,此算法的搜索空间是网络结构,不再是网络结构 $\times$ 参数空间。并且如果对网络模型进行标准的连续和平滑处理,Bayesian-SEM算法可以较容易应用于含连续变量的数据样本的学习。这三种算法在每次循环中,都需要在当前的期望统计因子下,计算每个可能的网络结构的分值,哪怕是分值很低的网络结构,每次都需要评估。当搜索空间巨大时,计算量难以想象,因此只能用于解决网络结构较少的领域学习问题。由于算法具有收敛性,所以一种解决方法可以是在每次循环中,不苛求在当前的期望统计因子下分值最高的网络结构,而是只要找到一个分值有增进的网络结构即可。这样,每次循环不必评估每个可能的网络结构,降低了计算的复杂性,增强了贝叶斯网学习的实用性。除Friedman外,其他一些学者也对EM算法如何应用于结构学习进行了探索。基于EM算法框架进行具有丢失数据的贝叶斯网结构学习,使用期望充分统计因子代替不存在的充分统计因子,在一些假设下,可使打分函数具有可分解形式(可进行局部搜索),并且在每次迭代中,结构都有所改进,使结构序列收敛。该方法能够在一定程度上提高学习效率,但一般是收敛到局部最优结构。

(3) 将遗传算法引入到缺失数据情况下的结构学习

Pedro讨论了利用遗传算法进行结构学习。他采用连接矩阵表示网络结构。首先假定网络结构的节点间存在序关系,以保证遗传操作构造的网络是合法结构。然后讨论了当节点间不存在序关系时,如何把不合法的贝叶斯网“修复”成合法结构。Pedro设计了基于贝叶斯方法的适应度函数及选择、交叉及变异算子,对遗传算法的控制参数进行了性能分析,并且模拟ASIA和ALARM两种典型网络,对节点间存在序关系和没有序关系的学习结果进行了比较。Pedro的方法应用于从完备数据中学习,能够找到好的网络结构,不完备数据的学习结果不理想。由于目前已经有几种贪婪算法可以得到类似的结果,因此与确定性的贪婪算法相比,Pedro的研究没有说明进化计算应用于网络结构学习的优势。不过Pedro确实为这个领域的研究拓展了思路,补充了新的方法。W. Myers等人改进了Pedro的工作。他们采用遗传操作把不完备数据转化成完备数据,把遗漏的数据编码成基因,同时进化网络结构和遗漏的变量值。这种做法的缺点在于它指数级地扩大了搜索空间(遗漏的变量值 $\times$ 网络结构)。当遗漏的数据量较小时,该算法能够找到较好的网络结构;遗漏的数据量较大时,搜索空间非常巨大,算法的运行效率大大降低,难以得出令人满意的结果。更大的缺点是此算法难以保证收敛性,遗传算子把不完备数据转化成完备数据,转换无依无据,随机性强,无法反映遗漏数据实际具有的概率分布,因而无法保证子代群体比不完备训练数据匹配得更好。

此外,国内吉林大学刘大有、王飞等人提出了EGA(exception & genetic algorithm)算法。该算法基于数学期望思想,在算法执行过程中始终保持一个当前最优的贝叶斯网,它把不完备数据转换成完备数据,从而使EGA能利用完备数据学习的优点,简化了算法的复杂度。该算法通过始终保留最优个体的遗传操作进化个体结构,并在数学期望的基础上利用最优个体进行数据转化,使每次进化都能提高群体的质量,从而保证算法可以收敛到全局最好

的网络结构,克服了确定性收敛到次优模型的问题。EGA 只对网络结构进行编码,避免了过于扩大搜索空间。和 Myers 等人的方法相比,适用于大型数据样本或遗漏的数据量较大时的结构学习。

