王双成 著

贝叶斯网络

学习、推理与应用

立信会计出版社
LIXIN KUAIJI CHUBANSHE

贝叶斯网络学习、推理与应用

王双成　著

立信会计出版社

图书在版编目(CIP)数据

贝叶斯网络学习、推理与应用/王双成著. —上海:立信会计出版社,2010.2

ISBN 978-7-5429-2470-4

Ⅰ. ①贝… Ⅱ. ①王… Ⅲ. ①贝叶斯推断 Ⅳ. ①O212

中国版本图书馆 CIP 数据核字(2010)第 027142 号

责任编辑 赵志梅
封面设计 周崇文

贝叶斯网络学习、推理与应用

出版发行	立信会计出版社		
地 址	上海市中山西路 2230 号	邮政编码	200235
电 话	(021)64411389	传 真	(021)64411325
网 址	www.lixinaph.com	E-mail	lxaph@sh163.net
网上书店	www.shlx.net	Tel:	(021)64411071
经 销	各地新华书店		

印 刷	上海申松立信印刷有限责任公司
开 本	890 毫米×1240 毫米 1/32
印 张	9.375
字 数	254 千字
版 次	2010 年 2 月 第 1 版
印 次	2010 年 2 月 第 1 次
书 号	ISBN 978-7-5429-2470-4/O
定 价	20.00 元

如有印订差错,请与本社联系调换

内容简介

贝叶斯网络是概率理论与图形理论的结合，围绕的一个基本问题是联合概率计算。基于贝叶斯网络可进行联合概率的条件和边缘分解，从而有效降低运算复杂性，并解决与联合概率计算有关的一系列问题。贝叶斯网络已在许多领域得到了广泛的应用，是不确定性知识表示和推理的有力工具。

本书按照贝叶斯网络基础、学习、推理、集成和应用的框架介绍贝叶斯网络的相关理论、方法和算法，有助于读者对贝叶斯网络理论体系的认识和理解，可供相关专业的高年级本科生、研究生和科研人员学习与参考。

前　　言

贝叶斯网络(Bayesian networks)是描述随机变量之间依赖关系的图形模式，被广泛用于不确定性问题的智能化求解。它具有多功能性、有效性和开放性(是一个能够集成其它智能技术与数据处理方法的平台)等特征，能够有效地转化数据为知识(具有形象直观的知识表示形式)，并利用这些知识进行推理(具有类似于人类思维的推理方式)，以解决分析、预测和控制等方面的问题。其有效性已在风险管理、信息融合、医疗诊断、系统控制和生物信息分析等许多领域得到验证。

自从 20 世纪 80 年代后期加利福尼亚大学计算机科学系 Pearl (1988)给出贝叶斯网络的严格定义并创建贝叶斯网络基础理论体系以来，贝叶斯网络获得了长足的发展。这些研究主要从贝叶斯网络学习、推理、集成和应用四个方面展开，出现了许多经典的方法和算法，也解决了大量的实际问题。

本书共分五个部分。

第一部分是贝叶斯网络基础，包括第 1、第 2、第 3 章。第 1 章介绍在贝叶斯网络研究中经常使用的一些概率公式和方法。第 2 章从概率模式、图形模式和它们之间联系的视角简要阐述贝叶斯网络的基础理论。第 3 章给出贝叶斯网络学习和推理中可能用到的一些量化方法和标准。

第二部分是贝叶斯网络学习，包括第 4 章至第 10 章。这几章分别从具有完整数据、丢失数据、隐藏变量、连续变量、噪声和小数据集等情况给出了一系列贝叶斯网络学习方法，以及随环境变化的贝叶斯网络更新算法。

第三部分是贝叶斯网络推理，包括第 11、第 12 章。第 11 章从贝叶斯网络信念更新和信念修正两个方面简要介绍经典的准确和近似推

理方法。第12章给出一系列贝叶斯网络分类器(分类预测推理)。

第四部分是贝叶斯网络集成,包括第13章至第16章。这几章介绍将贝叶斯网络与因果理论、决策理论、可能理论和时序过程相结合而得到的因果贝叶斯网络、决策贝叶斯网络(影响图)、可能贝叶斯网络(可能网)和动态贝叶斯网络。

