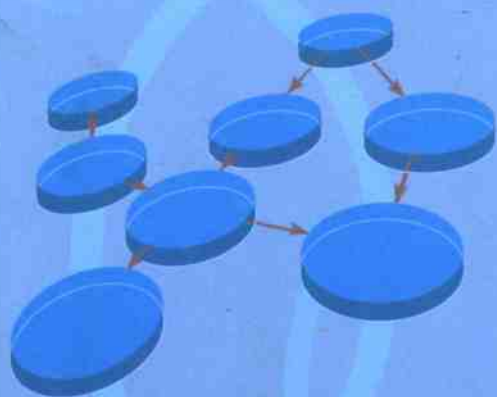



INTRODUCTION TO
BAYESIAN NETWORKS

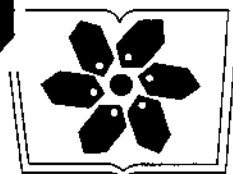
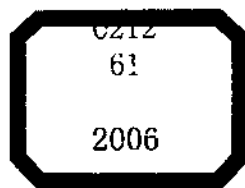
贝叶斯网引论

张连文
郭海鹏

著



 科学出版社
www.sciencep.com



中国科学院科学出版基金资助出版

贝叶斯网引论

张连文 郭海鹏 著

科学出版社

北京

内 容 简 介

贝叶斯网是将概率、统计应用于复杂系统的不确定性推理和数据分析的一种有效工具,它起源于20世纪80年代中期对人工智能中的不确定性问题的研究,近年来在国际上的影响不断扩大。本书是第一本系统论述贝叶斯网的基本理论、算法及其应用的中文专著。内容包括概率论及贝叶斯网基本概念、贝叶斯网推理、贝叶斯网学习,以及贝叶斯网在中医中的应用四大部分。本书从实例出发,由浅入深,直观与严谨相结合,并提供了详尽的参考文献。本书的读者对象是相关专业的高年级本科生、研究生和科研人员。

图书在版编目(CIP)数据

贝叶斯网引论/张连文,郭海鹏著. —北京:科学出版社,2006
ISBN 7-03-018170-0

I. 贝… II. ①张… ②郭… III. 贝叶斯推断 IV. O212

中国版本图书馆 CIP 数据核字(2006)第 125753 号

责任编辑:赵卫江 陈砾川/责任校对:赵 燕
责任印制:吕春珉/封面设计:耕者设计工作室

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>

双青印刷厂印刷

科学出版社发行 各地新华书店经销

*

2006年11月第一版 开本:B5(720×1000)

2006年11月第一次印刷 印张:19

印数:1--2 000 字数:360 000

定价:58.00元

(如有印装质量问题,我社负责调换(新欣))

销售部电话 010-62136131 编辑部电话 010-62138017(BI01)

序

这是一本系统论述由概率、统计与图论结合而发展起来的贝叶斯网的专著。贝叶斯网起源于 20 世纪 80 年代中期对人工智能中的不确定性问题的研究，已经成为人工智能的一个重要领域。近年来它的影响在国际上不断扩大，成为将概率、统计应用于复杂系统的不确定性推理和数据分析的一种有效工具。本书作者根据多年来在这方面的研究和与国际交流的体会，采取由实例出发、由浅入深、直观与严谨相结合等较易为读者理解的方式，对这一新的学科的基本理论、方法及其应用进行了系统的、较全面的介绍，列举了相当详尽的参考文献。我认为，本书在国内的出版是一桩在学术上有意义的事件，将促进国内对这一新兴而又有潜能的领域的进一步学习、研究和应用。作者希望我给本书写序，作为人工智能研究的一个热心观众，借这个机会谈一点有关人工智能以及它与数学、电子计算机（下面采用现在通用的称谓——电脑）的关系的粗浅看法，供有志于人工智能研究的读者参考。

