# 小数据集条件下基于数据再利用的 BN 参数学习

杨 宇 [1] 高晓光 [1] 郭志高 [1]

**摘 要** 着重研究了小数据集条件下结合凸约束的离散贝叶斯网络 (Bayesian network, BN) 参数学习问题, 主要任务是用先验知识弥补数据的不足以提高参数学习精度. 已有成果认为数据和先验知识是独立的, 在参数学习算法中仅将二者机械结合. 经过理论研究后, 本文认为数据和先验知识并不独立, 原有算法浪费了这部分有用信息. 本文立足于数据信息分类, 深入挖掘数据和先验知识之间的约束信息来提高参数学习精度, 提出了新的 BN 参数学习算法 — 凸约束条件下基于数据再利用的贝叶斯估计. 通过仿真实验展示了所提算法在精度和其他性能上的优势, 进一步证明数据和先验知识不独立思想的合理性.

**关键词** 贝叶斯网络, 参数学习, 小数据集, 数据信息分类, 数据再利用

**引用格式** 杨宇, 高晓光, 郭志高. 小数据集条件下基于数据再利用的 BN 参数学习. 自动化学报, 2015, **41**(12): 2058−2071

**DOI** 10.16383/j.aas.2015.c140838

## Learning BN Parameters with Small Data Sets Based by Data Reutilization

YANG Yu[1] GAO Xiao-Guang[1] GUO Zhi-Gao[1]

**Abstract** In this paper, parameters learning of discrete Bayesian networks (BNs) with small data sets with convex constraints is investigated, and the main task is improving the accuracy of parameter learning through offsetting the lack of data with prior knowledge. Data and prior knowledge are often mechanically integrated in most existing algorithms because they are treated independent. However, after a theoretical study, they are found dependent on each other, and the existing algorithms have dissipated this relevance. A novel parameter learning algorithm — Bayesian estimation based on data reutilization under convex constraints, is proposed with deeply mining the information between data and prior knowledge based on classification of data information. Finally, simulations demonstrate the advantages of novel algorithm in precision and other indexes, which in turn tells the dependance between data and prior information.

**Key words** Bayesian network (BN), parameter learning, small data sets, classification about data information, reutilization of data

**Citation** Yang Yu, Gao Xiao-Guang, Guo Zhi-Gao. Learning BN parameters with small data sets based by data reutilization. *Acta Automatica Sinica*, 2015, **41**(12): 2058−2071

1988 年, 美国加州大学的 Pearl 教授[1] 正式系统地提出贝叶斯网络 (Bayesian network, BN) 的概念. 因为在表示和推理不确定性知识方面拥有独特优势, BN 如今已广泛应用于各个领域, 如音频识别[2]、目标识别[3]、行为识别[4]、视频跟踪[5]、医疗诊断[6−7]、故障诊断[8]、风险评估[9] 和系统可靠性分析[10] 等, 所以对 BN 研究有着重大的意义.

贝叶斯网络由一个有向无环图和依附于图的条件概率集合组成, 在实际运用中, 需要针对具体的问题建立相应的网络, 称之为 BN 学习, 具体包括结构学习和参数学习. 依据经验知识, 领域专家比较容易确定一个 BN 的结构[11], 却很难准确地给出 BN 参数[12]. 倘若能获得充足的数据, 使用最大似然方法 (Maximum likelihood, ML)[13] 就能获得满足需求的参数. 然而, 因为物质代价或时间代价太大等原因, 从某些系统中获得充足的数据十分困难. 例如航空发动机故障诊断中的发动机故障数据和医疗诊断系统中的病例数据等, 这导致数据的数量较少, 不能满足参数学习需求, 这样的数据称为小数据集. 小数据集让估计参数精度降低, 可能使得 BN 的推理结果违背相关领域的常理, 显然这样的 BN 让领域专家难以接受. 在这种情况下, 提高小数据集条件下的 BN 参数学习精度很有必要. 幸运的是, 根据经验知识, 领域专家比较容易给出网络参数之间的定性约束信息[11, 14−15], 结合它们可以大幅提高小数据集条件下的 BN 参数学习精度.

目前, 结合专家约束的 BN 参数学习算法大致可以分为三类: 约束优化模型、保序回归模型和贝叶斯模型. 这三种模型从不同的角度挖掘专家约束中的信息, 以弥补小数据信息不足的缺点, 具体可归

收稿日期 2014-12-04 录用日期 2015-07-21
Manuscript received December 4, 2014; accepted July 21, 2015
国家自然科学基金 (60774064, 61573285), 教育部博士点基金 (20116102110026) 资助
Supported by National Natural Science Foundation of China (60774064, 61573285) and Research Fund for the Doctoral Program Higher Education of China (20116102110026)
本文责任编委 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

纳如下:

约束优化模型: 它将定性约束条件下的 BN 参数学习问题看作一种优化模型来处理, 根据约束使用方式的差异, 可以进一步分为惩罚函数优化模型和凸优化模型.

1) 惩罚函数优化模型: a) 目标函数: 先用数据构造似然函数, 因数据少, 导致似然函数精度较低, 不能独立作为优化的目标函数. 因此, 需要用专家约束构造惩罚函数来校正似然函数. 惩罚函数具有如下特点: 当参数满足专家约束时, 惩罚函数取值为 0, 即选择相信似然函数, 不作惩罚; 当参数违背约束时, 惩罚函数取值为负, 且随着违背的程度加重惩罚函数的绝对值越大. b) 可行域: 惩罚函数优化模型的可行域为参数空间 (即满足公理约束的高维欧氏空间). 关于惩罚函数优化模型的成果较多, Wittig 等[16]、Altendorf 等[17] 和 Liao 等[18] 都对该方法进行了探索, 其中 Liao 等将该思想引进到数据缺失情况下的参数学习问题, 提出了一种改良的 EM 算法. 2) 凸优化 (Convex optimization, CVX) 模型: a) 目标函数: 该方法的目标函数有似然函数[19] 和熵函数[20] 两种. b) 可行域: 不同于惩罚函数优化模型, 凸优化模型将专家约束作为公理约束的补充, 用以缩小搜索范围, 得到的参数自然满足约束信息. 目前仅有 de Campos 等[19−20] 做了相关的工作.