第五部分是贝叶斯网络应用,包括第17、第18章。第17章介绍基于贝叶斯网络的聚类分析方法。第18章给出贝叶斯网络在预警和评估等方面的应用。

本书是作者在多年从事贝叶斯网络研究基础上整理而成的,其撰写和出版得到国家自然科学基金(60675036)、上海市教委重点学科基金(J51702)和上海市教委科研创新重点项目(09zz202)的资助。

王双成

2009年11月于上海立信会计学院

目　录

第一部分　贝叶斯网络基础

第二部分　贝叶斯网络学习

第三部分　贝叶斯网络推理

第四部分　贝叶斯网络集成

第 一 部 分

贝叶斯网络基础

第1章 概率论基础

概率论是贝叶斯网络的理论基础，也正是为解决概率论中联合概率计算这一基本问题产生了贝叶斯网络。在贝叶斯网络理论体系中经常涉及一些概率计算公式、贝叶斯方法和贝叶斯概率等内容。

1.1 概率计算公式

条件概率、乘法公式、链式规则、全概率公式和贝叶斯公式是贝叶斯网络理论中经常用到的概率公式，下面简要介绍这些公式。

1.1.1 条件概率、乘法公式和链式规则

1）条件概率

定义1.1 设A,B是E（E为基本事件集）的两个事件，且$p(A)>0$，则称

$$p(B|A)=\frac{p(AB)}{p(A)} \tag{1.1}$$

为事件A发生的条件下，事件B的条件概率。

同理，当$p(B)>0$时，称

$$p(A|B)=\frac{p(AB)}{p(B)} \tag{1.2}$$

为事件B发生的条件下，事件A的条件概率。

2）乘法公式

设A,B是E的两个事件，若$p(A)>0$，则称

$$p(AB)=p(A)p(B|A)\text{或 }p(AB)=p(B)p(A|B) \tag{1.3}$$

为概率乘法公式。

3) 链式规则

事件形式链式规则：

对事件 $A_1, A_2, \cdots, A_n$，且 $p(A_1) \geqslant p(A_1A_2) \geqslant \cdots \geqslant p(A_1A_2\cdots A_{n-1}) > 0$，可得到推广的乘法公式

$$p(A_1A_2\cdots A_n)=p(A_1)p(A_2|A_1)p(A_3|A_1,A_2)\cdots p(A_n|A_1A_2\cdots A_{n-1})。\quad (1.4)$$

公式(1.4)又称为事件形式的链式规则。

随机变量形式链式规则：

$$\begin{aligned}p(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)=&p(X_1=x_1)p(X_2=x_2|X_1=x_1)\cdots\\&p(X_n=x_n|X_1=x_1,X_2=x_2,\cdots,X_{n-1}=x_{n-1}),\end{aligned}\quad (1.5)$$

或简写为 $p(x_1,x_2,\cdots,x_n)=p(x_1)p(x_2|x_1)\cdots p(x_n|x_1,x_2,\cdots,x_{n-1})$。

1.1.2　全概率公式

事件形式：

设 $B_1,B_2,\cdots,B_n$ 是 E 的事件，且满足① $\bigcup_{i=1}^{n}B_i=\Omega$（$\Omega$ 为样本空间）；② $B_iB_j=\phi, i\neq j$，则称

$$p(A)=\sum_{i=1}^{n}p(B_i)p(A|B_i)\quad (1.6)$$

为全概率公式。

离散随机变量形式：

$$p(Y=y)=\sum_{i=1}^{n}p(X_i=x_i)p(Y=y|X_i=x_i),\quad (1.7)$$

或简写为 $p(y)=\sum_{i=1}^{n}p(x_i)p(y|x_i)$。

连续随机变量形式：

$$p(y)=\int p(y\mid x)p(x)dx。\quad (1.8)$$

1.1.3　贝叶斯公式

事件形式：

设 $B_1,B_2,\cdots,B_n$ 是 E 的一个互不相容的完备事件组，且 $p(B_i)>$

0,A 是 E 的任意事件,则称

$$p(B_i \mid A)=\frac{p(B_i)p(A \mid B_i)}{\sum_{j=1}^{n}p(B_j)p(A \mid B_j)} \tag{1.9}$$

