

小数据集条件下基于双重约束的 BN 参数学习

郭志高¹ 高晓光¹ 邸若海¹

摘要 针对小数据集条件下的贝叶斯网络 (Bayesian network, BN) 参数学习问题, 提出了一种基于双重约束的贝叶斯网络参数学习方法. 首先, 对网络中的参数进行分析并将网络中的参数划分为: 父节点组合状态相同而子节点状态不同的参数和父节点组合状态不同而子节点状态相同的参数; 然后, 针对第一类参数提出了一种新的基于 Beta 分布拟合的贝叶斯估计方法, 而针对第二类参数利用已有的保序回归估计方法进行学习, 进而实现了对网络中参数的双重约束学习; 最后, 通过仿真实例说明了基于双重约束的参数学习方法对小数据集条件下贝叶斯网络参数学习精度提高的有效性.

关键词 贝叶斯网络, 参数学习, 小数据集, Beta 分布, 保序回归

引用格式 郭志高, 高晓光, 邸若海. 小数据集条件下基于双重约束的 BN 参数学习. 自动化学报, 2014, 40(7): 1509–1516

DOI 10.3724/SP.J.1004.2014.01509

Learning Bayesian Network Parameters under Dual Constraints from Small Data Set

GUO Zhi-Gao¹ GAO Xiao-Guang¹ DI Ruo-Hai¹

Abstract In this paper, a novel dual constraints based parameter learning algorithm is presented to overcome the problem of Bayesian network (BN) parameter learning from small data sets. First, the parameters in the network are analyzed and classified into classes as follows: parameters referring to different child states sharing the same parent configuration state and parameters referring to different parent configuration states sharing the same child state. Then, a novel beta distribution approximation based Bayesian estimation method is proposed, which is suitable for the learning of the first category parameters. Meanwhile, previously proposed isotonic regression estimation method is employed to compute the second category parameters. Finally, simulations demonstrate the effectiveness of the proposed algorithm on improving the precision of Bayesian network parameter learning from small data set.

Key words Bayesian network (BN), parameter learning, small data set, beta distribution, isotonic regression

Citation Guo Zhi-Gao, Gao Xiao-Guang, Di Ruo-Hai. Learning Bayesian network parameters under dual constraints from small data set. *Acta Automatica Sinica*, 2014, 40(7): 1509–1516

贝叶斯网络 (Bayesian network, BN) 是表示随机变量间依赖和独立关系的网络模型^[1], 由节点集、有向边集和条件概率分布集组成. 其中, 节点表示随机变量, 有向边表示

变量间的依赖关系, 条件概率分布表示变量间的依赖程度. 现实生活中的许多实例如医疗诊断^[2]、故障诊断^[3]、音频识别^[4]、目标识别^[5]、行为识别^[6]、视频跟踪^[7]、智能控制^[8] 等都可以通过 BN 进行建模. 由于具有较强的直观性, BN 已逐渐成为不确定性分析的重要工具并引起了国内外学者的广泛关注. BN 的学习包括结构学习和参数学习. 通常来说, 要构建一个 BN 须咨询相关领域的专家, 以获取该领域的经验知识. 事实上, 对于一个领域专家来说, 他 (或她) 能够更容易并且可靠地确定一个 BN 的网络结构^[9], 而很难给出具体的参数^[10]. 如果能够获得充足的样本数据, 那么利用经典的极大似然估计 (Maximum likelihood estimation, MLE) 方法, 就能够比较容易且准确地估计出 BN 参数. 但不幸的是, 从某些系统中获取大量样本数据非常困难或者代价十分昂贵, 如医疗诊断系统中的病例数据、金融操作风险管理系统中的案例数据、空战态势评估系统中的空战数据、航空发动机故障诊断系统中的发动机故障数据等. 在这种情况下, 由于所能获得的样本数据往往很少, 因而无法精确估计出 BN 中的参数. 而对 BN 而言, 不准确的 BN 参数将导致 BN 推理的结果差强人意, 甚至违背常理. 所以, 一直以来, 提高小数据集条件下 BN 参数学习的精度是一个受到广泛关注的问题.

领域专家虽然很难给出精确的网络参数, 却能够比较容易地给出网络中相关节点之间的关系信息, 即约束信息^[9, 11]. 而事实上, 专家提供的这些约束信息比具体的参数信息拥有更好的鲁棒性. 迄今为止, Wittig 等^[12] 已经证明了利用 BN 中的约束信息不仅可以提高 BN 参数估计特别是小数据集条件下的参数估计的精度, 而且利用约束信息学习得到的 BN 可以得到更加合理的推理结果, 也因此更加容易被人们接受. 目前, 众多学者已经利用约束信息对小数据集条件下 BN 参数学习开展了若干研究并取得了一些成果. 例如, Wittig 等^[12] 利用定性影响约束进行 BN 参数学习, 该方法首先构建定性影响约束模型, 然后利用定性影响约束构造惩罚函数, 并将最大熵函数与惩罚函数相结合构造目标优化函数, 最后利用梯度搜索算法求解参数. Altendorf 等^[13] 利用定性影响约束进行 BN 参数学习, 不同之处在于, 该方法在构建定性影响约束模型并将利用定性影响约束转化得到的惩罚函数与最大熵函数结合构建优化函数后, 利用 APN (Adaptive probabilistic networks) 算法进行求解. Feelders 等^[14] 利用定性影响约束进行 BN 参数学习, 不同的是, 该方法首先利用极大似然估计进行 BN 参数学习得到初始参数; 然后, 根据网络节点之间存在的定性影响构造参数之间的序约束; 最后, 利用保序回归算法校正初始参数使得参数满足序约束. Niculescu 等^[15] 将定性影响约束加到 BN 参数学习中, 该方法首先构建定性影响约束模型, 然后通过拉格朗日乘子将定性影响约束和最大熵函数结合起来, 通过求极值的方法求解 BN 参数. Campos 等^[16] 利用协同约束进行 BN 参数学习, 该方法首先构建协同约束模型; 然后, 将最大熵函数与协同约束相结合构造多线性规划问题; 最后, 利用 MINOS 工具箱进行求解. de Campos 等^[17] 利用非单调性约束进行 BN 参数学习, 该方法首先构建非单调性约束模型; 然后, 将最大熵函数与非单调性约束结合构造凸优化问题; 最后, 利用 Mosek 工具箱进行求解. de Campos 等^[18] 利用定性影响约束进行 BN 参数学习, 该方法首先构建定性影响约束模型; 然后, 将定性影响约束与最大熵函数结合构造凸优化问题; 最后, 利用凸优化工具箱进行求解. Isozaki 等^[19] 利用规范性约束进行 BN 参数学习, 该方法首先将包含 KL (Kullback-Leibler) 距离和熵函数的最小自由能函数作为目标优化函数; 然后,

收稿日期 2013-08-30 录用日期 2013-12-03
Manuscript received August 30, 2013; accepted December 3, 2013
国家自然科学基金 (60774064), 教育部博士点基金 (20116102110026) 资助
Supported by National Natural Science Foundation of China (60774064), Research Fund for the Doctoral Program of Higher Education of China (20116102110026)
本文责任编辑 王聪
Recommended by Associate Editor WANG Cong
1. 西北工业大学电子信息学院 西安 710129
1. School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710129

通过拉格朗日乘子将最小自由能函数与规范性约束结合起来, 通过求极值的方法求取参数. Liao 等^[20] 利用定性影响约束进行 BN 参数学习, 该方法首先构建定性影响约束模型; 然后, 引入权重将约束与最大熵函数结合起来; 最后, 利用改进 EM 算法进行参数求解. Chang 等^[21] 利用竞争性约束进行 BN 参数学习, 该方法首先利用拒绝-接受采样法抽取满足约束的参数; 然后, 用随机抽样数乘以参数以获取虚拟统计数据; 最后, 利用贝叶斯估计方法将真实统计数据和虚拟统计数据相结合学习得到 BN 参数.