保序回归模型: 该模型将定性约束条件下的 BN 参数学习问题视为保序回归的一个特例处理. 首先根据定性影响约束, 在查询节点的父节点组合空间上定义偏序规则和参数更新方法, 然后用 Pool adjacent violators (PAV) 算法或 Minimum lower sets (MLS) 算法求解参数, 其中 PAV 算法只能求解连续集合上全序的保序回归问题, 故该思路主要用 MLS 求解, 而 MLS 是建立在下集的概念上的算法. 该模型的本质为用偏序关系中较小的参数的数据统计值去校正较大的参数, 以尽量使二者满足约束关系. Feelders 等[21] 在该思想上做了开创性工作, 郭志高等[15] 对它进行了发展.

贝叶斯模型: 将定性先验知识转化为定量先验知识的取值区域, 再用贝叶斯 MAP 思想求参数, 具体分两种:

1) 首先用 Monte Carlo 法抽样参数可行域, 得到先验参数, 再用一个正因子乘以先验参数作为 Dirichlet 分布的超参, 最后用 MAP 算法的参数估计公式求解参数. 这由 Chang 等[7, 22] 率先提出, 称为定性最大后验概率 (Qualitative maximum a posterior, QMAP).

2) 首先根据约束求出各参数的取值范围, 再用 Monte Carlo 法基于均匀分布进行抽样, 接着用 Beta 分布拟合抽样参数得到虚拟统计值, 最后用贝叶斯公式求解参数. 郭志高等[15] 的文章中用到了该思想.

综合分析以上三种参数学习模型, 它们各有优缺点.

1) 约束优化模型虽然可以结合几乎所有的约束类型, 但它的缺点更明显: a) 针对不同的约束类型需要构造不同的惩罚函数, 且将惩罚函数加到目标函数中的权重是人为赋予的, 有时并不可靠; b) 凸优化模型的可行域比较精确, 但估计参数基本位于可行域边界, 此时不等约束退化为相等约束, 浪费了大量信息, 故算法精度并不高.

2) 保序回归模型只能使用定性影响约束, 且最后结果不一定都满足约束, 在三种模型中精度最低, 并很有可能使本来较好的估计参数往较坏的方向移动.

3) 贝叶斯方法精度最高, 但 Chang 等的方法中没有给出因子的确定方法, 需要用户根据信念选取; Guo 等的方法只适用于单调约束, 且其基于均匀分布抽样各参数取值区间的观点并不合理, 因为这样不能保证先验参数在可行域中均匀分布, 违背了最大熵原理.

更重要的是, 上述算法都认为数据和先验知识是独立的, 在参数学习过程中仅将二者机械结合. 但数据和专家约束并不独立, 它们之间存在约束信息. 因为根据似然性概念, 在参数空间中似然函数的最大值点为最大似然估计, 且随着参数与最大似然估计之间距离的增加, 似然函数迅速下降[23]. 而依据经验给出的先验知识并非完全正确, 在运用中不能全盘接受. 可以使用似然函数的这个性质对照甄别筛选后再行利用. 具体而言, 先验参数中离最大似然估计较远的参数和数据的拟合程度很小, 可以认为它们没有 (或很少) 携带真实参数的信息, 本文视之为先验知识中的干扰信息, 对参数学习有负面影响. 在参数学习过程中, 过滤掉先验知识中的干扰信息, 存在提高估计参数精度的潜力. 至此, 本文认为数据对先验参数和后验参数都有影响, 将数据的这两部分信息区分并综合利用, 可提高参数学习的精度.

依据上述分析, 本文提出了一种新的 BN 参数学习算法, 它适用于小数据集条件下给定凸约束的离散 BN 参数学习问题. 首先, 对专家约束确定的可行域基于均匀分布进行 Monte Carlo 抽样; 然后用数据构造关于先验参数空间位置的定性约束 (Qualitative constraint about spatial position of BN parameters, QCSP), 过滤掉与数据拟合程度较低的先验参数, 因为数据由这部分参数对应的 BN 产生的概率较小; 接着先将高维先验参数分离成一维先验参数, 再用 Beta 分布对其拟合得到虚拟统计值; 最后对数据再次利用, 以贝叶斯公式求解参

数. 本文算法也运用了贝叶斯模型的思想, 但相较于 Chang 等的方法没有类似因子的不确定因素, 也比郭志高等提出的方法的适用性高, 抽样分布更合理. 后文的实验证明本文算法在精度上要优于 ML、CVX 和 QMAP 等几种典型算法, 推理结果也更加准确.

本文组织结构为: 第 1 节简要介绍 BN 参数学习的基本理论知识及所用的约束模型; 第 2 节详细叙述了基于数据再利用的 BN 参数学习算法, 包括提出算法的思路和算法本身; 第 3 节用仿真实验验证本文算法的性能; 第 4 节简要地总结本文工作.

## 1 相关理论

### 1.1 贝叶斯网络及其参数学习

一个 BN 定义为二元组 $B=(G,\Theta)$, 它由有向无环图 (Directed acyclic graph, DAG) $G=(U,E)$ (其中节点集 $U=\{X_1,X_2,\cdots,X_n\}$ 表示随机变量集合, 弧集 $E$ 表示变量间的直接依赖关系) 和关联于每个变量的条件概率集合 $\Theta$ 组成. 离散变量的概率分布通常表示为节点概率表, 它包含在 BN 中给定父节点值时子节点取每个值的概率, 记作 $P(X_i|pa(X_i))$. 其中, $pa(X_i)$ 表示 $G$ 中 $X_i$ 的父节点集合. 因此 BN 在 $U$ 上定义了由下式给出的简化联合概率分布:

$$P(X_1,X_2,\cdots,X_n)=\prod_{i=1}^{n}P(X_i|pa(X_i)) \tag{1}$$

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. 用 $r_i$ 表示随机变量 $X_i$ 的状态数, 用 $q_i$ 表示 $X_i$ 父节点集合 $pa(X_i)$ 的状态数, 则 $P(X_i=i|pa(X_i)=j)$ 表示父节点状态为 $j$, 第 $i$ 个节点取第 $k$ 个状态的参数, 记为 $\theta_{ijk}$, 其中 $1\le i\le n$, $1\le j\le q_i$, $1\le k\le r_i$. 第 $i$ 个节点的所有参数构成了一个 $q_i\times r_i$ 维的矩阵, 称为该节点的概率分布表 (Node probability table, NPT).