为贝叶斯公式。

离散随机变量形式:

$$p(X_i=x_i \mid Y=y)=\frac{p(X_i=x_i)p(Y=y \mid X_i=x_i)}{\sum_{j=1}^{n}p(X_j=x_j)p(Y=y \mid X_j=x_j)}, \tag{1.10}$$

或简写为 $p(x_i \mid y)=\dfrac{p(x_i)p(y \mid x_i)}{\sum_{j=1}^{n}p(x_j)p(y \mid x_j)}$。

连续随机变量形式:

假设随机变量 ξ,η 的联合分布密度是 $p(x,y)=p_\xi(x)f_{\eta|\xi}(y \mid x)$,其中 $p_\xi(x)$ 是 ξ 的边缘密度,而 $f_{\eta|\xi}(y \mid x)$ 是当 $\xi=x$ 时,η 对 ξ 的条件密度,于是 ξ 对 η 的条件密度 $g_{\xi|\eta}(x \mid y)$ 可表示为:

$$g_{\xi|\eta}(x \mid y)=\frac{p(x)f_{\eta|\xi}(y \mid x)}{\int p_\xi(x)f_{\eta|\xi}(y \mid x)dx}。 \tag{1.11}$$

类似地,有

$$f_{\eta|\xi}(y \mid x)=\frac{q(y)g_{\xi|\eta}(y \mid x)}{\int q_\eta(x)g_{\xi|\eta}(x \mid y)dy}, \tag{1.12}$$

其中 $q_\eta(y)$ 是 y 的边缘分布密度。

1.2 贝叶斯方法

张尧庭和陈汉峰(1991)对贝叶斯方法归纳如下:

(1) 将未知参数看成随机变量(或随机向量),记它为 $\boldsymbol{\theta}$,于是当 $\boldsymbol{\theta}$ 已知时,样本 $x_1,\cdots,x_n$ 的联合分布密度 $p(x_1,\cdots x_n;\boldsymbol{\theta})$ 就看成是 $x_1,\cdots,x_n$ 对 $\boldsymbol{\theta}$ 的条件密度,记为 $p(x_1,\cdots,x_n|\boldsymbol{\theta})$或简写为 $p(x|\boldsymbol{\theta})$。

(2) 用 $\pi(\boldsymbol{\theta})$ 表示 $\boldsymbol{\theta}$ 的先验分布，一般根据以往对参数 $\boldsymbol{\theta}$ 的知识来确定先验分布(经常用于表示先验知识)，如果没有关于先验参数 $\boldsymbol{\theta}$ 的先验知识，$\pi(\boldsymbol{\theta})$ 应采用在 $\boldsymbol{\theta}$ 取值范围内的均匀分布(也称为**贝叶斯假设**)，这是贝叶斯方法中容易引起争议的一步。

(3) 利用条件分布密度 $p(x_1,\cdots,x_n|\boldsymbol{\theta})$ 和先验分布 $\pi(\boldsymbol{\theta})$，可以求出 $x_1,\cdots,x_n$ 与 $\boldsymbol{\theta}$ 的联合分布和样本 $x_1,\cdots,x_n$ 的分布，于是就可用它们求得 $\boldsymbol{\theta}$ 对 $x_1,\cdots,x_n$ 的条件分布密度，也就是用贝叶斯公式求得后验分布密度 $p(\boldsymbol{\theta}|x_1,\cdots,x_n)$。

(4) 利用后验分布密度 $p(\boldsymbol{\theta}|x_1,\cdots,x_n)$ 作出对 $\boldsymbol{\theta}$ 的推断(估计 $\boldsymbol{\theta}$ 或对 $\boldsymbol{\theta}$ 作检验)，则

$$p(\boldsymbol{\theta}|x_1,\cdots,x_n)=\frac{\pi(\boldsymbol{\theta})p(x_1,\cdots,x_n|\boldsymbol{\theta})}{p(x_1,\cdots,x_n)}, \tag{1.13}$$

其中 $p(x_1,\cdots,x_n)=\int\pi(\boldsymbol{\theta})p(x_1,\cdots,x_n \mid \boldsymbol{\theta})d\boldsymbol{\theta}$。