人工智能的兴起大致源于 20 世纪 40 年代电脑的出现。人们一方面兴奋地运用电脑强大的计算功能，去完成以往无法完成以及新提出的庞大计算任务；另一方面，对电脑所具有的类似于人脑的某些功能，包括一些逻辑判断功能，产生了巨大兴趣，从而对它寄予了很高的期望，期望它能够在智能方面起作用。于是人们开始研究自动机的理论、人类智能的机理以及智能行为的计算模型，进而期望构造与人脑相类似的智能系统，这大概是人工智能的最初想法。在当时，已经有人认为要制造完全代替人脑的电脑是不可能的。50 年代，我从著名的数理逻辑界前辈胡世华先生那里知道了一种相当有说服力的说法（我认为现在看来也是正确的）：机械是人手的延伸，而电脑是人脑的延伸。机械常常能够更好、更快地完成人手所能完成的机械性任务，但是不能完全代替人手；同样地，电脑能够完成人脑所能完成的机械性任务，但是不可能完全代替人脑。当然这种看法丝毫没有贬低人工智能的意义，相反，人工智能是一个引起人们浓厚兴趣、已经有相当成就而又被人们寄予巨大希望的科学领域。

与探索和模拟人类智能这一长远目标相比，人工智能研究更多的是集中在针对实际具体问题的智能系统的研究和开发。专家系统就是一个例子，它的目标是将专家们对某一复杂领域的知识和经验引入电脑系统，使得一般人能够借助电脑系统得到对问题的专家级解答。60 年代，在国内由于肝炎的流行，人们研制出学习和模仿著名中医专家关幼波的诊治系统就是一例。正如本书的 1.4 节所表述

的, 人工智能的实质进展有赖于不断针对人类的某种智能行为, 运用数学理论和方法, 结合计算机技术来建立适当的数学计算模型. 只有智能的目标和计算机技术而没有数学的深层次介入是不可能具有显著进展的. 不确定性是人工智能所面临的一个重要课题, 因此概率论自然就成为处理人工智能问题的一个工具. 对于静态的概率问题的研究和计算, 理论上只需要一个联合概率分布就足够了. 但是联合概率分布的复杂度相对于变量个数呈指数增长, 以致当变量个数很大时, 计算复杂度极高而不可行. 虽有电脑的强大计算功能也会成为不可行. 贝叶斯网的提出就是要解决这个问题. 它用图论的语言直观揭示问题的结构, 又按照概率论的原则对问题的结构加以分析, 降低推理计算的复杂度, 使得人们能够将概率论应用于大型问题. 近年来贝叶斯网能够在众多领域得到广泛应用, 原因就在于此. 从贝叶斯网的发展, 我们同样也看到计算机技术的广泛应用有赖于数学的深层次介入.

对于概率论与数理统计的广泛应用, 人们有普遍的共识. 但是在很长一段时期, 人们普遍地关注和运用概率的频率解释, 对于贝叶斯观点比较忽视, 认为将概率看作一种主观信任程度是没有客观标准的. 20 世纪中、后期以来, 贝叶斯观点的应用逐步得到了认同, 贝叶斯方法再度兴起, 贝叶斯网就与这次潮流密切相关, 这些是概率论与数理统计的应用中一个值得重视的趋势. 现在看来, 主观、客观观念属于哲学范畴, 应用于具体的对象时, 需要根据实践来判断. 主观信度或在贝叶斯观点下的先验概率, 实际上可以认为是在本次试验或观察以前或以外的某种经验积累的表述. 所以, 对于一种数学观点和方法, 应该依据实践来认识、判断和应用, 这样才不至于因噎废食.

本书作者之一张连文 1986 年在北京师范大学汪培庄教授的指导下获得应用数学硕士学位, 接着在我们的概率统计学科点攻读博士, 作为我和 Kansas 大学的 Glenn Shafer 教授联合指导的博士生, 以论文《人工智能中不确定性的三个模型》通过论文答辩 (1989). 我当时觉得他比较年轻, 对人工智能有兴趣, 就建议他再在国外攻读一个计算机博士学位, 这样对以后研究人工智能会有好处. 1994 年, 他在加拿大不列颠哥伦比亚大学 David Poole 教授的指导下获得计算机博士学位, 论文标题是《决策网络的计算理论》. 这些年来, 他实际上一直围绕人工智能中不确定性问题而工作, 对贝叶斯网的发展做出了一系列的重要工作; 而且, 他还和中医学者合作, 应用贝叶斯网研究中医的辨证施治. 这次能在科学出版社出版《贝叶斯网引论》, 是向内地对他的培养的一项汇报和宝贵的回报. 我希望通过本书的出版, 使他能够和内地建立更好的学术交流, 帮助推动贝叶斯网及相关方面在内地的应用和发展.