仔细分析可以发现: 已有文献都是基于单一约束的 BN 参数学习, 而单一约束对参数的约束力有限, 进而不能最大程度地提高参数学习的精度. 针对上述问题, 本文提出将网络中的参数划分为: 父节点组合状态相同而子节点状态不同的参数和父节点组合状态不同而子节点状态相同的参数. 然后, 对上述两类参数分别进行约束学习. 具体来说: 针对第一类参数利用一种新的基于 Beta 分布拟合的贝叶斯估计方法进行约束学习, 针对第二类参数利用一种已有的保序回归估计方法^[14] 进行约束学习, 实现了小数据集条件下 BN 参数的双重约束学习. 实验结果表明: BN 参数的双重约束学习使学习得到的 BN 参数, 特别是小数据集条件下学习得到的 BN 参数精度有了较大的提高.

本文的组织结构如下: 第 1 节, 简要介绍了 BN 的相关理论; 第 2 节, 详细介绍了基于双重约束的 BN 参数学习方法; 第 3 节, 仿真验证了本文方法对 BN 参数学习精度提高的有效性; 第 4 节, 简要总结了本文的工作.

1 相关理论

1.1 贝叶斯网络

一个 BN 可以定义为一个二元组 $B = (G, \theta)$, 其中 $G = (X, A)$ 为一个有向无环图, $\theta = (\theta_1, \theta_2, \dots, \theta_N)$ 为网络的参数集合, $X = \{X_1, X_2, \dots, X_N\}$ 表示节点变量集合, X_i 表示第 i 个节点变量, $A = [a_{ij}]_{N \times N}$ 为节点之间的有向弧集合, a_{ij} 表示变量 X_i 和 X_j 之间的依赖关系 $X_j \rightarrow X_i$, $\theta_i = p(x_i | x_{\pi(i)})$ 表示节点 X_i 的父节点 $X_{\pi(i)}$ 的组合状态为 $x_{\pi(i)}$ 的情况下, 节点 X_i 的状态为 x_i 的条件概率.

一个 BN 用网络结构 G 和网络参数 θ 确定了网络中节点 X 的联合概率分布 $\Pr(X)$. 具体来说, $\Pr(X)$ 可以表示为

$$\Pr(X) = \prod_{i=1}^N \theta_i = \prod_{i=1}^N p(x_i | x_{\pi(i)}) \quad (1)$$

1.2 贝叶斯网络参数学习

BN 参数学习是指在 BN 结构已知的情况下, 通过统计方法从样本数据集中估计网络参数的过程. 设 $n(x_{\pi(i)})$ 表示样本数据集中节点 X_i 的父节点 $X_{\pi(i)}$ 的组合状态为 $x_{\pi(i)}$ 的样本数目; $n(x_i, x_{\pi(i)})$ 表示样本数据集中节点 X_i 状态为 x_i 同时父节点 $X_{\pi(i)}$ 组合状态为 $x_{\pi(i)}$ 的样本数目. 根据马尔科夫独立性, BN 参数学习问题可以分解为不同节点与其父节点之间的参数学习问题. 参数 $p(x_i | x_{\pi(i)})$ 的最大似然估计方法公式为

$$p(x_i | x_{\pi(i)}) = \frac{n(x_i, x_{\pi(i)})}{n(x_{\pi(i)})} \quad (2)$$

这种估计由于使得在给定数据集的情况下 BN 参数 p 的对数似然函数值 $\ell(p|D)$ 最大, 而被称作最大似然估计方法.

$$\ell(p|D) = \sum_{i=1}^N \sum_{x_i, x_{\pi(i)}} n(x_i, x_{\pi(i)}) \log p(x_i | x_{\pi(i)}) \quad (3)$$

2 小数据集条件下基于双重约束的 BN 参数学习方法

针对一个由 N 个变量 $X = \{X_1, X_2, \dots, X_N\}$ 组成的贝叶斯网络, 设:

1) 节点 X_i 共有 r_i 个状态值且状态值集合为 $\chi_i = \{x_i^1, x_i^2, \dots, x_i^{r_i}\}$;

2) 节点 X_i 共有 p_i 个父节点且父节点集为 $X_{\pi(i)} = \{X_{\pi(i)}^1, X_{\pi(i)}^2, \dots, X_{\pi(i)}^{p_i}\}$;

3) 节点 X_i 的第 j 父节点 $X_{\pi(i)}^j$ 有 $d_{\pi(i)}^j$ 个状态值且状态值集合为 $\chi_{\pi(i)}^j = \{x_{\pi(i),j}^1, x_{\pi(i),j}^2, \dots, x_{\pi(i),j}^{d_{\pi(i)}^j}\}$;

4) 节点 X_i 的父节点有 $M = d_{\pi(i)}^1 \times d_{\pi(i)}^2 \times \dots \times d_{\pi(i)}^{p_i}$ 种组合状态值且组合状态值集合 $\chi_{\pi(i)} = \times_{j=1}^{p_i} \chi_{\pi(i)}^j = \{X_{\pi(i)}^1, X_{\pi(i)}^2, \dots, X_{\pi(i)}^M\}$.

2.1 网络中参数的划分

在进行 BN 参数学习之前, 首先对网络中的参数进行分析和分类. 经过分析发现 BN 中的参数可以分成两类: 1) 父节点组合状态相同而子节点状态不同所对应的参数, 这类参数满足参数之和为 1 的特性; 2) 父节点组合状态不同而子节点状态相同的参数, 这类参数并不满足参数之和为 1 的特性.

以图 1 为例, 图中 $\theta_{i1l}, \theta_{i2l}, \theta_{i3l}, \dots, \theta_{ir_i l}$ 为父节点组合状态相同子节点状态不同时的参数 (即第一类参数), 它们的具体大小分别为

$$\begin{aligned} \theta_{i1l} &= p(x_i = x_i^1 | x_{\pi(i)} = X_{\pi(i)}^l) \\ \theta_{i2l} &= p(x_i = x_i^2 | x_{\pi(i)} = X_{\pi(i)}^l) \\ \theta_{i3l} &= p(x_i = x_i^3 | x_{\pi(i)} = X_{\pi(i)}^l) \\ &\vdots \\ \theta_{ir_i l} &= p(x_i = x_i^{r_i} | x_{\pi(i)} = X_{\pi(i)}^l) \end{aligned}$$

其中, $X_{\pi(i)}^l$ 为第 l 组父节点组合状态值.