(4) 其他工作

吉林大学王双成、范森森提出了 BN-GS (Bayesian network & Gibbs sampling)。该方法使用 Gibbs 抽样修复丢失的数据,基于依赖分析方法进行贝叶斯网结构学习和调整。首先随机初始化丢失的数据,并建立最大似然树作为初始贝叶斯网结构,然后进行数据集和贝叶斯网的迭代修正调整,直到结构趋于稳定或满足给定的终止条件为止。每次数据集修正后进行贝叶斯网结构调整,使调整后的贝叶斯网适合于当前的数据集,并且不会陷入局部最优结构。Gibbs 抽样迭代收敛到平稳分布,因此结构序列将收敛到平稳分布的贝叶斯网结构。联合概率可按贝叶斯网结构进行分解,对一个变量的抽样只需考虑对应的条件概率因子即可。而条件概率因子中条件变量的数量与所有变量的数量没有联系,解决了满条件分布 (full conditional distribution) 所带来的问题,从而能够显著提高抽样效率。

为使学习模型具有统计鲁棒性和易计算性,经常需要在网络结构中引入隐藏变量。1999 年,Boyan 等人提出一种方法,不需要穷举所有的搜索就能发现隐藏变量的存在。实验表明,这个方法是一种计算可行的复杂动态系统的学习算法。但是,此方法仅触及了如何发现隐藏变量这个问题的表面,能够发现时态交互过程中的相互关系,对于非时态的交互关系不够敏感。另一极端情形是,此算法不支持发现长期的依赖关系及收集进化速度不同的变量间的相互影响。现有的工作中,FCI 算法基于条件独立关系的测试发现隐藏变量,对独立关系测试过程中可能产生的误差敏感,搜索过程中的早期误差具有级联效应,对学习结果影响较大。且该算法只能发现由条件独立关系限定的隐藏变量,不能动态地引入可以使网络结构更简洁、更便于推理和预测地隐藏变量。此外,Friedman 提出在学习过程开始前,先固定隐藏变量的数目,具有盲目性。Gal Elidan 等人提出剥离开网络结构的学习和隐藏变量的引入,在学习结果中如发现较复杂的结构,则引入新变量重新学习,计算复杂度高。在国内,吉林大学刘大有、王飞等人提出了 HGA 算法,该算法将遗传算法引入含有隐藏变量的贝叶斯网结构学习,根据网络结构复杂度引入使网络结构更简洁的新变量,并由适应度函数保证学到的可能含隐藏变量的网络结构确实比较准确。

基于经典统计学的方法理论成熟,计算简单,但它只利用了实例数据集所提供的信息,无法加入专家知识,对实例数据的依赖性大,基于贝叶斯网络有机结合了两类信息,对实例数据的依赖性降低,学习结果更加准确。参数学习是贝叶斯网络学习的基础,是贝叶斯网络结构学习必不可少的部分。

2. 参数学习

随机数据集 $D = [X[1], \dots, X[m]]$ 的每个事例 $X[i]$ 中都观察到了所有的变量,则称数据完备, D 称为样本。

MLE 最大似然估计,是实例数据完备情况下的学习方法,基本思想是:一个随机实验有若干个可能的结果 $C_1, C_2, \dots, C_n$,在一次实验中,结果 C_m 出现,即 C_m 出现的概率应该最大,可将似然函数 $p(c|\theta)$ 取极大值时的参数值 θ 作为对参数的估计值。似然性是判断具体 θ “好坏”的一种标准,依据 θ 产生样本的可能性,即似然性函数

$$L(\theta; D) = p(D|\theta) = \prod p(X[m]|\theta)$$

似然性越大,具体的 θ 越“好”,这可以推广到含有 n 个变量的情形。贝叶斯方法利用贝叶斯网表示数据取样过程中的不确定性。基本思想是:给定一个含有未知参数的分布以及一个完整的实例数据集 C , θ 是一个随机变量,具有一个先验分布 $p(\theta)$,可以根据以往的知识估计,或者认为 $p(\theta)$ 是一个均匀分布,参数 θ 的信息发生变化,表示为 $p(\theta|C)$,称为参数 θ 的后验概率。贝叶斯参数学习的任务就是计算这个后验概率,并作为参数估计的依据。通常采用条件期望估计方法,又称贝叶斯预测法。MLE方法是频度学习方法,偏差为0,是所有无偏差估计方法中方差最小的。由于没有利用先验知识,收敛速度较慢,贝叶斯方法采用顺序计算方式,不否认以往所有工作,克服了这个缺点。