严士健

于北京师范大学数学科学学院

前 言

贝叶斯网 (Bayesian networks) 是一种帮助人们将概率统计应用于复杂领域、进行不确定性推理和数据分析的工具。它起源于人工智能领域的研究, 近年来对众多其它领域也产生了重要影响。本书系统地介绍贝叶斯网的基本理论和方法, 为读者了解和进入这个新兴领域提供一条相对平坦的途径。

从技术层面上讲, 贝叶斯网是一种系统地描述随机变量之间关系的语言。构造贝叶斯网的主要目的是进行概率推理, 即计算一些事件发生的概率。要在一些随机变量之间进行概率推理, 理论上只需要一个联合概率分布即可。但是, 联合概率分布的复杂度相对于变量个数成指数增长, 所以当变量众多时不可行。贝叶斯网的提出就是要解决这个问题, 它把复杂的联合概率分布分解成一系列相对简单的模块, 从而大大降低了知识获取的难度和概率推理的复杂度, 使得人们可以把概率论应用于大型问题。

贝叶斯网是概率论与图论相结合的产物, 它一方面用图论的语言直观揭示问题的结构, 另一方面又按照概率论的原则对问题的结构加以利用, 降低推理的计算复杂度。近年来, 贝叶斯网之所以能够在众多不同领域得到广泛应用, 其根本原因就在于此。另外, 由于贝叶斯网直观易懂, 它也是学者们喜爱的讨论和交流工具。

统计学、系统工程、信息论以及模式识别等学科中许多经典的多元概率模型都是贝叶斯网的特例, 包括朴素贝叶斯模型 (naïve Bayes models) (Titterton et al., 1981)、隐类模型 (latent class models) (Lazarsfeld and Henry, 1968; Goodman, 1974)、混合模型 (mixture models) (Everitt and Hand, 1981; McLachlan and Basford, 1988)、隐马尔可夫模型 (hidden Markov models) (Baum and Petrie, 1966)、卡尔曼滤波器 (Kalman filters) (Kalman, 1960) 等等。贝叶斯网为这些模型提供了一个共同的框架, 使得在一个领域获得的结果可以推广到其它领域。更重要的是, 它也为发展新模型提供了一个自然的框架。例如, 动态贝叶斯网 (dynamic Bayesian networks) (Dean and Kanazawa, 1990; Russell and Norvig, 2003) 就是最近几年在这个框架下发展起来的一类新模型, 它主要用于对多维离散时间序列的监控和预测。另一个例子是多层隐类模型 (hierarchical latent class models) (见第 9 章), 它是对隐类模型的推广, 能够揭示观测变量 (observed variables) 背后的隐结构。

构造贝叶斯网的方法因问题而定。有时, 网络结构和参数由应用问题的定义

所决定。例如，在解码问题中（见 2.6.2 节），网络结构完全取决于编码器的设计，而网络参数则由编码器的设计和信道的特征所决定。在专家系统中，贝叶斯网的结构和参数往往通过咨询专家来获得。近年来，贝叶斯网越来越多地被用于数据分析，成了数据分析的工具。统计学视之为图模型的一种，而人工智能学科把从数据出发获得贝叶斯网的过程视为是机器学习的一个特例，称为贝叶斯网学习。

在国际上，贝叶斯网已经成为人工智能和机器学习教材的重要内容，同时还有多本专著已经或即将出版，这些均凸显贝叶斯网受到的重视。我们从事贝叶斯网研究多年，觉得有责任把它系统地介绍到国内，于是决定写这本书。

本书分四个部分。第一部分介绍贝叶斯网基础，包括第 1、2、3 章。第 1 章回顾一些概率论及信息论中与贝叶斯网密切相关的概念和结果。第 2 章以降低概率推理的复杂度为出发点，引入贝叶斯网的概念，介绍如何手工构造贝叶斯网，并给出诸多应用实例。第 3 章讨论贝叶斯网的图论侧面与它的概率论侧面之间的密切关系。