图中 $\theta_{ik1}, \theta_{ik2}, \theta_{ik3}, \dots, \theta_{ikp_i}$ 为父节点组合状态不同子节点状态相同时的参数 (即第二类参数), 它们的具体大小分别为

$$\begin{aligned} \theta_{ik1} &= p(x_i = x_i^k | x_{\pi(i)} = X_{\pi(i)}^1) \\ \theta_{ik2} &= p(x_i = x_i^k | x_{\pi(i)} = X_{\pi(i)}^2) \\ \theta_{ik3} &= p(x_i = x_i^k | x_{\pi(i)} = X_{\pi(i)}^3) \\ &\vdots \\ \theta_{ikp_i} &= p(x_i = x_i^k | x_{\pi(i)} = X_{\pi(i)}^{p_i}) \end{aligned}$$

其中, x_i^k 为节点 X_i 的第 k 个状态值.

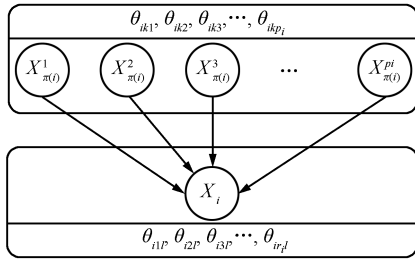


图 1 BN 参数划分示意图

Fig. 1 Parameters classification in Bayesian network

注 1. 上述参数的划分方式是基于普适性的参数划分, 适用于所有贝叶斯网络. 在实际应用中, 领域专家并不需要参与网络参数的划分, 而是由计算机按照上述划分方式自动完成.

注 2. 除上述两类参数以外, 还存在另外两种情况: 第一种情况, 父节点组合状态相同且子节点状态相同的参数; 第二种情况, 父节点组合状态不同且子节点状态不同的参数. 其中: 1) 第一种情况的参数为同一个参数, 而同一个参数与自身不具备可比性, 因而无法产生约束; 2) 第二种情况的参数之间不存在任何联系, 同样不具备可比性, 因而无法产生约束.

基于上述分类, 本文提出一种基于双重约束的参数估计 (Dual constrained estimation, DCE) 方法. 该方法的主要思路为: 首先, 利用本文提出的基于 Beta 分布拟合的贝叶斯估计方法对网络中父节点组合状态相同而子节点状态不同的参数进行第一重约束学习; 然后, 利用保序回归估计方法^[22]对网络中父节点组合状态不同而子节点状态相同的参数进行第二重约束学习, 进而完成对网络参数的双重交叉约束学习. 具体过程如下:

2.2 第一重约束学习: 网络中父节点组合状态相同, 子节点状态不同的参数学习

定义 1. 如果网络中父节点组合状态相同而子节点状态不同的参数 $\theta_{i1l}, \theta_{i2l}, \theta_{i3l}, \dots, \theta_{ir_i l}$ 满足:

$$\theta_{i1l} < \theta_{i2l} < \dots < \theta_{ikl} < \dots < \theta_{ir_i l}$$

则称 θ_{ikl} 是关于 k 单调递增的; 反之, 则称 θ_{ikl} 是关于 k 单调递减的.

当参数 θ_{ikl} 关于 k 单调递增时, 由参数的规范性和单调性可得:

$$\begin{cases} \theta_{i1l} + \theta_{i2l} + \dots + \theta_{ir_i l} = 1 \\ \theta_{i1l} < \theta_{i2l} < \dots < \theta_{ir_i l} \end{cases} \Rightarrow \begin{cases} 0 < \theta_{i1l} < \frac{1}{r_i} \\ \theta_{i1l} < \theta_{i2l} < \frac{1 - \theta_{i1l}}{r_i - 1} \\ \vdots \\ \theta_{ir_i l} = 1 - \sum_{k=1}^{r_i-1} \theta_{ikl} \end{cases} \quad (4)$$

相反, 当参数 θ_{ikl} 关于 k 单调递减时, 由参数的规范性

和单调性可得:

$$\begin{cases} \theta_{i1l} + \theta_{i2l} + \dots + \theta_{ir_i l} = 1 \\ \theta_{i1l} > \theta_{i2l} > \dots > \theta_{ir_i l} \end{cases} \Rightarrow \begin{cases} 0 < \theta_{ir_i l} < \frac{1}{r_i} \\ \vdots \\ \theta_{i3l} < \theta_{i2l} < \frac{1 - \sum_{k=3}^{r_i} \theta_{ikl}}{2} \\ \theta_{i1l} = 1 - \sum_{k=2}^{r_i} \theta_{ikl} \end{cases} \quad (5)$$

这样就得到了每个参数的取值范围.

1) 当参数 θ_{ikl} 关于 k 单调递增时, 参数 θ_{ikl} 的约束学习过程如下:

a) 当 $k = 1$ 时, 在无任何其他先验的情况下, 我们可以认为 θ_{i1l} 的取值服从均匀分布. 而由式 (4) 可知 θ_{i1l} 服从均匀分布 $U(0, 1/r_i)$, 则由均匀分布期望和方差的计算方法可以得到:

$$\begin{cases} E_{\theta_{i1l}} = \frac{1}{2r_i} \\ D_{\theta_{i1l}} = \frac{1}{12r_i^2} \end{cases} \quad (6)$$

由于 Beta 分布便于处理, 所以我们用 Beta 分布 $B(a_1, b_1)$ 逼近均匀分布. 为此, 我们将均匀分布与 Beta 分布两者方差之间的距离最小作为目标函数, 期望值相等作为约束, 得到 Beta 分布参数的获取方法:

$$\begin{aligned} & \min (D_{B1} - D_{\theta_{i1l}})^2 \\ & \text{s.t.} \begin{cases} E_{B1} = E_{\theta_{i1l}} \\ a_1 > 0, b_1 > 0 \end{cases} \end{aligned} \quad (7)$$

其中, a_1, b_1 为 Beta 分布的两个参数, E_{B1} 和 D_{B1} 为 Beta 分布的期望和方差, 它们的大小分别为

$$\begin{aligned} E_{B1} &= \frac{a_1}{a_1 + b_1} \\ D_{B1} &= \frac{a_1 b_1}{(a_1 + b_1)^2 (a_1 + b_1 + 1)} \end{aligned} \quad (8)$$

将式 (6) 和 (8) 带入式 (7) 中可以得到如下目标优化问题:

$$\begin{aligned} & \min \left(\frac{a_1 b_1}{(a_1 + b_1)^2 (a_1 + b_1 + 1)} - \frac{1}{12r_i^2} \right)^2 \\ & \text{s.t.} \begin{cases} \frac{a_1}{a_1 + b_1} = \frac{1}{2r_i} \\ a_1 > 0, b_1 > 0 \end{cases} \end{aligned} \quad (9)$$