不完备数据集下的参数学习。

现实中的数据常常不完备,这时似然函数的计算将变得很复杂,精确计算极大值几乎不可能,只能近似求出似然函数的极大值,并将该点的参数作为估计值。常用的有期望最大化(EM)和梯度上升算法、蒙特·卡洛、高斯等方法。

EM算法的基本思想:当实例数据不完备时,利用贝叶斯网的推理算法,可以对不完备实例中没有观测值的变量进行估计,从而形成一个完整的实例数据集。Little和Rubin证实当能够建造统计模型时,EM算法是一种好的学习算法。

蒙特·卡洛方法、最大梯度法和EM算法与数据样本中遗漏数据的比例密切相关,比例越高,蒙特·卡洛方法越不精确,达到收敛时间越长,最大梯度法和EM算法陷入局部极大的可能性越大。最大梯度法和EM算法在计算上的“瓶颈”是贝叶斯推理,网络结构复杂时精确计算几乎不可能,只能进行模拟计算或近似计算。

EM算法比简单梯度算法收敛速度快,但和共轭梯度方法比较尚不明朗。Lauritzen说明在EM算法应用有困难时,梯度方法可能是一种替代方案。Thiesson综合运用EM算法和共轭梯度法,当接近最大值时采用共轭梯度法以加速收敛速度。蒙特·卡洛方法和EM算法都需要很大的资源,收敛速度较慢,算法的执行时间和数据遗漏的数目密切相关,而且要求满足MAR假设。Ramoni和Sebastiani引入一种不依赖MAR假设的确定性方法,称为定界和塌缩(BC),首先给出和数据库中的观察相一致的参数可能近似值的区域,然后通过凸函数的组合模型在该区域中找出一个近似点。实验表明,满足MAR假设时,BC和Gibbs抽样计算出的结果等价,具有更好的鲁棒性,并且BC算法的执行时间短,不依赖数据遗漏的数目。Ramoni和Sebastiani还把BC算法扩展用于计算边际似然性,并进一步用于不完备数据的网络模型的抽取。从算法的计算复杂性来看,Gibbs抽样算法的效率是最低的,且只能对条件概率表进行批量更新。当选用相同的推理算法时,最大梯度法和EM算法的计算复杂度相同,既可以对条件概率表进行批量更新,又可以用于对条件概率表的顺序更新。

1.2.2 贝叶斯网络的推理

贝叶斯推理是在一个不确定环境下和不完全信息下的进行决策支持和因果发现的工具。贝叶斯推理提供了一种以概率为基础的推理算法,它对人们感兴趣的受概率控制变量,结合观察的数据,对这些概率进行推理做出最优的决策。因为它提供了一种通过加权证据支持可用假设的定量方法,这不仅为那些直接操纵概率的算法提供理论基础,而且也分析那些没

有明确的运算概率的算法提供了理论构架。

利用网络表示概率信息，可以追溯到 1921 ~ 1934 年期间，遗传学者 Sewall Wright 在对于基因遗传和动物生长因素的概率分析工作中，首次采用有向无环图表示的概率模型。其后，在人工智能领域相继产生了各种不同的概率图模型。1961 年，I. J. Good 和 Alan Turing 合作发展了概率的表示方法以及贝叶斯推理方法，可以认为是现代贝叶斯网络的先驱。

在 20 世纪 70 年代后期，以随机变量的有向无环图形式出现的影响图 (influence diagram) 或决策网在决策分析问题中有广泛的应用。1982 年，Judea Pearl 提出消息传递方法处理树形网络中的推理问题。1983 年，JinKim 把该方法扩展到单连通图。为解决多连通图上的推理，1985 年，Judea Pearl 提出环路割集 (cutset) 方法，1988 年，David Spiegelhalter 和 Steffen Lauritzen 提出了集团树消息传播 (clique-tree propagation) 算法，最早在多连通概率网推理问题中使用了聚类的方法。该方法将多连通图转化为集团树，令信任度在集团之间传播。推理效果对稀疏网络很有效，对于密集网络运行相当缓慢，复杂性随最大集团节点个数呈指数级变化。变量消去算法 (variable elimination) 概念上比集团树方法更简单，最早由 Shachter 提出该思想。1994 年，Zhang 和 Pool 明确将其作为算法提出。该方法执行期间，消去与查询变量无关的节点。1990 年，Cooper 证明一般贝叶斯网模型上的精确概率推理是 NP 难解的，于是研究者们开始寻找近似的推理算法。现有的近似推理算法包括随机模拟方法，模型简化方法，基于搜索的方法和循环信任度转播方法等。