若用 $N_{ijk}$ 表示样本数据中父节点状态为 $j$、第 $i$ 个节点取第 $k$ 个状态的统计值, 则参数 $\theta_{ijk}$ 的最大似然估计 (Maximum likelihood estimation, MLE) 估计为

$$\theta_{ijk}=\frac{N_{ijk}}{\sum\limits_{k}N_{ijk}} \tag{2}$$

假设 BN 参数的先验分布为 Dirichlet 分布 (超参数 $T=\{\tau_{ijk}\}$), 因为 Dirichlet 为共轭多项式分布, 所以其后验分布也为 Dirichlet 分布, 若数据的统计值为 $N=\{N_{ijk}\}$, 则有

$$P(\theta|D,T,G)\propto\prod_{i=1}^{n}\prod_{j=1}^{q_i}\prod_{k=1}^{r_i}\theta_{ijk}^{\tau_{ijk}+N_{ijk}-1} \tag{3}$$

当参数的后验分布取极大值时, 得估计参数为

$$\theta_{ijk}=\frac{N_{ijk}+\tau_{ijk}}{\sum\limits_{k}(N_{ijk}+\tau_{ijk})} \tag{4}$$

上式中, 先验 Dirichlet 分布的超参数 $\tau_{ijk}$ 可认为是虚拟数据统计值, $N_{ijk}$ 是样本统计值. 式 (4) 为贝叶斯方法和最大后验概率 (Maximum a posterior, MAP) 结合的产物, 称之为贝叶斯 MAP 估计.

### 1.2 约束模型介绍

一般地, 在 BN 中, 领域知识可以定义为条件概率上的不同约束, 这些约束可以描述成线性回归函数 $f(\theta_{ijk})\le c$[22] ($c$ 是常数). 特别地, 从线性回归函数中可以总结出如下两种约束形式:

1) 范围约束 (Range constraints), 定义了 BN 中任意参数的上下界, 如式 (5) 所示.

$$0\le\alpha_{ijk}\le\theta_{ijk}\le\beta_{ijk}\le 1 \tag{5}$$

2) 不等式约束 (Inequation constraints), 定义了一对参数的相对大小关系, 如式 (6) 所示.

$$\theta_{ijk}\le\theta_{i'j'k'},\quad i\ne i'\vee j\ne j'\vee k\ne k' \tag{6}$$

除以上两种约束模型外, 定性约束中还有一类常见约束 — 近似相等约束 (Approximate-equality constraints)[24]. 领域专家可能给出这样的断言: 某参数和另外一个参数大致相等. 这个断言非常符合人类的思维方式, 又避开了直接给出准确参数值的不可靠性. 它可以用数学符号表示为

$$\theta_{ijk}\approx\theta_{i'j'k'},\quad i\ne i'\vee j\ne j'\vee k\ne k' \tag{7}$$

式 (7) 不便利用, 现引入一个很小的正有理数 $\varepsilon$, 当参数间的差距小于它时, 认为它们近似相等. 则可将式 (7) 改写为

$$|\theta_{ijk}-\theta_{i'j'k'}|<\varepsilon,\quad i\ne i'\vee j\ne j'\vee k\ne k' \tag{8}$$

当然, 一定还存在其他形式的定性约束 (如本文第 2.1.1 节用数据构造的约束), 不一一赘述.

**定义 1.** 设 $O$ 为非空集合, 若 $\alpha\in O$ 且 $\beta\in O$ 推出 $t\alpha+(1-t)\beta\in O$, 则称集合 $O$ 为凸集.

**定义 2.** 若 BN 参数之间的约束确定的参数可行域为凸集, 则称这样的约束为凸约束.

**定义 3.** 若关于 BN 参数的约束集合由不同类型的凸约束组成, 则容易证明其确定的参数可行域亦为凸集. 本文称这样的约束集合为凸约束集合, 简称凸约束.

BN 参数之间的凸约束符合人类对数据的直观思维习惯 (例如领域专家可能说两个参数谁大一些或者大概相等), 且其包含的约束种类多, 信息量大, 约束力强, 相较单一或少数种类约束有极大的优势. 前文重点分析的 CVX 和 QMAP 算法和提出的新算法都是基于凸约束的参数学习方法, 相比其他算法, 它们对约束类型有更好的通用性.

## 2 基于数据再利用的 BN 参数学习

实际中, 数据是由真实 BN 产生的, 而正确的先验知识刻画了一部分 BN 的性质, 可将先验知识理解为产生该数据的条件之一. 从概率论的角度看, 因和果都会制约对方发生的概率, 故二者不独立. 而先验知识是领域专家根据经验给出, 并不精确, 在利用前可以先行甄选, 而数据与先验参数的似然度恰能完成这个任务. 为充分挖掘数据携带的信息以提高参数学习的精度, 本章首先探讨数据与先验知识的不独立, 接着根据数据对先验参数的甄别能力构造 QCSP 过滤先验知识中的干扰信息, 最后提出一种精度更高的新算法.

### 2.1 数据对先验参数分布的影响

BN 参数与数据的对数似然函数可以表示为

$$l(\theta, D) \propto \sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i} N_{ijk}\log\theta_{ijk} \tag{9}$$

已经证明[23], 当 $\theta_{ijk}$ 取如下值 (即最大似然估计) 时, 似然函数取得最大值:

$$\theta_{ijk}^{ML} = \begin{cases} \dfrac{N_{ijk}}{\sum\limits_{k} N_{ijk}}, & \text{若} \sum\limits_{k} N_{ijk} > 0 \\ \dfrac{1}{r_i}, & \text{若} \sum\limits_{k} N_{ijk} = 0 \end{cases} \tag{10}$$

显然, 最大似然估计 $\theta^{ML}$ 是参数空间的一个内点. 假设 BN 参数的分布 $p(\theta)$ 在 $\theta^{ML}$ 周围光滑且不为零, 则有式 (11).

$$\begin{aligned} p(\theta|D) &= \frac{p(\theta)p(\theta)|D}{p(D)} = \frac{p(\theta)\exp\{l(\theta,D)\}}{p(D)} \propto \\ &\frac{p(\theta)}{p(D)}\exp\left\{\sum_{i=1}^{n}\sum_{j=1}^{q_i}\sum_{k=1}^{r_i} N_{ijk}\log\theta_{ijk}\right\} = \\ &\frac{p(\theta)}{p(D)}\left(\exp\left\{\sum_{i=1}^{n}\sum_{j=1}^{q_i} N_{ij\cdot}\left(\sum_{k=1}^{r_i}\theta_{ijk}^{ML}\log\frac{\theta_{ijk}}{\theta_{ijk}^{ML}} + \right.\right.\right. \\ &\left.\left.\left.\sum_{k=1}^{r_i}\theta_{ijk}^{ML}\log\theta_{ijk}^{ML}\right)\right\}\right) \end{aligned} \tag{11}$$