求解上式可以得到 Beta 分布的参数 a_1 和 b_1 , 然后, 将 a_1 和 b_1 作为先验虚拟样本数据加入到后验样本数据中由贝叶斯估计公式:

$$\theta_{i1l} = \frac{N_{i1l} + a_1}{N_{il} + a_1 + b_1} \quad (10)$$

计算得到约束条件下 θ_{i1l} 的参数值. 其中, N_{il} 代表样本数据中节点 X_i 的父节点组合状态为 $X_{\pi(i)}^l$ 的样本数量, N_{i1l} 代表样本数据中节点 X_i 的状态为 x_i^1 且其父节点组合状态为 $X_{\pi(i)}^l$ 的样本数量.

b) 当 $1 < k < r_i$ 时, 在无任何先验的情况下, 我们可以认为 θ_{ikl} 的取值服从均匀分布. 而由式 (4) 可知, θ_{ikl} 服从均匀分布 $U(\theta_{i(k-1)l}, (1 - \sum_{m=1}^{k-1} \theta_{iml})/r_i - k + 1)$, 则由均匀分布期望和方差的计算方法可以得到:

$$\begin{cases} E_{\theta_{ikl}} = \frac{\theta_{i(k-1)l}}{2} + \frac{1 - \sum_{m=1}^{k-1} \theta_{iml}}{2(r_i - k + 1)} \\ D_{\theta_{ikl}} = \frac{[\theta_{i(k-1)l} - \frac{1 - \sum_{m=1}^{k-1} \theta_{iml}}{r_i - k + 1}]^2}{12} \end{cases} \quad (11)$$

同样, 为了用 Beta 分布 $B(a_k, b_k)$ 逼近均匀分布, 我们将均匀分布与 Beta 分布两者方差之间的距离最小作为目标函数, 期望值相等作为约束, 得到 Beta 分布参数的获取方法:

$$\begin{cases} \min (D_{Bk} - D_{\theta_{ikl}})^2 \\ \text{s.t.} \begin{cases} E_{Bk} = E_{\theta_{ikl}} \\ a_k > 0, b_k > 0 \end{cases} \end{cases} \quad (12)$$

其中, a_k, b_k 为 Beta 分布的两个参数, E_{Bk} 和 D_{Bk} 为 Beta 分布的期望和方差, 它们的大小分别为

$$\begin{aligned} E_{Bk} &= \frac{a_k}{a_k + b_k} \\ D_{Bk} &= \frac{a_k b_k}{(a_k + b_k)^2 (a_k + b_k + 1)} \end{aligned} \quad (13)$$

将式 (11) 和 (13) 带入式 (12) 中可以得到如下目标优化问题:

$$\begin{cases} \min \left\{ \frac{a_k b_k}{(a_k + b_k)^2 (a_k + b_k + 1)} - \frac{[(r_i - k + 1)\theta_{i(k-1)l} - 1 + \sum_{m=1}^{k-1} \theta_{iml}]^2}{12(r_i - k + 1)^2} \right\}^2 \\ \text{s.t.} \begin{cases} \frac{a_k}{a_k + b_k} = \frac{\theta_{i(k-1)l}}{2} + \frac{1 - \sum_{m=1}^{k-1} \theta_{iml}}{2(r_i - k + 1)} \\ a_k > 0, b_k > 0 \end{cases} \end{cases} \quad (14)$$

求解式 (14) 得到 Beta 分布的参数 a_k 和 b_k , 然后将 a_k 和 b_k 作为先验虚拟样本数据加入到后验样本数据中由贝叶斯估计公式:

$$\theta_{ikl} = \frac{N_{ikl} + a_k}{N_{il} + a_k + b_k} \quad (15)$$

计算得到约束条件下 θ_{ikl} 的参数值. 其中, N_{il} 代表样本数据中节点 X_i 的父节点组合状态为 $X_{\pi(i)}^l$ 的样本数量, N_{ikl} 代表样本数据中节点 X_i 的状态为 x_i^k 且其父节点组合状态为 $X_{\pi(i)}^l$ 的样本数量.

c) 当 $k = r_i$ 时, 由式 (4) 可直接计算得到 $\theta_{ir_i l}$:

$$\theta_{ir_i l} = 1 - \sum_{m=1}^{r_i-1} \theta_{iml} \quad (16)$$

2) 当参数 θ_{ikl} 关于 k 单调递减时, 参数 θ_{ikl} 的约束学习过程与当上述过程相似, 不同之处在于该过程: 首先, 对 $\theta_{ir_i l}$ 进行拟合估计; 然后, 对参数 θ_{ikl} ($1 < k < r_i$) 进行拟合估计; 最后, 利用参数规范性计算得到 θ_{i1l} .

当参数 θ_{ikl} 关于 k 单调递减时, 参数 θ_{ikl} 的约束学习过程与当上述过程相似, 不同之处在于该过程: 首先, 对 $\theta_{ir_i l}$ 进行拟合估计; 然后, 对参数 θ_{ikl} ($1 < k < r_i$) 进行拟合估计; 最后, 利用参数规范性计算得到 θ_{i1l} .

至此, 我们便完成了对父节点组合状态相同而子节点状态不同的参数的约束学习过程. 为了在第 3 节中比较分析, 我们将第一重约束学习方法命名为基于 Beta 分布拟合的贝叶斯估计 (Beta distribution approximation based Bayesian estimation, BABE) 方法. BABE 算法步骤总结如下:

步骤 1. 首先, 根据领域专家知识确定 θ_{ikl} 关于 k 的单调性, 如果 θ_{ikl} 关于 k 单调递增, 则转入步骤 2; 如果 θ_{ikl} 关于 k 单调递减, 计算过程类似于步骤 2~4.

步骤 2. 首先依据式 (4) 确定 θ_{i1l} 的均匀分布区间; 然后, 优化目标函数 (9) 得到 Beta 分布的参数 a_1 和 b_1 , 最后根据式 (10) 计算得到参数 θ_{i1l} 并转入步骤 3.

步骤 3. 首先, 依据式 (4) 确定 θ_{ikl} ($1 < k < r_i$) 的均匀分布区间; 然后, 优化目标函数 (14) 得到 Beta 分布的参数 a_k 和 b_k ; 最后, 根据式 (15) 计算得到参数 θ_{ikl} 并转入步骤 4.

步骤 4. 根据式 (16) 计算得到参数 $\theta_{ir_i l}$, 算法结束.

