Bayesian 网中概率参数学习方法

薛万欣[1] ,刘大有[2] ,张 弘[2]
(1. 北京联合大学管理学院 ,北京 100101 ; 2. 吉林大学计算机科学技术学院 ,吉林长春 130021)

摘 要: Bayesian 网已经成为 AI 领域的研究热点 ,并在现代专家系统、诊断系统及决策支持系统中发挥着至关重要的作用. Bayesian 网的研究主要集中在三个方面 :知识表示、学习与推理. 概率知识是 Bayesian 网坚实的数学基础 ,从数据中学习分布参数使得 Bayesian 网逐步走向现实应用. 本文介绍和比较了概率参数学习的各种常用方法 ,并探求了它们在不同应用背景下的优缺点. 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合所提供的信息 ,无法加入专家知识 ,对实例数据的依赖性大 ;基于 Bayesian 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 Bayesian 网学习的基础 ,是 Bayesian 网结构学习必不可少的部分.

关键词: Bayesian 网; 条件概率; 参数分布
中图分类号: TP181 **文献标识码**: A **文章编号**: 0372-2112 (2003) 11-1686-04

Learning with a Bayesian Networks a Set of Conditional Probility Tables

XUE Wan-xin[1] ,LIU Da-you[2] ,ZHANG Hong[2]
(1. *Department of Management Bei Jing Union University , Beijing* 100101 ;
2. *Department of Computer Science , JI LIN University , Changchun , Jilin* 130021 , *China*)

Abstract : Bayesian network is becoming more remarkable in AI research fields ,which plays important role in modern expert system ,diagnoses system and decision system. Bayesian network works on three points as below :knowledge representation ,learning and inference. Prababilistic methods are its mathematical fundamental which helps learning distribution from data and leads Bayesian theory to real application. This paper introduces various common methods in probability data learning and make comparison among them under various application background. The methods based on classical statistics have a matured theory and a set of simple and direct calculation. But they reply heavily on sample data ,which apply only those information from sample data with expert knowledge left aside. Bayesian Network combines information of expert knowledge and sample data together. It can give more accurate learning result and rely less on sample data. Parameter learning is main part of learning Bayesian Network models ,and it 's the basis of Bayesian Network learning.

Key words : bayesian networks ;conditional probability ;parameter distribution

1 引言

理论和实践中的许多问题都可以通过 Bayesian 网建模实现 ,目前典型的 Bayesian 网经常需要几十或上百个变量 ,需要成百上千的概率参数 ,概率参数的获取不是一次过程所能解决的 ,需要的大量概率参数是阻碍 Bayesian 网实际应用的原因. 从实际数据中归纳、学习参数分布是 Bayesian 网实际应用必须解决的一个问题.

2 Bayesian 网

定义(Bayesian 网) Bayesian 网是一个二元组 $S = \langle G, P\rangle$,其中 :

(1) G 是一个有向非循环图 ,网中节点与知识领域的随机变量一一对应;网中的有向弧表示变量间的因果关系 ,从节点 X 到节点 Y 有向弧的直观含义是 X 对 Y 有直接的因果影响;

(2) $P = \{ p(X| \pi_X) \}$条件概率表示因果影响的强度 ,其中 π_X 代表节点 X 的父亲节点集合. 可由领域专家给出 ,但条件概率表都比较大 ,专家很难逐项给出 ,实际上获取概率参数既复杂又费时 ,这使从大量样本数据中发现 Bayesian 网 ,即解决 Bayesian 网的学习问题变得很有意义.

引理 Bayesian 网是概率信息的载体 ,是联合概率分布的图形表示方式. 给定 Bayesian 网 T,则存在一个离散变量集合 $X = \{ X_1 , X_2 , \cdots, X_n\}$上的联合概率分布 P, X 的变量和 T 中的节点间存在一一对应关系 ,并且 P 中存在递归展开式 : $p(X_1 , \cdots, X_n) = \prod_i p(X_i| \pi_{X_i})$,其中 π_{X_i}是 X_i 在 T 中父亲节点集合.