第二部分介绍贝叶斯网推理，包括第 4、5、6 章。第 4 章揭示贝叶斯网推理的原理，并给出最基本的推理算法，即变量消元算法。第 5 章介绍另一个推理算法，即团树传播算法。与变量消元法相比，团树传播法的主要优点是它使得两次不同推理的中间结果可以共享。因此，当需要做多次推理时，团树传播法比变量消元法更为合适。第 6 章介绍近似推理算法，包括随机抽样法和变分法以及最新的研究进展。

第三部分讨论贝叶斯网学习，包括第 7、8、9 章。第 7 章假设已知网络结构，讨论参数估计方法。第 8 章探讨在不知道网络结构的情况下，如何通过数据分析同时确定贝叶斯网的结构和参数。第 9 章假设观测变量背后有一个隐结构，研究如何通过数据分析对这个隐结构进行推测。

第四部分只有一章，即第 10 章，它介绍的是贝叶斯网在中医研究中的一个应用实例。贝叶斯网的应用有很多，之所以选用中医的例子是因为贝叶斯网有机会对中医现代化起重要推动作用，而中医也为贝叶斯网研究提出了一些有趣的新问题。

本书的读者对象是相关专业的高年级本科生、研究生和科研人员。我们选择贝叶斯网最核心的内容，基于我们自己的体会将其组织起来，因此与国际上现有的专著相比，本书更适合作为贝叶斯网的入门教材。当然，本书也可以用于自学，书中不仅详细阐述贝叶斯网的基本理论，还简要介绍它的各种应用实例，综述最新研究进展，并提供大量的参考文献以便读者进一步深入学习。在注意理论深度和严谨性的同时，我们也注重理论背后的直观含意以及如何将理论应用于实际。

本书的筹划始于2002年秋。那时，张连文在香港科技大学开设了一门关于贝叶斯网的研究生课程，备课时已为将来把讲稿扩充成书做了许多考虑。具体写作开始于2003年秋，初稿完成于2005年夏。陈骏和王斌几次仔细阅读初稿，提出了许多改进意见，并参与了一些章节的写作，我们表示衷心感谢。严士健教授是张连文的博士导师，长期关心和支持张连文的工作，这次又拨冗为本书写序，在此特表感激之意。本书的写作得到香港研究资助局项目622105、国家重点基础研究发展计划（973计划）No. 2003CB517101，以及香港科技大学博士后基金的资助，本书的出版得到中国科学院科学出版基金的资助，史忠植教授、刘际明教授提供了协助，陈砺川、赵卫江女士做了大量细致的编辑工作，这里一并致谢。

由于我们水平有限，书中难免有疏漏和错误，请读者不吝指正。错误更正会在本书的网页（<http://www.cs.ust.hk/bnbook/>）上公布。在此网页上，读者还可以找到一些关于贝叶斯网的信息，如相关学术组织、学术会议、学术期刊、学术团队以及常用软件等。

张连文 郭海鹏

2006年春于香港清水湾

目 录

第一部分 贝叶斯网基础

第 1 章 概率论基础	3
1.1 随机事件与随机变量.....	3
1.2 概率的解释.....	6
1.2.1 古典解释.....	6
1.2.2 频率解释.....	6
1.2.3 主观解释.....	7
1.2.4 特性解释与逻辑解释.....	9
1.3 多元概率分布.....	10
1.3.1 联合概率分布.....	10
1.3.2 边缘概率分布.....	11
1.3.3 条件概率分布.....	13
1.3.4 边缘独立与条件独立.....	16
1.3.5 贝叶斯定理.....	18
1.4 概率论与人工智能.....	20
1.5 信息论基础.....	21
1.5.1 Jensen 不等式.....	22
1.5.2 熵.....	24
1.5.3 联合熵、条件熵和互信息.....	25
1.5.4 相对熵.....	28
1.5.5 互信息与变量独立.....	29
第 2 章 贝叶斯网	31
2.1 不确定性推理与联合概率分布.....	31
2.2 条件独立与联合分布的分解.....	33
2.3 贝叶斯网的概念.....	34
2.4 贝叶斯网的构造.....	36
2.4.1 确定网络结构.....	36