2.3 第二重约束学习: 网络中父节点组合状态不同, 子节点状态相同的参数学习

定义 2. 如果网络中父节点组合状态不同而子节点状态相同的参数 $\theta_{ikl_1}, \theta_{ikl_2}, \theta_{ikl_3}, \dots, \theta_{ikl_M}$ 满足:

$$\theta_{ikl_1} < \theta_{ikl_2} < \theta_{ikl_3} \dots < \theta_{ikl_M}$$

则称 $\theta_{ikl_1}, \theta_{ikl_2}, \theta_{ikl_3}, \dots, \theta_{ikl_M}$ 所对应的父节点组合状态 $X_{\pi(i)}^{l_1}, X_{\pi(i)}^{l_2}, X_{\pi(i)}^{l_3}, \dots, X_{\pi(i)}^{l_M}$ 满足如下优先关系: $X_{\pi(i)}^{l_1} \preceq X_{\pi(i)}^{l_2} \preceq X_{\pi(i)}^{l_3} \dots \preceq X_{\pi(i)}^{l_M}$.

当网络中父节点组合状态不同而子节点状态相同的参数所满足的大小关系已知时, 参数的约束学习过程如下:

首先, 根据参数所满足的大小关系由定义 2 可以得到不同父节点组合状态所满足的优先关系; 然后, 根据优先关系可以构造出包含所有最小下限集的集合 $\ell = \{L_1, L_2, L_3, \dots, L_S\}$ [22]. 值得注意的是, 每个最小下限集的元素为不同的父节点组合状态值.

然后, 根据式 (17) 计算每个最小下限集的加权平均值:

$$Av(L_i) = \begin{cases} \frac{\sum_{l \in L_i} N_{il} \theta_{ikl}}{\sum_{l \in L_i} N_{il}}, & \text{若 } \sum_{l \in L_i} w(z) > 0 \\ 0.5, & \text{若 } \sum_{l \in L_i} w(z) = 0 \end{cases} \quad (17)$$

选取加权平均值最小的下限集为 B , 即

$$B = \cup \{A \in \ell | Av(A) = \min_{L_i \in \ell} Av(L_i)\} \quad (18)$$

并令 B 中所有父节点组合状态值所对应的参数值为 $Av(B)$, 即: $\theta_{ikl} = Av(B)$, $X_{\pi(i)}^l \in B$.

删除每个最小下限集 L_i 中所包含的与集合 B 重复的父节点组合状态值, 同时将集合 B 中所包含的父节点组合状态值从 $\chi_{\pi(i)}$ 中删除, 即

$$L_i = L_i \setminus B \quad (19)$$

$$\chi_{\pi(i)} = \chi_{\pi(i)} \setminus B \quad (20)$$

最后, 判断集合 $\chi_{\pi(i)}$ 是否为空集.

以上便为网络中父节点组合状态不同而子节点状态相同的参数约束学习过程. 同样为了便于第 3 部分仿真分析比较, 我们将第二重约束学习方法命名为保序回归估计 (Isotonic regression estimation, IRE) 方法. IRE 方法步骤总结如下:

步骤 1. 首先根据领域专家知识确定父节点状态不同而子节点状态相同的参数所满足的大小关系, 然后根据大小关系得到不同父节点组合状态之间的优先关系并转入步骤 2.

步骤 2. 根据优先关系构造出所有最小下限集并根据式 (17) 计算每个最小下限集的加权平均值并转入步骤 3.

步骤 3. 依据式 (18) 得到集合 B , 并令集合 B 中父节点组合状态值所对应的参数值为 $Av(B)$, 并转入步骤 4.

步骤 4. 对每个最小下限集和所有父节点组合状态值集合 $\chi_{\pi(i)}$ 依据式 (19) 和 (20) 进行相应处理并转入步骤 5.

步骤 5. 判断集合 $\chi_{\pi(i)}$ 是否为空集, 如果是, 结束计算; 否则, 返回步骤 2.

3 仿真实验

仿真实验平台为 Windows 7 系统, 处理器为 Intel CPU 1.8 GHz, 仿真软件为 Matlab 7.1.

3.1 仿真模型

仿真模型为如图 2 所示的草坪湿润推理贝叶斯网络^[23], 该网络模型是一种典型的贝叶斯网络, 被广泛用于评估各种参数学习算法^[23-26], 对 MLE 算法、IRE 算法、BABE 算法、DCE 算法应该具有良好的无偏性. 该网络模型的真实参数如表 1 和 2 所示.

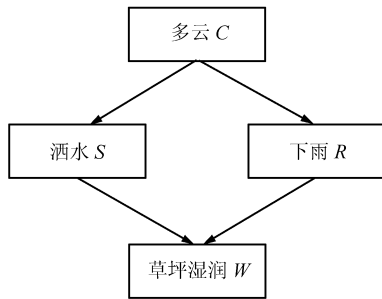


图 2 草坪湿润推理贝叶斯网络

Fig.2 Lawn wet inference Bayesian network

表 1 节点 C 的初始概率值

Table 1 Initial probability values of node C

	$C = 0$	$C = 1$
初始概率	0.5	0.5

表 2 网络真实参数

Table 2 True parameters of the network

	$S = 0$	$S = 1$	$R = 0$	$R = 1$	$S = 0, R = 0$	$S = 0, R = 1$	$S = 1, R = 0$	$S = 1, R = 1$	
$C = 0$	0.1	0.9	0.99	0.01	$W = 0$	0.99	0.2	0.1	0.01
$C = 1$	0.8	0.2	0.2	0.8	$W = 1$	0.01	0.8	0.9	0.99

表 3 实验约束

Table 3 Experimental constraints

编号	约束	编号	约束	编号	约束
1	$P_{S C}(1 0) > P_{S C}(0 0)$	6	$P_{S C}(0 1) > P_{S C}(1 1)$	11	$P_{R C}(1 1) > P_{R C}(0 1)$
2	$P_{R C}(0 0) > P_{R C}(1 0)$	7	$P_{S C}(1 0) > P_{S C}(1 1)$	12	$P_{R C}(1 1) > P_{R C}(1 0)$
3	$P_{W S,R}(1 1,1) > P_{W S,R}(1 1,0)$	8	$P_{W S,R}(1 1,1) > P_{W S,R}(1 0,0)$	13	$P_{W S,R}(1 0,1) > P_{W S,R}(0 0,1)$
4	$P_{W S,R}(1 1,1) > P_{W S,R}(0 1,1)$	9	$P_{W S,R}(1 0,1) > P_{W S,R}(1 0,0)$	14	$P_{W S,R}(1 1,0) > P_{W S,R}(1 0,0)$
5	$P_{W S,R}(1 1,1) > P_{W S,R}(1 0,1)$	10	$P_{W S,R}(0 0,0) > P_{W S,R}(1 0,0)$	15	$P_{W S,R}(1 1,0) > P_{W S,R}(0 1,0)$

3.2 仿真内容

首先, 根据经验知识可以得到关于草坪湿润推理贝叶斯网络的约束信息, 如表 3 所示. 分析可以发现, 第 1、2、4、6、10、11、13 和 15 条约束的约束对象即为父节点组合状态相同而子节点状态不同的参数, 第 3、5、7、8、9、12 和 14 条约束的约束对象即为父节点组合状态不同而子节点状态相同的参数.

注 3. 第 9 约束信息的对立面: 洒水却不下雨的情况下, 草坪干燥的概率小于不洒水不下雨的情况下草坪干燥的概率, 可由第 9 条约束信息和规范性约束保证, 第 3、5、8 和 14 条约束信息类似.