于是,

$$\begin{aligned} p(\theta|D) = \frac{p(\theta)}{p(D)} \times \\ \prod_{i=1}^{n}\prod_{j=1}^{q_i}\left(\exp\left\{\left(\sum_{k=1}^{r_i}\theta_{ijk}^{ML}\log\frac{\theta_{ijk}}{\theta_{ijk}^{ML}} + \right.\right.\right. \\ \left.\left.\left.\sum_{k=1}^{r_i}\theta_{ijk}^{ML}\log\theta_{ijk}^{ML}\right)\right\}\right)^{N_{ij\cdot}} \end{aligned} \tag{12}$$

其中, $N_{ij\cdot} = \sum_k N_{ijk}$. 当数据已定时, 根据假设, $p(\theta|D)$ 作为 $\theta$ 的函数, 在 $\theta^{ML}$ 达到唯一的最大值. 由式 (12) 可知, 当 $\theta$ 与 $\theta^{ML}$ 的距离增大时, $p(\theta|D)$ 的值将减小, 并且随着样本量的增加而加速减小.

由此可见, 数据对先验参数的分布有明显的影响, 故数据和先验知识不独立. 此外, 离最大似然估计较远的先验参数和数据的拟合程度很小, 它们没有 (或很少) 携带真实参数的信息, 反而拉低了先验参数与真实参数的契合度, 对参数学习有负面影响, 本文视之为干扰信息, 学习参数时应构造约束将其过滤, 让先验参数与真实参数的契合度更高, 以便更有效地驱使估计参数向靠近真实参数的方向运动. 式 (12) 虽然刻画了数据对可行域中参数分布的影响, 但其本身过于复杂, 用它来构造约束显得繁琐, 而根据前面的分析, $p(\theta|D)$ 的值和 $\theta$、$\theta^{ML}$ 的距离是负相关的, 故可以采用 $\theta$ 与 $\theta^{ML}$ 之间的欧氏距离来构造约束.

### 2.2 过滤先验知识中的干扰信息

#### 2.2.1 构造 QCSP

可行域由专家约束 $\Omega$ 确定, 为方便叙述, 本文记之为 $\Theta^{\Omega}$. 过滤干扰信息需要分割, 但 $\Theta^{\Omega}$ 是一个高维空间, 无法像在现实空间中一样形象直观地实施分割操作, 为了表达清晰, 现引入度量空间的相关概念.

**定义 4.** $O$ 是非空集合, 若 $\forall x, y \in O$, 有一实数 $d(x, y)$ 与之对应, 且满足:

1) $d(x, y) \geq 0$ (非负性);

2) $d(x, y) = 0$ 当且仅当 $x = y$ (非退化性);

3) $d(x, y) = d(y, x)$ (对称性);

4) $d(x, y) \leq d(x, z) + d(z, y)$, $\forall z \in O$ (三角不等式).

则 $d(x, y)$ 称为 $O$ 中的一个度量 (距离), 定义了 $d$ 的集合 $O$ 称为度量空间, 记为 $(O, d)$, 简记 $O$.

**定义 5.** 设 $(O, d)$ 为度量空间, $x_0 \in O$, $r > 0$. 令

$$B(x_0, r) = \{x \in O | d(x_0, x < r)\}$$

$$S(x_0,r)=\{x\in O|d(x_0,x=r)\}$$
$$\bar{B}(x_0,r)=\{x\in O|d(x_0,x\le r)\}$$

称 $B(x_0,r)$ 为以 $x_0$ 为中心 $r$ 为半径的开球, $\bar{B}(x_0,r)$ 为以 $x_0$ 为中心 $r$ 为半径的闭球, $S(x_0,r)$ 为以 $x_0$ 为中心 $r$ 为半径的球面.

由此, 本文在参数空间 $\Theta$ 中定义 $ed(\theta^{(1)},\theta^{(2)})$:

$$ed\left(\theta^{(1)},\theta^{(2)}\right)=\sqrt{\sum_i\sum_j\sum_k\left(\theta_{ijk}^{(1)}-\theta_{ijk}^{(2)}\right)^2} \quad (13)$$

容易证明, $ed(\theta^{(1)},\theta^{(2)})$ 满足定义 4 中的 4 个基本性质, 则它是 $\Theta$ 中的一个度量, 同时将参数空间记为 $(O,ed)$. 因可行域 $\Theta^{\Omega}$ 包含于 $\Theta$, 故可用度量 $ed$ 辅助构造约束, 过滤可行域中的干扰信息.

用数据构造 QCSP 的步骤为 (参见图 1):

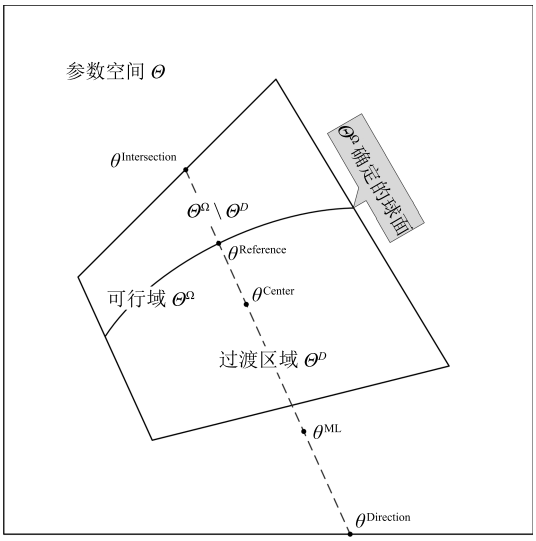



图 1 用数据构造 QCSP 的二维示意图

Fig. 1 Planar diagrammatic sketch for building QCSP with data

**步骤 1.** 首先用 MLE 算法估计参数得到 $\theta^{\mathrm{ML}}$, 特别地, 当同父节点状态参数都为 0 时, 将它们设置为 $1/n$, $n$ 为子节点的状态数 (否则 $\theta^{\mathrm{ML}}\notin\Theta$, 导致后面的步骤无法继续);