从一个应用领域中构筑 Bayesian 网涉及三个步骤 : (1) 必须分辨出所要建模领域中具有重要性的变量和这些变量的所有可能取值 ,并以节点表示 ; (2) 判断节点间的依赖或独立关

收稿日期 :2002-11-30 ;修回日期 :2003-06-16
基金项目 :国家 863 项目(No. 863-306-ZD05-01-2) ;国家自然科学基金(No. 69883003) ;吉林省科委项目(No. 20010588) ;教育部高校博士点专项科研基金项目 ;教育部符号计算与知识工程重点实验室

系,并以图方式表示;(3) 获得 Bayesian 网定量部分所需要的概率参数,概率参数的获取比较繁琐. Bayesian 网的构建一般在三个步骤间反复循环. Bayesian 网的学习分为结构和参数分布学习. Bayesian 网的结构通常可由领域专家给出.

3　参数分布学习方法

Bayesian 网的条件概率学习问题可以归结为统计学中的参数估计问题. 在大部分的实际应用领域中,概率信息可以通过多种方式获得,包括从统计数据中学习,从文献中查阅,或者咨询领域专家. 统计学中将参数估计问题从方法上分为两大类:基于经典统计学的方法和基于贝叶斯统计学的方法. 经典统计学,是基于对概率的频度性理解而建立起来的一系列理论. 贝叶斯统计学是统计学中的一个分支,与传统统计学的本质区别在于对概率的理解. 从观察的数据集学习概率参数,数据集分为完备数据集和不完备数据集,两种情况下学习的难易程度不一样.

3.1　完备数据集下概率参数的学习

随机数据集 $D=\{X[1],\cdots,X[m]\}$ 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.

3.1.1　MLE 方法　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 $C^1, C^2, \cdots, C^n$,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 $L(\theta:D)=p(D|\theta)=\prod_m p(X[m]|\theta)$,似然性越大,具体的 θ 越"好".

推广到含有 n 个变量的一般的贝叶斯似然性函数为:

$$\begin{aligned}L(\theta:D)&=\prod_m p(X_1[m],\cdots,X_n[m]|\theta)\\&=\prod_m\prod_i p(X_i[m]|\pi_{X_i}[m],\theta_i)\\&=\prod_i\prod_m p(X_i[m]|\pi_{X_i}[m],\theta_i)\\&=\prod_i L_i(\theta_i:D)\end{aligned}$$

在 $p(X_i|\pi_X)$ 是多项分布的一般 Bayesian 网中,局部似然函数可进一步分解:

$$\begin{aligned}L_i(\theta_i:D)&=\prod_m p(X_i[m]|\pi_{X_i}[m],\theta_i)\\&=\prod_m\prod_{\pi_{X_i}}\prod_{x_i^k} p(x_i^k[m]|\pi_{X_i}^j[m],\theta_i)\\&=\prod_{\pi_{X_i}}\prod_{x_i^k}\theta_{x_i^k|\pi_{X_i}}^{N(x_i^k,\pi_{X_i}^j)}\end{aligned}$$

因为数据集完备,在父亲集合 π_{X_i} 每种具体的取值 $\pi_{X_i}^j$ 下,$p(X_i|\pi_{X_i}^j)$ 分布是和 π_{X_i} 的其他取值 $\pi_{X_i}^l\ (l\neq j)$ 无关的独立多项分布问题,MLE 方法计算出 $\theta_{X_i^k}=\left[N\left(x_i^k,\pi_{X_j}^i\right)\right]\Big/\left[N\left(\pi_{X_j}^i\right)\right]$,利用该公式,可以很容易地计算出给定结构的条件概率表.

3.1.2　Bayesian 方法　Bayesian 方法利用 Bayesian 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合 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. Bayesian 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 Bayesian 预测法.