然后, 根据文献 [27] 可以计算出学习得到草坪湿润推理贝叶斯网络精确参数所需的最小样本数据量:

$$N^2 \log N^2 \log N^{k+1} = 4^2 \log 4^2 \log 4^{(2+1)} \approx 35$$

其中, N 为网络节点的数量, k 为网络中最大父节点数. 因此, 数据量小于 35 的数据集称为小数据集.

最后, 从网络中随机抽样获得一个数据量为 10 的小数据集进行参数学习, 抽样数据如表 4 所示.

表 4 实验数据

Table 4 Experimental data

	1	2	3	4	5	6	7	8	9	10
C	0	0	1	0	1	0	0	1	0	0
S	1	1	0	1	0	1	1	0	1	1
R	0	0	1	0	1	0	0	1	0	0
W	1	1	1	1	1	1	1	1	1	1

3.3 仿真结果

在上述约束和数据条件下, 分别利用 MLE、BABE、IRE 和 DCE 算法进行参数学习, 具体学习结果如表 5~表 8 所示.

分析上述学习结果发现: 1) MLE 算法的学习结果违背了第 3、4、5、8、10 条约束; 2) BABE 算法的学习结果违背了第 3、5 条约束; 3) IRE 算法的学习结果违背了第 3 条约束; 4) DCE 算法的学习结果满足了所有约束.

3.4 算法分析

3.4.1 算法优越性分析

1) 算法优越性分析之一

为了分析本文算法的优越性, 我们在不同数据量条件下进行不同算法参数学习实验. 实验结果采用 KL 散度^[28] 进行度量, 实验结果如图 3 所示, 其中图 3(a) 为实验结果的全

表 5 MLE 算法学习得到的参数

Table 5 Parameters learned by MLE algorithm

	$S = 0$	$S = 1$	$R = 0$	$R = 1$	$W = 0$	$S = 0, R = 0$	$S = 0, R = 1$	$S = 1, R = 0$	$S = 1, R = 1$
$C = 0$	0	1	1	0	0	0	0	0	0
$C = 1$	1	0	0	1	1	0	1	1	0

表 6 BABE 算法学习得到的参数

Table 6 Parameters learned by BABE algorithm

	$S = 0$	$S = 1$	$R = 0$	$R = 1$	$W = 0$	$S = 0, R = 0$	$S = 0, R = 1$	$S = 1, R = 0$	$S = 1, R = 1$
$C = 0$	0.1465	0.8535	0.8641	0.1359	0	0.7410	0.1443	0.0941	0.2418
$C = 1$	0.8240	0.1760	0.1509	0.8491	1	0.2590	0.8557	0.9059	0.7582

表 7 IRE 算法学习得到的参数

Table 7 Parameters learned by IRE algorithm

	$S = 0$	$S = 1$	$R = 0$	$R = 1$	$W = 0$	$S = 0, R = 0$	$S = 0, R = 1$	$S = 1, R = 0$	$S = 1, R = 1$
$C = 0$	0	1	1	0	0	1	0.025	0	0
$C = 1$	1	0	0	1	1	0	0.975	1	1

表 8 DCE 算法学习得到的参数

Table 8 Parameters learned by DCE algorithm

	$S = 0$	$S = 1$	$R = 0$	$R = 1$	$W = 0$	$S = 0, R = 0$	$S = 0, R = 1$	$S = 1, R = 0$	$S = 1, R = 1$
$C = 0$	0.1465	0.8535	0.8641	0.1359	0	0.7410	0.1443	0.1375	0.0875
$C = 1$	0.8240	0.1760	0.1509	0.8491	1	0.2590	0.8557	0.8625	0.9125

局显示, 图 3 (b) 为实验结果的局部显示.

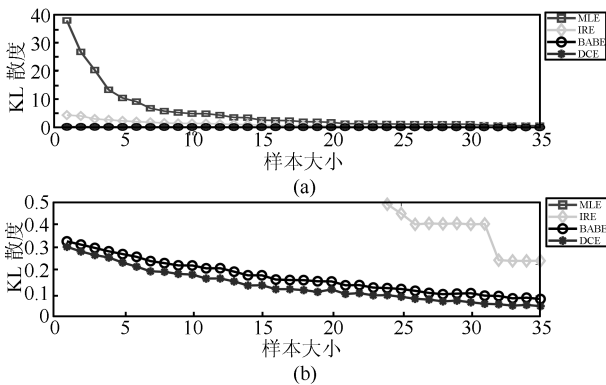


图 3 不同算法优越性分析之一

Fig. 3 Superiority analyses of different algorithms-I

分析可以发现: 1) BABE 和 DCE 算法学习结果的 KL 散度值小于 MLE 和 IRE 算法^[14], 体现了 BABE 和 DCE 算法的优越性; 2) DCE 算法学习结果的 KL 散度值小于 BABE 算法, 进一步体现了 DCE 算法的优越性.

2) 算法优越性分析之二

为了进一步分析本文算法的优越性, 我们进行不同算法学习得到不同 KL 散度值所需数据量实验. 实验结果如图 4 所示.

分析可以发现: 1) MLE 和 IRE 算法学习得到较小 KL 散度值 (如 0.3、0.4、0.5) 都需要较大的数据量 (≈ 30), 而 BABE 和 DCE 算法则需较小的数据量 (< 10), 体现了 BABE 和 DCE 算法的优越性; 2) DCE 算法学习得到较小 KL 散度值 (如 0.1、0.2、0.3) 所需数据量少于 BABE 算法, 进一步体现了 DCE 算法的优越性.

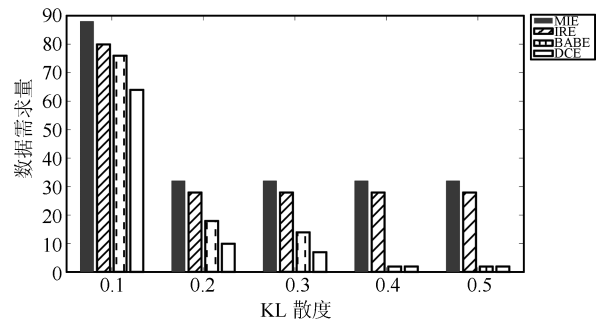


图 4 不同算法优越性分析之二

Fig. 4 Superiority analyses of different algorithms-II

3.4.2 算法实时性分析

为了分析本文算法的实时性, 我们在不同数据量条件下进行不同算法参数学习耗时分析实验. 实验结果如图 5 所示.

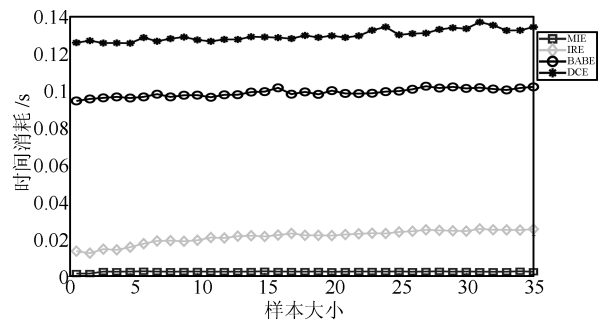


图 5 不同算法实时性分析

Fig. 5 Timeliness analyses of different algorithms

分析可以发现: MLE 算法耗时最少, BABE 和 DCE 算法耗时较长, DCE 算法耗时最长.