1.3 贝叶斯概率

简单地说，贝叶斯概率是观测者对某一事件发生的信任程度(一般称为主观概率，相对而言，经典概率称为客观概率或物理概率)。观测者根据先验知识和现有的统计数据，用概率的方法来预测未知事件发生的可能性。贝叶斯概率不同于事件的客观概率，客观概率是多次重复试验，然后统计事件发生的频率，而贝叶斯概率则是利用现有的知识对未知事件的预测。

记 $D=\{X_1=x_1,X_2=x_2,\cdots,X_m=x_m\}$ 为重复 m 次试验所得的观测样本。其中 X 为事件变量，x 为变量的值或状态，记参数 θ 为事件 $X=x$ 发生的先验概率[$\theta=p(x\mid\xi)$]，$p(\theta\mid\xi)$ 为它的概率密度函数，其中 ξ 为观测者的先验知识。现在贝叶斯概率的计算问题可以表述如下：已知先验概率 $p(\theta\mid\xi)$ 和样本 D，求第 $m+1$ 次试验中的事件 $X_{m+1}=x_{m+1}$ 发生的概率 $p(X_{m+1}=x_{m+1}\mid D,\xi)$。

由全概率公式得：

$$p(X_{m+1}=x_{m+1}\mid D,\xi)=\int p(X_{m+1}=x_{m+1}\mid\theta,D,\xi)p(\theta\mid D,\xi)d\theta=$$

$$\int\theta p(\theta\mid D,\xi)d\theta=E_{p(\theta\mid D,\xi)}(\theta)。\tag{1.14}$$

这表明，事件 $X_{m+1}=x_{m+1}$ 发生的贝叶斯概率为先验概率 θ 相对于后验概率分布 $p(\theta\mid D,\xi)$ 的期望值。根据贝叶斯公式，由先验概率 $p(\theta\mid\xi)$ 计算后验概率 $p(\theta\mid D,\xi)$ 的公式为：

$$p(\theta\mid D,\xi)=\frac{p(\theta\mid\xi)p(D\mid\theta,\xi)}{p(D\mid\xi)}=\frac{p(\theta\mid\xi)p(D\mid\theta,\xi)}{\int p(\theta\mid\xi)p(D\mid\theta,\xi)d\theta}。\tag{1.15}$$

在先验概率 $p(\theta\mid\xi)$ 已知的情况下，样本 D 中的各事件 $X=x$ 条件独立。如果事件变量 X 为二元分布，即事件只有发生或不发生两种情况，则 $p(D\mid\theta,\xi)=\theta^h(1-\theta)^t$，其中 h 为样本 D 中事件发生的次数，$h+t=m$。现在设先验概率为 β 分布，即

$$p(\theta\mid\xi)=Beta(\theta\mid\alpha_h,\alpha_t)=\frac{\Gamma(\alpha)}{\Gamma(\alpha_h)\Gamma(\alpha_t)}\theta^{\alpha_h-1}(1-\theta)^{\alpha_t-1},\tag{1.16}$$

其中 $\alpha_h>0,\alpha_t>0$ 为 β 分布的参数，$\alpha=\alpha_h+\alpha_t$。Γ 是伽玛(Gamma)函数，$\Gamma(x+1)=x\Gamma(x)$，$\Gamma(1)=1$。由上面的三个式子可得后验概率 $p(\theta\mid D,\xi)$ 也为 β 分布：

$$p(\theta\mid D,\xi)=Beta(\theta\mid\alpha_h+h,\alpha_t+t)=$$

$$\frac{\Gamma(\alpha+m)}{\Gamma(\alpha_h+h)\Gamma(\alpha_t+t)}\theta^{\alpha_h+h-1}(1-\theta)^{\alpha_t+t-1},\tag{1.17}$$

β 分布的期望值是已知的，即

$$\int\theta Beta(\theta\mid\alpha_h,\alpha_t)d\theta=\frac{\alpha_h}{\alpha},$$

于是，预测事件的贝叶斯概率为：

$$p(X_{m+1}=x_{m+1}\mid D,\xi)=\int\theta p(\theta\mid D,\xi)d\theta=$$

$$\int\theta Beta(\theta\mid\alpha_h+h,\alpha_t+t)d\theta=\frac{\alpha_h+h}{\alpha+m}。\tag{1.18}$$