先验分布 $p(\theta)$ 取作 Dirichlet 分布:$p(\theta)=Dir(\theta|\alpha_1,\cdots,\alpha_r)=\dfrac{\Gamma(\alpha)}{\prod_{k=1}^r\Gamma(\alpha_k)}\prod_{k=1}^r\theta_k^{\alpha_k-1}$

其中 $\alpha=\sum_{k=1}^r\alpha_k$, $\alpha_k>0$, $k=1,\cdots,r$, $\alpha_1,\cdots,\alpha_r$ 称作超参数;$\Gamma(\cdot)$ 称为 Game 函数,$\Gamma(x)=\int_0^\infty t^{x-1}e^{-t}dt$,满足:$\begin{cases}\Gamma(x+1)=x\Gamma(x)\\\Gamma(1)=1\end{cases}$,那么,样本发生的概率为:

$$\begin{aligned}p(D)&=\int p(\theta)p(D|\theta)d\theta=\int\frac{\Gamma(\alpha)}{\prod_{k=1}^r\Gamma(\alpha_k)}\prod_{k=1}^r\theta_k^{\alpha_k-1}\times\prod_{k=1}^r\theta_k^{N_k}d\theta\\&=\frac{\Gamma(\alpha)}{\Gamma(\alpha+N)}\prod_{k=1}^r\frac{\Gamma(\alpha_k+N_k)}{\Gamma(\alpha_k)}\end{aligned}$$

参数 θ 的后验分布也是 Dirichlet 分布:

$$\begin{aligned}p(\theta|D)&=\frac{p(\theta)p(D|\theta)}{p(D)}=\frac{\dfrac{\Gamma(\alpha)}{\prod_{k=1}^r\Gamma(\alpha_k)}\prod_{k=1}^r\theta_k^{\alpha_k-1}\times\prod_{k=1}^r\theta_k^{N_k}}{\dfrac{\Gamma(\alpha)}{\Gamma(\alpha+N)}\prod_{k=1}^r\dfrac{\Gamma(\alpha_k+N_k)}{\Gamma(\alpha_k)}}\\&=\frac{\Gamma(\alpha+N)}{\prod_{k=1}^r\Gamma(\alpha_k+N_k)}\prod_{k=1}^r\theta_k^{\alpha_k+N_k-1}=Dir(\alpha_1+N_1,\cdots,\alpha_r+N_r)\end{aligned}$$

一般 Bayesian 网的边际似然性的计算公式为:

$$\begin{aligned}p(D|S)&=\int p(\theta_s)p(D|\theta_s)d\theta_s\\&=\prod_{i=1}^n\prod_{j=1}^{q_i}\frac{\Gamma(\alpha_{ij})}{\Gamma(\alpha_{ij}+N_{ij})}\prod_{k=1}^{r_i}\frac{\Gamma(\alpha_{ijk}+N_{ijk})}{\Gamma(\alpha_{ijk})}\end{aligned}$$

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,Bayesian 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.

3.2　不完备数据集下的学习

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 EM 和梯度上升算法及蒙特卡洛、高斯等方法[1].

3.2.1　蒙特·卡洛方法　Geman 等人提出的 Gibbs 取样方法是众多蒙特·卡洛方法中较有名的一种. 基本思想是:以 Gibbs 取样估算变量集 X 上联合概率分布 $p(X)$ 的数学期望 $f(X)$. 算法的具体步骤:(1) 初始化:随机初始化不完备的数据集 D 得到一个完备的数据集 D_c;(2) 在初始样本 D 中选取一个未观察到的变量 X_{il}(第 l 个事例的第 i 个变量),按照下面的概率分布给 X_{il} 的状态随机地赋值:$p(x'_{il}|D_c\setminus x_{il},S)=$

$\frac{p(x'_{il} \mid D_c \setminus x_{il}, S)}{\sum_{x''_{il}} p(x''_{il} \mid D_c \setminus x_{il}, S)}$，其中 $D_c \setminus x_{il}$ 表示 D_c 去掉观察 X_{il}；(3) 按照 2 逐个为 D 中每个事例中未观察到的变量赋值，直至产生一个新的随机完备数据集 D'_c；(4) 计算 $p(\theta_s \mid D, S)$；(5) 重复 (1) 到 (4) 足够多次，取众多 $p(\theta_s \mid D, S)$ 的平均值作为目标估计值.