**步骤 2.** 依据最大熵原理, 基于均匀分布用 Monte Carlo 法抽样可行域得到先验参数, 求先验参数的平均值得到可行域中心 $\theta^{\mathrm{Center}}$;

**步骤 3.** 用搜索算法求高维直线 $\overline{\theta^{\mathrm{ML}}\theta^{\mathrm{Center}}}$ 与可行域边界的一个交点 $\theta^{\mathrm{Intersection}}$, 它在高维直线上与 $\theta^{\mathrm{ML}}$ 位于 $\theta^{\mathrm{Center}}$ 的异侧;

**步骤 4.** 用搜索算法求高维直线 $\overline{\theta^{\mathrm{ML}}\theta^{\mathrm{Center}}}$ 与参数空间的一个交点 $\theta^{\mathrm{Direction}}$, 它在高维直线上与 $\theta^{\mathrm{ML}}$ 位于 $\theta^{\mathrm{Center}}$ 的同侧;

**步骤 5.** 求高维线段 $\overline{\theta^{\mathrm{ML}}\theta^{\mathrm{Center}}}$ 中点 $\theta^{\mathrm{Reference}}$, 它是一个重要的参考点;

**步骤 6.** 构造约束形如式 (14) 的 QCSP 约束.

$$ed(\theta,\theta^{\mathrm{Direction}})\le ed(\theta^{\mathrm{Reference}},\theta^{\mathrm{Direction}}) \quad (14)$$

用式 (14) 在 $(O,ed)$ 中定义一个以 $\theta^{\mathrm{Direction}}$ 为中心的闭球 $\bar{B}(\theta^{\mathrm{Direction}},r)$ (图 1 中的弧线即为其球面), 它和可行域 $\Theta^{\Omega}$ 的交集即是过滤干扰信息后的先验参数的集合, 本文称之为有效区域 $\Theta^D$, 称 $\Theta^D$ 中的参数为有效先验参数. 由于球是凸集, 所以式 (14) 是凸约束. QCSP 将可行域分割成两部分, 满足式 (13) 的部分即所求的有效区域 (如图 1 中 $\Theta^D$ 所示), 不满足的部分即视为干扰信息 (如图 1 中 $\Theta^{\Omega}\setminus\Theta^D$ 所示).

综上, 有效区域 $\Theta^D$ 可表示为

$$\left\{\theta\in\Theta^{\Omega}|ed(\theta,\theta^{\mathrm{Direction}})\le ed(\theta^{\mathrm{Reference}},\theta^{\mathrm{Direction}})\right\} \quad (15)$$

### 2.2.2 构造 QCSP 合理性的实验验证

前文阐述了用数据构造 QCSP 的思路, 下面将用三个变量控制实验 (实验 1 ∼ 3) 来进一步说明构造 QCSP 的合理性. 首先介绍一下试验网络模型.

本文实验所用的模型为草坪湿润模型 (如图 2 所示), 它由加州大学伯克利分校 Murphy 教授设计[25], 被广泛应用于各种参数学习算法的评估. 该模型由多云 $C$、洒水 $S$、下雨 $R$ 和草坪湿润 $W$ 四个节点组成, 节点间的因果关系如图 2 中箭头所示, 每个节点都是可能状态为 $\{0,1\}$ 的二值节点. 虽然草坪湿润模型非常简单, 但根据马尔科夫独立性 (链规则), 复杂 BN 的参数学习可以分解为同父节点参数学习问题, 并且任何 BN 都有顺连、分连和汇连三种基本结构构成, 在具体的参数学习过程中, 可以将复杂 BN 分解为一系列的基本结构, 进而分解为每个节点进行参数学习, 该模型包含三种基本结构, 故具有代表性. 本次实验学习由草坪湿润节点 $W$ 及其父节点洒水节点 $S$ 和下雨节点 $R$ 组成的子网络的参数 $P(W|S,R)$, 其他网络参数都事先设置, 并列于

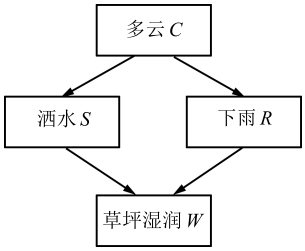



图 2 草坪湿润 BN

Fig. 2 BN of lawn wetness

表 1. 表 2 为约束备选集合, 实验时会根据要求从中抽取约束, 由第 2.2 节可知, 备选约束均为凸约束.

表 1　草坪湿润 BN 参数预置

Table 1　Default parameters for BN of lawn wetness

| 节点 | 参数 | |
|---|---|---|
| 多云 $C$ | $p(c=1)=0.5$ | $p(c=0)=0.5$ |
| 洒水 $S$ | $p(s=1\|c=1)=0.8$ | $p(s=0\|c=1)=0.2$ |
| | $p(s=1\|c=0)=0.9$ | $p(s=0\|c=0)=0.1$ |
| 下雨 $R$ | $p(r=1\|c=1)=0.8$ | $p(r=0\|c=1)=0.2$ |
| | $p(r=1\|c=0)=0.01$ | $p(r=0\|c=0)=0.99$ |

表 2　草坪湿润模型W节点参数的约束备选集合

Table 2　Alternative parameter constraints of BN of lawn wetness

| 编号 | 约束 |
|---|---|
| 1 | $p(w=1\|s=1,r=0)>p(w=0\|s=1,r=0)$ |
| 2 | $p(w=1\|s=0,r=1)>p(w=1\|s=0,r=0)$ |
| 3 | $p(w=1\|s=1,r=1)>p(w=0\|s=1,r=1)$ |
| 4 | $p(w=1\|s=1,r=1)>p(w=1\|s=1,r=0)$ |
| 5 | $p(w=0\|s=0,r=0)>p(w=1\|s=0,r=0)$ |
| 6 | $p(w=1\|s=1,r=1)>p(w=1\|s=0,r=0)$ |
| 7 | $p(w=1\|s=1,r=0)>p(w=1\|s=0,r=0)$ |
| 8 | $p(w=1\|s=1,r=1)>p(w=0\|s=1,r=1)$ |
| 9 | $p(w=1\|s=1,r=1)>p(w=1\|s=0,r=1)$ |

为了阐述更清晰, 这里定义一个概念 — QCSP 的正确率, 其定义式为

$$R=\frac{N_{\text{Correct}}}{N_{\text{Sampled}}} \tag{16}$$

其中, $R$ 为 QCSP 的正确率; $N_{\text{Sampled}}$ 为对同一 BN (同结构和参数, 下同) 抽样并构造 QCSP 的次数, 本文设置为 20; $N_{\text{Correct}}$ 为对同一 BN 抽样 20 次中真实参数满足 QCSP 的次数. 显然, $R$ 越大, 越能说明用数据构造约束的思路的合理性.