随机模拟方法是近似推理算法中最常见的一类，通过反复模拟网络中的事件取得采样，计算采样中发生的相关事件的频率来确定查询节点的概率值。1988 年，最早由 Max Henrion 将统计学中的拒绝采样 (rejection sampling) 算法应用于贝叶斯网络。1989 年，Fung 和 Chang 以及 Shachter 和 Peot 提出的似然加权 (likelihood weighting) 算法，该算法是重要性采样算法的一个实例。1990 年，Shachter 和 Peot 提出了重要性采样算法的两个变形：自身重要性采样 (SIS) 和启发式重要性采样 (HIS)，改进了似然加权算法的运行状况。1993 年，Paul Dagum 和 Mike Luby 证明，近似推理过程的一般形式也属于 NP-hard 复杂性类。1996 年，Dagum 和 Luby 设计提出有界差异算法和 AA 算法。这两个算法是似然加权的变形，是基于似然加权算法和停止规则原理而提出，效率优于似然加权算法。2000 年，Cheng 和 Druzdzel 提出了一种自适应重要性采样 (adaptive importance sampling) 算法，是至今为止最有效的随机采样算法。以上所述随机采样算法都属于重要性采样算法这一范畴之内，各个采样是相互独立的。另外还有一类随机采样算法，称为马尔可夫链蒙特·卡洛 (MCMC) 方法。在这一算法中，采样是相互依赖的。MCMC 方法始于 Metropolis 等人所提出的 Metropolis 算法。1984 年，S. Geman 和 D. Geman 针对无向马尔可夫网络中的推理问题提出了 Gibbs 采样算法。1987 年，J. Pearl 将 MCMC 算法应用于贝叶斯网络精确算法。从 20 世纪 90 年代开始，各种模型简化的方法纷纷提出来，用于降低模型复杂度，提高推理效率。例如 1994 年提出的弱依赖消去方法，1997 年提出的弧消去方法。1995 年提出定位局部估计算法从网络中删除选中的节点。1989 年提出的有界调节算法忽略一些切集实例来计算概率的大致范围，考虑更多的实例来提高精确度。1994 年提出的状态空间抽象算法，通过减少 CPT 的基数来简化模型。基于搜索的方法，适用于相对较小的联合概率空间包含大部分概率这样的概率分布。这些算法搜索概率值较高的实例，用它们获得合理的近似。具体算法包括 1991 年由 Henrion 提出的 Top - N 基于搜索的方法；1993 年 Poole 提出的使用冲突的搜索方法；1996

年 Santos 提出的确定性近似和采样积累方法等。1999 年, 由 K. P. Murphy, Y. Weiss 和 M. Jordan 提出循环信任度传播 (Loopy belief propagation) 算法, 该算法是 Pearl 的多叉树传播算法在贝叶斯网中循环应用。经验表明, 计算机视觉方面, 该算法有很好的应用。至今, 各种贝叶斯网络的推理算法已达数十种之多。

对于贝叶斯网络的精确推理算法常用的方法主要有桶消元法、连接树算法以及信息传递算法, 其中连接树算法是应用最为广泛和计算速度最快的方法, 对变量进行消除、道德化以及三角化以后就形成一个删除树 (elimination trees), 删除的节点满足运行交叉属性 (running intersection property) 后, 便可以从中构造一个连接树。RIP 是指一个有序的集合序列 $C_1, \dots, C_k$ 对所有的 $1 < j \leq k$, 存在一个 $i < j$, 使得 $C_j \cap (C_1 \cup \dots \cup C_{j-1}) \subseteq C_i$ 。贝叶斯推理分为精确和近似算法, 对贝叶斯网络的推理有很多, 应用较为广泛的就是连接树算法 (junction tree algorithm), 构造连接树是进行其他算法的基础。