2.4.2	因果关系与贝叶斯网	39
2.4.3	确定网络参数	41
2.5	贝叶斯网的应用	45
2.5.1	医疗诊断	45
2.5.2	工业应用	48
2.5.3	金融分析	48
2.5.4	计算机系统	49
2.5.5	军事应用	50
2.5.6	生态学	51
2.5.7	农牧业	53
2.6	贝叶斯网对其它领域的影响	53
2.6.1	生物信息学	53
2.6.2	编码学	57
2.6.3	机器学习	61
2.6.4	时序数据和动态模型	62
2.7	文献介绍	65
第3章	图分隔与变量独立	66
3.1	直观分析	66
3.1.1	基本情况	66
3.1.2	一般情况	68
3.2	有向分隔与条件独立	70
3.2.1	几个引理	70
3.2.2	马尔可夫性	71
3.3	有向分隔与无向分隔	74
3.4	有向无圈图与联合概率分布	77
3.5	文献介绍	78

第二部分 贝叶斯网推理

第4章	贝叶斯网与概率推理	81
4.1	推理问题	81
4.1.1	后验概率问题	81
4.1.2	最大后验假设问题	82
4.1.3	最大可能解释问题	83

4.2	变量消元算法	83
4.2.1	概率分布的分解与推理复杂度	83
4.2.2	消元运算	85
4.2.3	算法描述	86
4.2.4	一个例子	88
4.3	复杂度分析	89
4.3.1	复杂性的度量	89
4.3.2	复杂度的计算	90
4.4	消元顺序	94
4.4.1	最大势搜索	95
4.4.2	最小缺边搜索	95
4.5	推理问题简化	97
4.6	MAP 假设问题	99
4.6.1	两个运算	100
4.6.2	分解与计算复杂度	102
4.6.3	变量消元 MAP 算法	102
4.7	文献介绍	105
第 5 章	团树传播算法	106
5.1	团树	106
5.2	一个变量后验概率的计算	107
5.3	团树传播的正确性	111
5.4	团树传播与计算共享	113
5.5	每个变量的后验概率的计算	115
5.6	团树的构造	118
5.6.1	图消元与团树	118
5.6.2	图消元构造团树算法的正确性	121
5.6.3	极小团树	122
5.6.4	t -团与 g -团	123
5.7	文献介绍	124
第 6 章	近似推理	125
6.1	随机抽样算法	125
6.1.1	重要性抽样法	125
6.1.2	MCMC 抽样	131
6.2	变分法	133

6.2.1 朴素平均场法	134
6.2.2 循环传播算法	136
6.3 其它近似推理算法	138
6.4 文献介绍	138

第三部分 贝叶斯网学习

第7章 参数学习	143
7.1 贝叶斯网与数据分析	143
7.2 单参数最大似然估计	144
7.3 单参数贝叶斯估计	146
7.4 单变量网络参数估计	150
7.5 一般网络最大似然估计	151
7.5.1 最大似然估计的计算	152
7.5.2 最大似然估计的性质	155
7.6 一般网络贝叶斯估计	157
7.7 缺值数据最大似然估计	160
7.7.1 EM算法的基本思想	161
7.7.2 EM算法的基本理论	163
7.7.3 EM算法	165
7.7.4 EM算法的收敛性	166
7.8 缺值数据贝叶斯估计	168
7.9 文献介绍	171
第8章 结构学习	172
8.1 似然函数与模型选择	172
8.2 贝叶斯模型选择	175
8.3 大样本模型选择	178
8.4 其它模型选择标准	180
8.5 模型优化	182
8.5.1 评分函数的分解	183
8.5.2 K2算法	184
8.5.3 爬山法	186
8.6 缺值数据结构学习	188
8.6.1 SEM算法的基本思想	188