**实验 1.** QCSP 正确率之样本量影响实验. 1) 样本量 (控制变量) 从 1 到 35 逐渐增加; 2) 每个样本量条件下, 约束数量 (随机变量) 在 1～9 之间随机产生, 并且约束从约束备选集合中随机抽取; 3) 每个样本量条件下, 真实参数位置 (随机变量) 未知, 每次实验都从由所有备选约束确定的可行域中随机抽取; 4) 在每个条件 1) 下, 按条件 2) 和条件 3) 构造 BN 求 QCSP 的正确率并重复 10 次, 并将最后结果求平均值. 实验结果如图 3 所示.

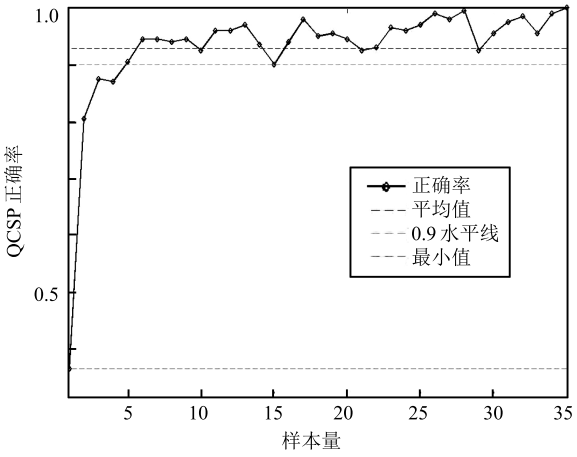



图 3　QCSP 正确率的影响因素之样本量

Fig. 3　Accuracy of QCSP for altering data size

**实验 2.** QCSP 正确率之约束数量影响实验: 1) 约束数量 (控制变量) 量从 1～9 逐渐增加; 2) 每个约束数量条件下, 样本量 (随机变量) 在 1 到 35 之间随机产生; 3) 每个约束数量条件下, 真实参数位置 (随机变量) 未知, 每次实验都从由所有备选约束确定的可行域中随机抽取; 4) 在每个条件 1) 下, 按条件 2) 和条件 3) 构造 BN 求 QCSP 的正确率并重复 20 次, 并将最后结果求平均值. 实验结果如图 4 所示.

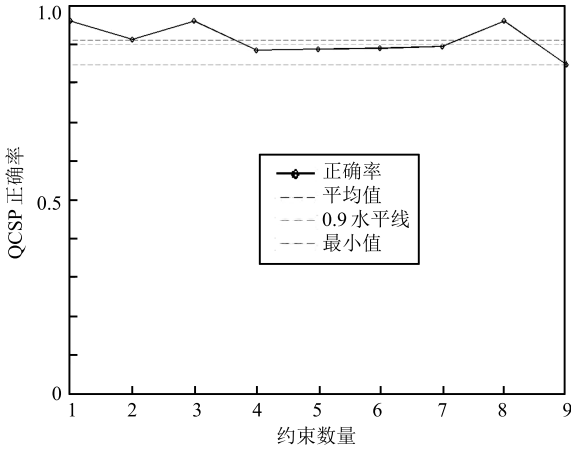



图 4　QCSP 正确率的影响因素之约束数量

Fig. 4　Accuracy of QCSP for altering the number of constraints

**实验 3.** QCSP 正确率之真实参数位置影响实验. 1) 真实参数位置 (控制变量) 分三组 — 可行域的中心、可行域中心与边界之间以及可行域边界; 2) 每个真实参数位置条件下, 约束数量 (随机变量) 在 1～9 之间随机产生, 并且约束从约束备选集合中随机抽取; 3) 每个真实参数位置条件下, 样本量 (随机变量) 在 1～35 之间随机产生; 4) 在每个条件 1)

下, 按条件 2) 和条件 3) 构造 BN 求 QCSP 的正确率并重复 100 次, 并将最后结果求平均值. 实验结果如图 5 所示.

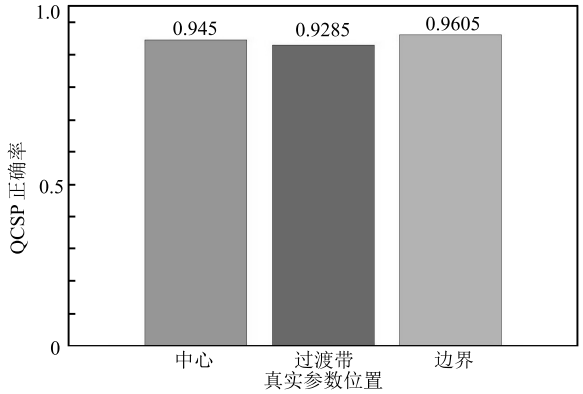



图 5 QCSP 正确率的影响因素之真实参数位置

Fig. 5 Accuracy of QCSP for altering true parameters

由图 3 可知, 1) 当样本量不超过 4 时, QCSP 的正确率不超过 0.9. 其中, 当样本量为 2、3 和 4 时, QCSP 的正确率大约为 0.8 ∼ 0.9; 当样本量为 1 时, QCSP 的正确率约为 0.37, 此时最好对 QCSP 进行调整, 否则应该放弃使用本算法; 2) 随着样本量增加, QCSP 的正确率逐渐趋近 1. 因为样本数据提供真实参数在可行域中的方位信息, 且样本量越大, 方位越准确, 又因 QCSP 是基于参数在可行域中的方位定义的, 所以随着样本量增加, QCSP 的正确率逐渐增加, 最后趋近 1.

从图 4 可以看出, 约束数量对 QCSP 的正确率影响较小, 这是因为约束很少 (或者没有) 携带参数在可行域中的方位信息. 所以, 本文算法对于任意约束数量的情况都适用.

由图 5 可知, 真实参数位于可行域中心 (图 5 左边分支)、真实参数位于可行域中心与边界之间 (图 5 中间分支) 和真实参数位于可行域边界 (图 5 右边分支) 时, QCSP 的正确率略有差异, 但都明显超过 0.9. 这说明, 本文算法对真实参数在可行域的任意位置都适用.