蒙特·卡洛方法在其他方法不适用时，可以应用，非常灵活. 样本越大，运行时间越长，结果越精确，但计算复杂度是事例数目的指数幂.

Neal、Madigan 与 York 介绍了 Gibbs 抽样和其他蒙特卡洛方法，包括对算法初始化和收敛的讨论.

比 Gibbs 抽样速度更快的一种蒙特卡洛方法是重要性抽样. 当处理非常大的网络时，只有最复杂的蒙特卡洛方法：成块 Gibbs 抽样和基于预计算的重要性抽样，能提供好的结果.

似然性加权和反向模拟[2]是两种模拟图中某些变量的取值，不考虑图中其他部分存储的信息对这些变量影响的重要抽样方法.

Zhang 和 Poole 提出基于剪枝的重要性抽样，使用重要性抽样方法估计出近似的后验概率，按照它抽取出样本，并且比较了基于剪枝与基于预计算方法的差别.

Hernandez 等人描述的预计算近似算法使用概率表示抽样分布，其主要问题是概率表的大小与变量的可能取值数目成比例.

两种稀疏表示方法 - 规则库和概率树都试图在资源有限的条件下，以更少的空间表示更多的信息. 它们利用条件概率分布中存在的上下文独立关系，简化网络以达简便推理的目的.

Antonio Salmero 等人提出一种基于预计算的二阶段重要性抽样方法：第一阶段基于概率树表示形式，通过节点的顺序删除进行近似计算；第二阶段在第一步抽样的基础上抽取一个样本，不同的样本具有不同的权重. 这个算法直接在概率树中计算，因此比概率表示的基于预计算的重要性抽样在性能上有极大的改进.

3.2.2 高斯近似法 高斯近似法是一种比蒙特卡洛方法计算复杂度低，在处理大型样本也可以得到较准确结果的近似方法. 其基本思想是：对于大规模数据，可以用多元高斯分布近似模拟 $p(\theta_s \mid D, S) \propto p(D \mid \theta_s, S)$，$g(\theta_s) = \log(p(D \mid \theta_s, S)\, p(\theta_s \mid S))$，定义 $\tilde{\theta}_s = \arg\max_{\theta_s}\{g(\theta_s)\}$，则 $\tilde{\theta}_s$ 亦使 $p(\theta_s \mid D, S)$ 最大；以 $g(\theta_s)$ 在 $\tilde{\theta}_s$ 点的二阶泰勒展开式近似计算 $g(\theta_s) \approx g(\tilde{\theta}_s) - 1/2(\theta_s - \tilde{\theta}_s)\,A\,(\theta_s - \tilde{\theta}_s)^T$，其中 $(\theta_s - \tilde{\theta}_s)^T$ 是行向量 $(\theta_s - \tilde{\theta}_s)$ 转置矩阵，A 是 $g(\theta_s)$ 在 $\tilde{\theta}_s$ 点的赫斯行列式. 由此，可以得出 $p(\theta_s \mid D, S) \propto p(D \mid \theta_s, S)\, p(\theta_s \mid S) \approx p(D \mid \tilde{\theta}_s, S)\, P(\tilde{\theta}_s \mid S)$ $\exp\{-\frac{1}{2}(\theta_s - \tilde{\theta}_s)\,A\,(\theta_s - \tilde{\theta}_s)^T\}$.