Pearl 提出的算法在贝叶斯网络的推理中也应用较为广泛, 其基本思想就是每一节点分配一个处理器, 每个节点既可以发送证据 (distribute evidence) 也可以收集证据 (collect evidence), 因此经过这样的过程后, 每个节点的信念度就会发生改变从而得到了更新。

1.3 贝叶斯网络的研究与应用现状

贝叶斯网络由于其对推理模式的扩充和改进, 以及以严谨的概率数学为基础, 在专家系统的不确定性推理领域, 逐步取代了早期基于规则的方法, 已广泛应用于诊断、预测、决策、信息分类、数据挖掘、文本检测、模式识别、系统控制等领域。基于贝叶斯网的应用软件, 大致可以分为应用于特定领域的专家系统和通用的贝叶斯网络开发工具软件两类。

美国通用公司已经开发出基于贝叶斯网络的故障诊断系统, 美国航天局和 Rockwell 公司利用贝叶斯网络开发出太空船的推进器故障诊断系统, Lewis. R. W. 和 Ransing. R. S. 将贝叶斯网络用于制造业压制过程中的检测和诊断, Nikovski D. 已将贝叶斯网络用于医学中。

因此, DBN 已成为大规模数据库中进行知识发现的重要手段。Sebastiani 等人用贝叶斯网为慈善机关的客户建模, 能更好地预测捐赠人的特征和希望得到的礼物, 帮助慈善机构获得最大的效率。Cheeseman 等人使用朴素的贝叶斯网设计了一个无监督的分类系统 Autoclass System, 取得了很好的效果。美国 NBA 教练, 利用 IBM 公司提供的贝叶斯网络数据挖掘技术, 临场决定替换队员, 获得了大的成功。国内在知识发现中应用贝叶斯网的研究相对较少, 慕春棣等人对贝叶斯网学习方法的数学推导进行了分析总结, 并介绍了贝叶斯网在数据挖掘中的应用。王玮等人实现了一种基于贝叶斯网方法进行知识发现的算法, 取得了一定的效果。总之, 由于贝叶斯网的预测能力以及它具有显示领域变量间最重要的直接关联关系和阐明领域结构的优势, 贝叶斯网已经成功应用于许多数据挖掘的应用系统中。

第 2 章 静态贝叶斯网络

2.1 静态贝叶斯网络的定义

贝叶斯网络 (Bayesian network) 又称为信念网络 (belief network), 贝叶斯网络是一种有向无环图, 于 1986 年由 Pearl 提出, 是人工智能非精确推理领域近十几年来研究的热点。贝叶斯网络使用概率理论来处理知识的不确定性, 提供了一种将知识直觉地可视化的方法, 是一种新的知识表示模型和推理方法。

贝叶斯网络理论将先验知识与样本信息相结合、依赖关系与概率表示相结合, 是数据挖掘和不确定知识表示的理想模型, 下面我们举一个简单的贝叶斯网络的例子。表 2.1 为各个节点之间的条件概率表, 没有边的节点可视为两者没有联系, 图 2.1 方框内的数字表示节点的先验概率值。贝叶斯网络就是在给定观察值的情形下推理得出节点的状态, 图 2.1 表示各个节点的先验概率, 当我们观察到 $P(\text{Walson} = \text{yes}) = 1$ 以及 $P(\text{Icy} = \text{no}) = 1$ 时, 推理得出的结果如图 2.2 所示。

表 2.1 贝叶斯网络的条件概率表

	$P(\text{Holems} \text{Icy})$		$P(\text{Walson} \text{Icy})$	
	yes	no	yes	no
yes	0.8	0.2	0.8	0.2
no	0.1	0.9	0.1	0.9