本小节实验不仅进一步证明了用数据构造 QCSP 的合理性, 也确定了本文算法的适用条件 (即样本量不能过少). 对比三个实验发现, 仅有样本量对构造 QCSP 的正确率有较大影响, 但当样本量大于等于 2 时, 影响就很小. 因此, 只在样本量极少时不宜使用本文算法, 具体的量化标准如式 (17) 所示. 例如, 本文仿真模型的样本量至少为 2 (该数字只针对本模型), 本文算法才具有明显优势 (第 3 节的仿真正好说明该问题).

$$R = \frac{N_{\text{Correct}}}{N_{\text{Sampled}}} \geq 0.8 \tag{17}$$

## 2.3 提出新算法

### 2.3.1 基于数据在利用的 BN 参数学习

在求有效区域 $\Theta^D$ 时, 数据已经被利用了一次, 如果认为数据携带的信息已经完全挖掘, 即数据对于 $\Theta^\Omega$ 中的参数分布无影响, 则 $\Theta^\Omega$ 中的参数分布规律为均匀分布, 但事实并非如此. 要说明这个问题, 必须要分辨先验参数和后验参数两个概念. 在过虑干扰信息时, 是数据作用于先验参数, 这个步骤一定会改变 BN 参数的后验分布, 但如式 (3) 所示, 参数的后验分布由数据和 Dirichlet 分布超参两部分决定, 被改变的只是后者, 而不是消除了数据对参数后验分布的影响. 因此, 在最终的参数估计过程中, 本文算法将再次用到数据, 以充分利用其携带的信息.

上述内容说明, 样本数据携带的信息可以分为两部分: 1) 刻画先验参数的后验概率, 转换成约束后可过滤先验知识中的干扰信息; 2) 刻画后验参数的概率分布, 作为 MAP 思想求解参数的基础. 这两部分信息是独立存在的, 并不会因为运用了一类信息后另一类信息消失. 该思想是本文率先提出, 称之为数据信息分类.

过滤了先验知识中的干扰信息后, 就能进行参数估计, 这需要先从过滤干扰信息后的先验知识中提炼出 Dirichlet 分布超参. 首先, 将过滤干扰信息后的先验参数投影到每一维上, 得到 $\sum_i q_i r_i$ 个一维参数集合; 接着, 基于二阶矩法用 Beta 分布去拟合每一个参数集合, 得到对应的 Beta 分布的参数和, 其数学模型如下式所示:

$$\begin{aligned} &\min_{a_{ijk},b_{ijk}} (D_{\text{Beta}} - D_{\theta_{ijk}})^2 \\ &\text{s.t.} \quad E_{\text{Beta}} = E_{\theta_{ijk}}, \ \ a_{ijk}, b_{ijk} > 0 \end{aligned} \tag{18}$$

其中, $D_{\text{Beta}}$ 和 $D_{\theta_{ijk}}$ 分别为 Beta 分布和过渡参数的方差, $E_{\text{Beta}}$ 和 $E_{\theta_{ijk}}$ 分别为 Beta 分布和过渡参数的均值, $a_{ijk}$ 和 $b_{ijk}$ 就是所求的 Beta 分布的参数.

约束条件下求解 BN 参数的过程可以用如下的优化模型表示:

$$\begin{aligned} &\arg\max \log P(\theta|D, \Omega, G) = \\ &\quad \log P(D|\theta, G) + \log P(\theta|\Omega, G) + \log P(D) \end{aligned} \tag{19}$$

为了上式的最优解 (即所求参数) 一定满足专家约束, 本文将优化模型的搜索空间限制在可行域 $\Theta^\Omega$, 即后验参数的分布函数可表示为式 (20) (参见图 6).

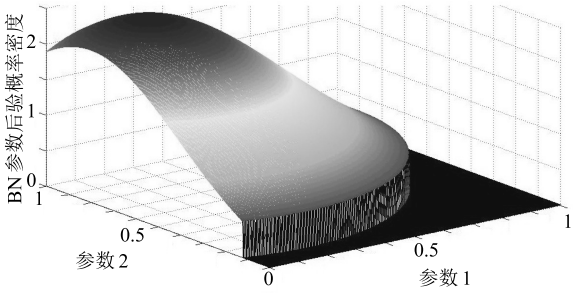



图 6　有效区域中后验参数分布的三维示意图

Fig. 6　Three-dimensional diagrammatic sketch for posterior distribution of parameters over effective area

$$P(\theta|D,\Omega,G)=\begin{cases}f(\theta), & \text{若 } \theta\in\Theta^{\Omega}\\ 0, & \text{若 } \theta\in\Theta\setminus\Theta^{\Omega}\end{cases}\tag{20}$$

上式中, $f(\theta)\geq 0$ 且 $\int_{\Theta^{\Omega}} f(\theta)\mathrm{d}\theta=1$.

设 $\theta\in\Theta^{\Omega}$, 在离散 BN 中, 它的后验分布可表示为

$$P(\theta|D,\Omega,G)\propto\prod_{i=1}^{n}\prod_{j=1}^{q_i}\prod_{k=1}^{r_i}\theta_{ijk}^{\tau_{ijk}+N_{ijk}-1}\tag{21}$$

为了保证估计参数不违背专家约束, 把后验参数限制在可行域中, 将式 (21) 在可行域中作归一化处理, 即可得 BN 参数的后验分布概率密度:

$$f(\theta)=\frac{1}{K}\prod_{i=1}^{n}\prod_{j=1}^{q_i}\prod_{k=1}^{r_i}\theta_{ijk}^{\tau_{ijk}+N_{ijk}-1}\tag{22}$$

式 (21) 中, 归一化因子

$$K=\int_{\Theta^{\Omega}}\prod_{i=1}^{n}\prod_{j=1}^{q_i}\prod_{k=1}^{r_i}\theta_{ijk}^{\tau_{ijk}+N_{ijk}-1}\mathrm{d}\theta$$

因此, 后验参数在参数空间中的概率密度函数为

$$P(\theta|D,\Omega,G)=\begin{cases}\dfrac{1}{K}\displaystyle\prod_{ijk}\theta_{ijk}^{\tau_{ijk}+N_{ijk}-1}, & \text{若 } \theta\in\Theta^{\Omega}\\ 0, & \text{若 } \theta\in\Theta\setminus\Theta^{\Omega}\end{cases}\tag{23}$$

由式 (23) 可知, 当 BN 参数的后验概率取最大值时, 就得到了参数在约束条件下的 MAP 估计, 因本文的 Dirichlet 超参取得合适, 使得 MAP 估计基本位于可行域中, 即得式 (4). 令 $\tau_{ijk}=a_{ijk}$ 且 $\sum_k\tau_{ijk}=a_{ijk}+b_{ijk}$, 得到最终的参数估计公式为

$$\theta_{ijk}=\frac{a_{ijk}+N_{ijk}}{(a_{ijk}+b_{ijk})+\sum\limits_{k}N_{ijk}}\tag{24}$$

从上述求解参数的过程可知, 数据在两处得到利用: 一是用数据构造 QCSP 过滤先验知识中的干扰信息; 二是用式 (24) 求解参数, 这就是数据再利用, 它和数据信息分类是相辅相成的. 鉴于新算法用到了贝叶斯思想以及其数据再利用特点, 本文称之为凸约束条件下基于数据再利用的贝叶斯估计 (Bayesian estimation based on reutilization of data under convex constraints, BERD).