从上式可以看出，为完成高斯近似计算必须找到 $\tilde{\theta}_s$，并计算 $g(\theta_s)$ 在 $\tilde{\theta}_s$ 点的赫斯行列式 $\exp\{-\frac{1}{2}(\theta_s - \tilde{\theta}_s)\,A\,(\theta_s - \tilde{\theta}_s)^T\}$. Meng 与 Rubin 讨论了计算技巧，Raftery 介绍了如何利用很多统计包计算出似然比率来近似计算赫斯行列式 $\exp\{-\frac{1}{2}(\theta_s - \tilde{\theta}_s)\,A\,(\theta_s - \tilde{\theta}_s)^T\}$，Thiesson 使用 Bayesian 网推理计算无限制的多项分布.

3.2.3 EM 算法 EM 算法的基本思想：当实例数据不完备时，利用 Bayesian 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合. Little 和 Rubin[3]证实当能够建造统计模型时，EM 算法是一种好的学习算法. 算法的具体步骤：(1) 初始化：随机为 θ_s 赋初值；(2) 求期望：计算每个可能事件 $X_i = x_i^k$，$\pi_{X_i} = \pi_{X_i}^j$ 在 θ_s 条件下发生的期望统计因子. 以离散变量为例：$E_{p(x \mid D), \theta_s, S}(N_{ijk}) = \sum_{l=1}^{M} p(x_i^k, \pi_{X_i}^j \mid y_l, \theta_s, S)$，其中 N_{ijk} 表示 $X_i = x_i^k$，$\pi_{X_i} = \pi_{X_i}^j$ 的充分统计因子；(3) 取最大：按照期望充分因子把不完备数据 D 转换成完备数据样本；(4) 如果达到精度要求，则算法结束，否则返回 2.

EM 算法利用 Bayesian 网推理算法来计算无观测值变量的期望值，选择高效的推理算法是 EM 算法的关键. Lauriten 采用关联树算法就减少了 EM 算法的复杂性.

Dempster 等人在证明了 EM 算法存在局部极值问题的情况下，为加快收敛速度，定义了广义 EM 算法 GEM 保证“取部分最大”仍具有收敛的性质. “取部分最大”是增大完备数据的似然性或后验概率，而不是使其最大，Meng 和 Rubin 提出的期望条件最大算法是加速的实例. EM 算法的主要开销在“求期望”步骤，不是在“取最大”步骤，上述的加速算法没有显著效果，不能应用于大型数据库. Neal 和 Hinton 提出了增量 EM 算法和 Bothiesson 等人提出了迟缓 EM 算法，增量 EM 算法采用“求部分期望”的做法，在数据的每一次循环遍历中增大充分统计因子及参数，较快地开拓出新信息，加快了收敛速度. 迟缓 EM 算法在选择的数据子集上执行几次连续的“求期望-取最大”过程，具有收敛性. 实践证明可以应用于大型数据库的学习，并且显著地降低了计算开销.

3.2.4 梯度上升算法 梯度上升算法是一种利用“梯度”来快速计算函数极大值的数学方法，其执行过程就是沿着函数的表面一步步向函数的极大值逼近. 梯度方法是解决非线性优化的方法，适应性强、灵活，可以借鉴神经元网络的训练方法. 算法的具体步骤：(1) 初始化：给参数 θ 随机地赋值，并指定步长 α；(2) 计算似然函数在点 θ 的梯度向量 $\Delta\theta$. $\frac{\partial \ln p(C \mid \theta)}{\partial \theta_{ijk}} = \sum_{l=1}^{m} \frac{p(X_i^k, pa_i^j \mid C_l, \theta)}{\theta_{ijk}}$，该公式由 Russel[4] 证明；(3) 如果该点的梯度向量为 0，则算法结束，θ 为求得的条件概率表；如果梯度向量不为 0，则对参数进行修改，获得新的参数 $\theta = \theta + \alpha \times \Delta\theta$，并进行归一化处理，使参数 θ 符合条件概率表的约束条件，返回 (2).

梯度方法需要在所有合理的参数范围空间中搜索，存在局部极值，需要结合优化技巧以求得理想结果.