3.4.3 算法时效性分析

本文算法在提高参数学习精度的同时增加了学习时间. 通常来说, 学习精度与学习时间是一对矛盾体: 学习精度的提高往往伴随着学习时间的增加. 因此, 为了分析本文算法学习精度与学习时间之间的关系即时效性, 我们进行不同数据量条件下本文算法与其他算法的时效性对比实验. 为此, 我们定义了时效性指标, 该指标计算方法如下:

$$E_1 = \frac{KL_{MLE} \cdot T_{MLE}}{KL_{DCE} \cdot T_{DCE}} \quad (21)$$

$$E_2 = \frac{KL_{IRE} \cdot T_{IRE}}{KL_{DCE} \cdot T_{DCE}} \quad (22)$$

$$E_3 = \frac{KL_{BABE} \cdot T_{BABE}}{KL_{DCE} \cdot T_{DCE}} \quad (23)$$

其中, KL_{MLE} 、 KL_{IRE} 、 KL_{BABE} 、 KL_{DCE} 分别为 MLE、IRE、BABE、DCE 算法学习结果的 KL 散度值, T_{MLE} 、 T_{IRE} 、 T_{BABE} 、 T_{DCE} 分别为 MLE、IRE、BABE、DCE 算法的学习时间. 实验结果如图 6 所示.

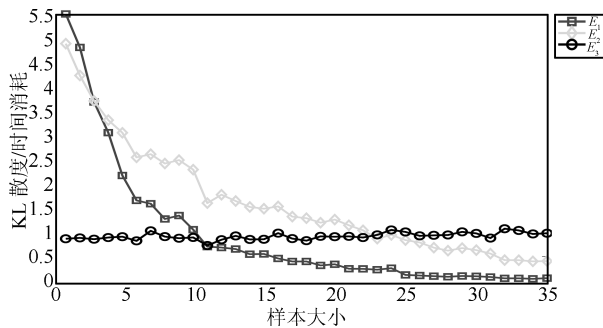


图 6 不同算法时效性分析

Fig. 6 Timeliness-effectiveness analyses of different algorithms

分析可以发现: 1) DCE 算法时效性与 BABE 算法时效性相近; 2) DCE 算法时效性在小数据集条件下特别是数据小于 15 的情况下明显高于 MLE 算法; 3) DCE 算法时效性在小数据集条件下特别是数据小于 25 的情况下也明显高于 IRE 算法. 上述结果表明 BABE 算法和 DCE 算法在小数据集条件下具有较好的时效性.

3.4.4 算法实用性分析

为了分析本文算法的实用性, 我们进行基于不同算法学习参数的推理实验, 具体来说: 首先, 从真实网络中分别抽取训练数据和测试数据; 然后, 利用训练数据和约束信息基于不同算法进行参数学习; 最后, 基于不同算法学习得到的网络和测试数据利用联接树推理算法进行推理实验. 实验结果如图 7 所示.

分析可以发现: 1) 基于 BABE 和 DCE 算法学习参数的推理精度高于 IRE 算法和 MLE 算法, 体现了 BABE 和 DCE 算法良好的实用性; 2) 基于 DCE 算法学习参数的推理精度高于 BABE 算法, 进一步体现了 DCE 算法的实用性.

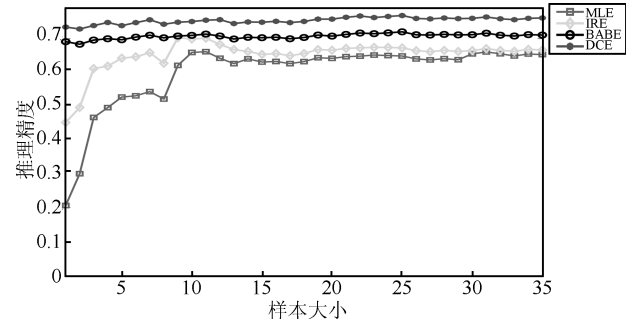


图 7 不同算法实用性分析

Fig. 7 Practicability analyses of different algorithms

4 结论

针对小数据集条件下 BN 参数学习问题, 本文提出了一种基于双重约束的 BN 参数学习方法, 实现了对网络中参数的双重交叉约束学习, 提高了参数学习的精度. 本文贡献可以概括如下:

1) 提出将 BN 网络参数的学习分为对父节点组合状态相同而子节点状态不同参数的学习和父节点组合状态不同而子节点状态相同参数的学习. 上述分类学习方法: 一方面, 使相关研究人员更加明确地开展针对每一类参数的学习研究; 另一方面, 使 BN 参数学习由局部参数学习转为全局参数学习, 提高了参数学习精度.

2) 提出了一种针对 BN 中父节点组合状态相同而子节点状态不同的参数的学习方法即 BABE 算法. 值得注意的是, 针对 BN 中父节点组合状态相同而子节点状态不同的参数的学习方法尚不多见. 因此, BABE 算法的提出为改善小数据集条件下 BN 参数的学习提供了更多途径.

3) 在 BABE 算法的基础上, 结合文献 [14] 提出的 IRE 算法, 提出了一种基于双重约束的 BN 参数学习方法即 DCE 算法. 该算法具有以下特性: 1) 相同数据条件下可以提高参数学习的精度; 2) 相同参数精度要求下可以减少对数据的需求; 3) 相同观测证据条件下可以提高推理精度.

尽管本文算法存在上述优点, 但算法仍然存在进一步提升的空间: 1) 本文算法实时性还不尽如人意, 而造成这一问题的主要原因是第一重参数约束学习过程中 Beta 分布参数的求解比较费时. 因此, 寻找更加高效的求解方法以提高本文算法的实时性将是下一步研究的重点之一; 2) DCE 算法学习精度的提高较 BABE 算法还不够明显, 而造成这一问题的主要原因是第二重参数约束学习对参数改善的力度还不够. 因此, 通过改善第二重参数约束学习的结果以提高双重参数约束学习的精度将是下一步研究的另外一个重点.