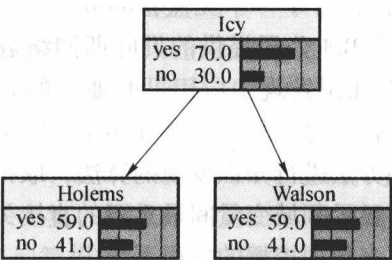


图 2.1 简单贝叶斯网络举例

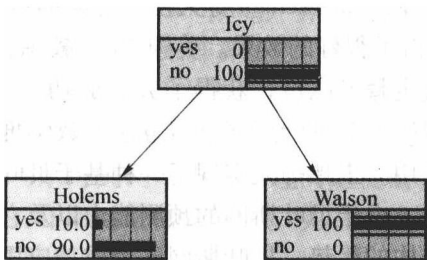


图 2.2 贝叶斯网络推理结果

与数据挖掘中的其他方法: 如规则表示、决策树、人工神经网络等相比, 贝叶斯网络具有下列优点: 贝叶斯网络能够方便地处理不完全数据。例如, 考虑具有相关关系的多个输入变量的分类或回归问题, 对标准的监督学习算法而言, 变量间的相关性并不是它们处理的关键因素, 当这些变量中有某个缺值时, 它们的预测结果就会出现很大的偏差。而贝叶斯网络

则提供了较为直观的概率关联关系模型。

贝叶斯网络能够学习变量间的因果关系。因果关系是数据挖掘中的极为重要的模式。原因有二：在数据分析中，因果关系有利于对领域知识的理解；在干扰较多时，便于作出精确的预测。例如市场销售分析人员想知道增加广告投入是否能提高产品的销量。为回答这个问题，分析人员必须知道，在某种程度上，广告投入是否是提高销量的原因。即使没有这方面的实验数据，贝叶斯学习对这类问题的回答也是相当简单的，因为这种因果关系已经包含在贝叶斯网络模型中了。贝叶斯网络与贝叶斯统计相结合能够充分利用领域知识和样本数据的信息。任何从事过实际建模任务的人都会知道，先验信息或领域知识在建模方面的重要性，尤其是在样本数据稀疏或数据较难获得的时候，一些商业方面的专家系统完全根据领域专家知识来构建就是一个很好的例证。贝叶斯网络用弧表示变量间的依赖关系，用概率分布表来量化变量之间的依赖关系。

贝叶斯网络理论在人工智能领域处理不确定性问题具有其特有的优点，它所具有的突出的特点有：

- ① 能够处理不确定性和概率性的事件与事物。
- ② 能够用于学习因果或其他类型的关系。
- ③ 是一种将（专家）先验知识和数据进行综合的理想表达模式。
- ④ 能够处理不完全（或部分数据丢失的）数据集。

总之，BN 模型灵活性好，能自然地将专家知识融入模型中，具有令人瞩目的从数据中导出模型的能力，并能继续用专家知识和数据改进模型的性能，而且它的模型结构和参数具有明确的含义。因此，BN 是应用于智能分析、智能学习、智能控制等智能领域的强有力工具之一。

贝叶斯网络是一个具有以下特征的图形结构。

- ① 贝叶斯网络是一个带有条件概率的有向无环图 DAG (directed acyclic graph)。
- ② 节点表示随机变量，节点间的弧反映了随机变量间的条件依赖关系，指向节点 X 的所有节点称为 X 的父节点。
- ③ 与每个节点相联系的条件概率表 CPT (conditional probability table) 列出了此节点相对于其父节点所有可能的条件概率。由此可以看出，贝叶斯网络是基于概率分析和图论的一种不确定性知识表达和推理的模型。从直观上讲，贝叶斯网络表现为一个赋值的复杂因果关系网络图，网络中的每一个节点表示一个变量，即一个事件，各变量之间的弧表示事件发生的直接因果关系。贝叶斯网络的结构表达了定性知识，即事件之间的因果联系。后验概率和条件概率表达了定量知识，即原因对结果的影响程度。

2.2 静态贝叶斯网络的推理

2.2.1 静态贝叶斯网络推理介绍

贝叶斯网络的推理原理基于贝叶斯概率理论，推理过程实质上就是概率计算过程。贝叶斯网络利用随机变量间的条件独立性，将一个联合概率分布直观地表达为一个图形结构和一系列的条件概率表，经消元（消除变量）计算可求出任一单变量的概率分布（边缘概率）