John Binder 等人描述了一种局部计算梯度的方法，使用标准 Bayesian 网推理算法可以得出的“副产品”信息以局部方式计算梯度，实验说明这种方法扩展了参数较少的噪音-或

和动态 Bayesian 网,以及含有连续变量的混合模型都达到了加速学习的目的.

4 比较

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 Bayesian 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 Sebastian[5]还把算法 BC 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.

5 今后发展方向

多数情况下,大部分的概率值来源于领域专家,由于耗时太多很难获取,较实用的方法是减少需要评估的概率数目.

Bayesian 网需要的概率数目直接依赖于网络的图形拓扑结构,网络链接越密集,需要的概率参数越多,对于每个变量需要提供的条件概率数一般是其父亲变量的指数幂. 可以通过采取改换网络图形拓扑结构及采用参数化的概率分布方法,减少一个网络中需要评估的概率数. 变换拓扑结构的方法有:引入中间变量分离父亲节点集或者删除表示弱关系的有向弧. Eddie Schwalb 提出了结构转换方法,借用神经网络结构,根据专家给定的模型和部分观察集使用后向传播对条件概率进行估计. 对一个变量使用参数化概率分布的目的是减少直接评估的概率数目,提供一些简单规则计算一些概率值,常用的参数化概率分布有噪音 - 或、噪音 - 和模型等,这些模型的基础是变量和其父亲变量集之间发生因果关系的模式,这些模式中需要直接评估的概率值是父亲变量数目的线性倍数,其余的指数数目的概率值可以利用模型中的基本独立关系推导出来. 改变拓扑结构和采用参数化的概率分布都可能降低准确度,简化后的网络占用领域专家的时间较少,节省的时间可用于网络的证实和精化.

图形工具提供直方图和饼型统计图方式使得领域专家能够更准确地给出概率值. 可以对推理行为做出自动解释的工具也有助于网络的量化过程,对网络推理行为的详细解释能够指出网络定量部分存在的概率问题,以便纠正和改进. 随着 Bayesian 网应用范围的逐渐扩大,会越来越需要一种工程化的原则指导 Bayesian 网的建造. 随着支持全局建造过程的高效重复过程和相关图形工具的出现,Bayesian 网的量化过程会逐渐简化.

6 总结

Bayesian 网以其丰富的概率表达能力、综合先验知识的学习特性及其稳固的数学基础引起了研究人员的关注. 本文探讨了概率学习的方法,并对它们进行了相应的比较. 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合所提供的信息,无法加入专家知识,对实例数据的依赖性大;基于 Bayesian 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 Bayesian 网学习的基础,是 Bayesian 网结构学习必不可少的部分.

参考文献:

[1] Ruseel S, et al. Local learning in probabilistic network with hidden variables[A]. Proc. of The 14th Ijcai Montreal[C]. Canada, Morgan Kaufmann, 1995. 1146 - 1152.

[2] R Fung, B D Favero. Backward simulation in Bayesian networks[A]. R L De Mántaras, D Poole. Proceeding of the 10th Conference on Uncertainty in Artificial Intelligence[C]. (San Mateo) Morgan & Kauffman, 2000. 227 - 234.

[3] Little R, D Rubin. Statistical Analysis with Missing Data[M]. New York, John Wiley & Sons. 1997.

[4] Ruseel S, et al. Local learning in probabilistic network with hidden variables[A]. Proc. of The 14th Ijcai Montreal[C]. Canada, Morgan Kaufmann, 1995. 1146 - 1152.

[5] M Ramoni, P Sebastiani. Learning Bayesian networks from incomplete databases. KMi[R]. KMi-TR-43. 2001.

作者简介:

薛万欣 女,1966 年 5 月生于辽宁,博士生,主要研究方向:不确定性推理,分布式人工智能,电子商务等.

刘大有 男,1942 年 7 月生于河北,教授,博士导师,研究方向:人工智能,不确定性推理,神经元网络与遗传算法等.

张 弘 男,1968 年 1 月生于长春,博士生,研究方向:遗传算法与神经元网络等.