8.6.2	SEM算法	189
8.6.3	SEM的收敛性	191
8.7	文献介绍	192
第9章	隐结构模型学习	194
9.1	隐变量与隐变量模型	194
9.2	可分辨性及几个相关概念	195
9.3	隐变量模型选择	196
9.4	隐类模型	197
9.4.1	正则性	198
9.4.2	隐变量模型选择与大样本理论	199
9.4.3	隐类模型学习算法	200
9.4.4	仿真试验	201
9.5	多层隐类模型	204
9.5.1	走根运算与模型等价	205
9.5.2	无根 HLC 模型	206
9.5.3	正则性	207
9.5.4	正则模型空间的有限性	209
9.6	多层隐类模型学习算法	211
9.6.1	势学习算法	211
9.6.2	模型学习算法	212
9.6.3	复杂度分析	216
9.6.4	仿真试验	217
9.7	文献介绍	220

第四部分 贝叶斯网应用

第10章	隐结构模型与中医辨证	225
10.1	中医辨证的客观化、量化	225
10.1.1	相关工作回顾	225
10.1.2	隐结构法	227
10.2	肾虚数据收集	229
10.3	数据分析原理	232
10.4	肾虚数据分析	234
10.5	结果模型定性内容的质量	236

10.6	结果模型定量内容的质量	239
10.7	结果模型与辩证论治	250
10.8	模型辩证的质量	252
10.9	讨论	260
参考文献		263
英汉词汇对照		278
索引		287

第一部分
贝叶斯网基础

第 1 章 概率论基础

本章简要介绍概率论中与贝叶斯网密切相关的一些基本概念. 与一般的概率论教材不同, 我们将侧重于概念的直观含意. 深入理解诸如条件独立、贝叶斯定理等概念和结果, 对以后理解贝叶斯网至关重要. 另外, 本章还将简述概率方法在人工智能研究中的崛起过程, 并介绍在第三部分中会被用到的一些信息论知识.

1.1 随机事件与随机变量

世界上许多事情都具有不确定性. 例如掷硬币, 其结果可能正面朝上, 也可能反面朝上, 在抛掷之前无法预知. 又如赌马, 理论上每匹马都有跑第一的可能, 事先无法预料哪匹马一定会赢. 再如火星上是否曾有生命存在? 答案有两种可能, 是或不是, 但根据目前掌握的证据判断, 无法给出绝对的答复. 概率论是研究处理这类现象的数学理论. 本节介绍概率论中几个最基本的概念, 包括样本空间、事件、概率测度、随机变量以及概率函数.

1. 样本空间和事件

在概率论中, **随机试验**指的是事先不能完全预知其结果的试验. 随机试验的所有可能结果组成该试验的**样本空间**, 通常记为 Ω . 样本空间可以是离散的, 也可以是连续的. 如无特殊说明, 本书所论及的样本空间都将是离散的.

样本空间中的点, 即随机试验的可能结果, 称为**样本点**, 或**原子事件**, 记为 ω . 样本空间的子集称为**事件**, 通常用大写字母表示: A, B, \dots . 如果随机试验的结果包含在一个事件之中, 则称该事件发生了. 样本空间 Ω 本身也是一个事件, 而且是一定会发生的**必然事件**. 空集也是一个事件, 是**不可能事件**, 通常记为 \emptyset . 事件之间可以进行交 (\cap)、并 (\cup)、差 (\setminus) 等各种集合运算. 若两事件 A 和 B 交空, 即 $A \cap B = \emptyset$, 则称它们为**互斥事件**, 又称**不相容事件**. 两互斥事件不能同时发生. 若 A 和 B 互斥, 且 $A \cup B = \Omega$, 则称它们为**互补事件**.

例 1.1 考虑掷硬币试验, 其结果有正反面两种可能, 因此样本空间为 $\Omega = \{h, t\}$, 其中 h 表示正面, t 表示反面, h 和 t 为两个互补的原子事件. □

例 1.2 考虑掷骰子试验, 有 6 种可能的结果, 样本空间为 $\Omega = \{1, 2, 3, 4, 5, 6\}$. 子集合 $\{1, 3, 5\}$ 表示的是“掷出结果为奇数”这一事件, 其互补事件为 $\{2, 4, 6\}$, 即“掷出结果为偶数”. □

例 1.3 考虑从香港科技大学的所有研究生中随机抽样的试验, 其结果可能