References

- 1 Peal J. *Probabilistic Reasoning in Intelligent Systems*. Massachusetts: Morgan Kaufmann, 1988
- 2 Tamda Y, Imoto S, Araki H. Estimating genome-wide gene networks using nonparametric Bayesian network models on massively parallel computers. *IEEE Transactions on Computational Biology and Bioinformatics*, 2011, **3**(8): 683–697
- 3 Ibrahim W, Beiu V. Using Bayesian networks to accurately calculate the reliability of complementary metal oxide semiconductor gates. *IEEE Transactions on Reliability*, 2011, **60**(3): 538–549
- 4 Jin Nai-Gao, Yin Fu-Liang, Chen Zhe. Audio-visual speaker tracking based on dynamic Bayesian network. *Acta Auto-*

- matica Sinica*, 2008, **34**(9): 1083–1089
(金乃高, 殷福亮, 陈喆. 基于动态贝叶斯网络的音视频联合说话人跟踪. 自动化学报, 2008, **34**(9): 1083–1089)
- 5 Chen Hai-Yang, Gao Xiao-Guang, Fan Hao. Inference algorithm of variable structure DDBNs and multi-target recognition. *Acta Aeronautica et Astronautica Sinica*, 2010, **31**(11): 2222–2227
(陈海洋, 高晓光, 樊昊. 变结构 DDBNS 的推理算法与多目标识别. 航空学报, 2010, **31**(11): 2222–2227)
 - 6 Du You-Tian, Chen Feng, Xu Wen-Li. Approach to human activity multi-scale analysis and recognition based on multi-layer dynamic Bayesian network. *Acta Automatica Sinica*, 2009, **31**(11): 225–232
(杜友田, 陈峰, 徐文立. 基于多层动态贝叶斯网络的人的行为多尺度分析及识别方法. 自动化学报, 2009, **31**(11): 225–232)
 - 7 Wan Jiu-Qing, Liu Qing-Yun. Data association in visual sensor networks based on high-order spatial-temporal model. *Acta Automatica Sinica*, 2012, **38**(2): 236–247
(万九卿, 刘青云. 基于高阶时空模型的视觉传感网络数据关联方法. 自动化学报, 2012, **38**(2): 236–247)
 - 8 Infantes G, Ghallab M, Ingrand F. Learning the behavior model of a robot. *Autonomous Robots*, 2011, **30**(3): 157–177
 - 9 Druzdzel M. Probabilistic Reasoning in Decision Support Systems: from Computation to Common Sense [Ph.D. dissertation], Carnegie Mellon University, USA, 1993
 - 10 Druzdzel M, van der Gaag L C. Building probabilistic networks: where do the numbers come from? *IEEE Transactions on Knowledge and Data Engineering*, 2000, **12**(4): 481–486
 - 11 Helsper E, Gaag L, Groenendal F. Designing a procedure for the acquisition of probability constraints for Bayesian networks. In: Proceedings of the 14th Conference on Engineering Knowledge in the Age of the Semantic Web. Northampton, UK: Springer, 2004. 280–292
 - 12 Wittig F, Jameson A. Exploiting qualitative knowledge in the learning of conditional probabilities of Bayesian networks. In: Proceedings of the 16th International Conference on Uncertainty in Artificial Intelligence. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000. 644–652
 - 13 Altendorf E, Restificar A, Dieterich T. Learning from sparse data by exploiting monotonicity constraints. In: Proceedings of the 21st International Conference on Uncertainty in Artificial Intelligence. Arlington, Virginia: AUAI Press, 2005. 18–26
 - 14 Feelders A, van der Gaag L C. Learning Bayesian network parameters under order constraints. *Journal of Approximate Reasoning*, 2006, **42**(1–2): 37–53
 - 15 Niculescu R, Mitchell T, Rao R B. Bayesian network learning with parameter constraints. *Journal of Machine Learning Research*, 2006, **7**(1): 1357–1383
 - 16 Campos C, Cozman F. Belief updating and learning in semi-qualitative probabilistic networks. In: Proceedings of the 21st International Conference on Uncertainty in Artificial Intelligence. Edinburgh, Scotland: AUAI Press, 2005. 153–160
 - 17 de Campos C P, Tong Y, Ji Q. Constrained maximum likelihood learning of Bayesian networks for facial action recognition. In: Proceedings of the Tenth European Conference on Computer Vision. Marseille, France: Springer, 2008. 168–181
 - 18 de Campos C P, Ji Q. Improving Bayesian network parameter learning using constraints. In: Proceedings of the 19th International Conference on Pattern Recognition. Tampa FL: IEEE, 2008. 1–4
 - 19 Isozaki T, Kato N, Ueno M. Minimum free energies with “data temperature” for parameter learning of Bayesian networks. In: Proceedings of the 20th International Conference on Tools with Artificial Intelligence. Ohio, USA: IEEE, 2008. 371–378
 - 20 Liao Wen-Hua, Qiang Ji. Learning Bayesian network parameters under incomplete data with domain knowledge. *Pattern Recognition*, 2009, **42**(11): 3046–3056
 - 21 Rui Chang, Wei Wang. Novel algorithm for Bayesian network parameter learning with informative prior constraints. In: Proceedings of the 2010 International Joint Conference on Neural Networks. Barcelona, Spain: IEEE, 2010. 1–8
 - 22 Brunk H. Maximum likelihood estimates of monotone parameters. *Annals of Mathematical Statistics*, 1955, **26**(11): 607–616
 - 23 Murphy K P. Dynamic Bayesian Networks: Representation, Inference and Learning [Ph.D. dissertation], University of California, Berkeley, USA, 1993
 - 24 Bidyuk P I, Terent A N, Gasanov A S. Construction and methods of learning of Bayesian networks. *Cybernetics and Systems Analysis*, 2005, **41**(4): 587–598
 - 25 Nielsen U, Pellet J, Elisseeff A. Explanation trees for causal Bayesian networks. In: Proceedings of the 24th International Conference on Uncertainty in Artificial Intelligence. Helsinki, Finland: AUAI Press, 2008. 427–434
 - 26 Thwaites T. Causal identifiability via chain event graphs. *Artificial Intelligence*, 2013, **195**(2): 291–315
 - 27 Höffgen K. Learning and robust learning of product distributions. In: Proceedings of 6th Annual Conference on Computational Learning Theory. New York, USA: ACM, 1993. 77–83
 - 28 Kullback S, Leibler R. On information and sufficiency. *Annals of Mathematical Statistics*, 1951, **22**(1): 79–86
- 郭志高 西北工业大学电子信息学院博士研究生. 主要研究方向为小数据集条件下贝叶斯网络参数学习.
E-mail: guozhigao2004@163.com
(GUO Zhi-Gao Ph.D. candidate at the School of Electronics and Information, Northwestern Polytechnical University. His research interest covers parameter learning of Bayesian networks from small data set.)
- 高晓光 西北工业大学电子信息学院教授. 主要研究方向为智能决策, 复杂系统建模与效能分析. 本文通信作者.
E-mail: cxg2012@nwpu.edu.cn
(GAO Xiao-Guang Professor at the School of Electronics and Information, Northwestern Polytechnical University. Her research interest covers intelligent decision, complex system modeling and effective analysis. Corresponding author of this paper.)
- 邱若海 西北工业大学电子信息学院博士研究生. 主要研究方向为小数据集条件下贝叶斯网络结构和参数学习.
E-mail: xfwtdrh@163.com
(DI Ruo-Hai Ph.D. candidate at the School of Electronics and Information, Northwestern Polytechnical University. His research interest covers structure and parameter learning of Bayesian networks from small data set.)