### 2.3.2　新算法流程

总结前文求解参数的过程, 可将本文算法归纳为如下流程 (参见图 7):

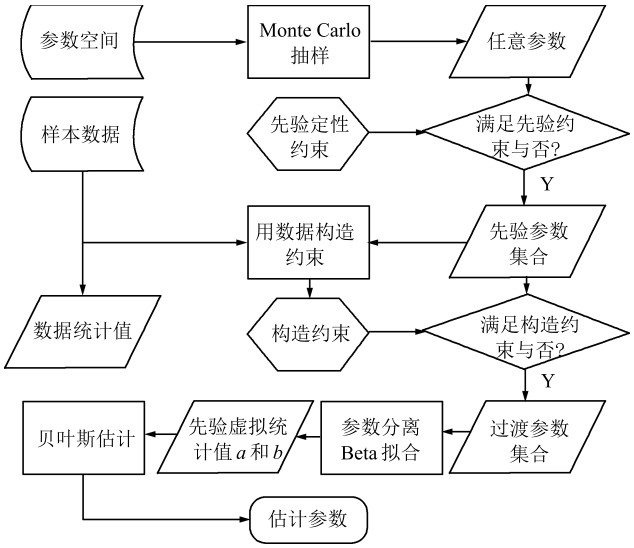



图 7　本文算法的流程图

Fig. 7　The flow chart of novel algorithm

**步骤 1.** 统计样本, 获取样本统计值 $N_{ijk}$;

**步骤 2.** 基于均匀分布规律, 用 Monte Carlo 法抽样可行域 $\Theta^{\Omega}$, 获得先验参数集合;

**步骤 3.** 根据第 2.2.1 节中的步骤构造 QCSP;

**步骤 4.** 用 QCSP 过滤先验参数集合, 获取有效参数集合.

**步骤 5.** 基于二阶矩法用 Beta 分布拟合有效先验参数, 得到对应的 Beta 分布参数 $a_{ijk}$ 和 $b_{ijk}$;

**步骤 6.** 将 $a_{ijk}$、$b_{ijk}$ 和 $N_{ijk}$ 带入式 (23) 求解参数.

## 3　仿真实验

本文从精度[1]、实用性和时间复杂度三个方面对 ML、CVX、QMAP 和 BERD 算法进行分析和对比实验, 以验证新算法是否具有优势. 实验 BN 模型仍为第 2.2.2 节中的草坪湿润模型, 只是引入范围约束和近似相等约束, 将约束备选集合替换为表 3.

[1]本文的 BN 参数学习精度均用 KL 散度表示, 它是一种概率分布之间的差异的度量.

表 3 仿真实验草坪湿润模型W节点参数的约束备选集合

Table 3 Alternative parameter constraints of BN of lawn wetness for simulation experiments

| 编号 | 约束 |
|---|---|
| 1 | $0.9 > p(w=0\|s=0,r=0) > 0.55$ |
| 2 | $p(w=0\|s=0,r=0) > p(w=1\|s=0,r=0)$ |
| 3 | $p(w=1\|s=0,r=1) > p(w=0\|s=0,r=1)$ |
| 4 | $p(w=1\|s=1,r=0) > p(w=1\|s=0,r=0)$ |
| 5 | $p(w=1\|s=1,r=1) > p(w=1\|s=1,r=0)$ |
| 6 | $p(w=0\|s=1,r=0) \approx p(w=0\|s=0,r=1)$ |
| 7 | $p(w=1\|s=1,r=0) > p(w=0\|s=1,r=0)$ |
| 8 | $p(w=1\|s=1,r=1) > p(w=0\|s=1,r=1)$ |
| 9 | $p(w=1\|s=0,r=1) > p(w=1\|s=0,r=0)$ |
| 10 | $p(w=1\|s=1,r=1) > p(w=1\|s=0,r=1)$ |

### 3.1 精度分析

为验证新算法 BN 参数学习精度是否具有优势, 用样本渐增实验、约束渐增实验和随机实验对比 ML、CVX、QMAP 和 BERD 参数学习精度. 实验设置分别如下:

**实验 4.** 精度之样本渐增实验. 1) 样本量 (控制变量) 由 1 到 35 逐渐增加; 2) 在每个样本量条件下, 约束数量 (随机变量) 在 1～9 之间随机产生, 并且约束从约束备选集合中随机抽取; 3) 在每个样本量条件下, 真实参数位置 (随机变量) 未知, 每次实验都从由所有备选约束确定的可行域中随机抽取; 4) 在每个条件 1) 下, 按条件 2) 和条件 3) 构造 50 个 BN 并抽样进行参数学习, 结果取平均值. 实验结果如图 8 所示.

**实验 5.** 精度之约束渐增实验. 1) 约束数量 (控制变量) 由 1～9 逐渐增加, 并且约束从约束备选集合中随机抽取; 2) 在每个约束数量条件下, 样本量 (随机变量) 在 1～35 之间随机产生; 3) 在每个约束数量条件下, 真实参数位置 (随机变量) 未知, 每次实验都从由所有备选约束确定的可行域中随机抽取; 4) 在每个条件 1) 之下, 按条件 2) 和条件 3) 构造 50 个 BN 并抽样进行参数学习, 结果取平均值. 实验结果如图 9 所示.

**实验 6.** 精度之随机实验. 1) 约束数量在 1～9 之间随机产生, 并且约束从约束备选集合中随机抽取; 2) 样本量在 1～35 之间随机产生; 3) 真实参数位置未知, 每次实验都从由所有备选约束确定的可行域中随机抽取; 4) 进行 100 次随机实验, 每次按条件 1)、条件 2)、条件 3) 构建一个 BN 并抽样进行参数学习, 结果取平均值. 实验结果如图 10 所示.

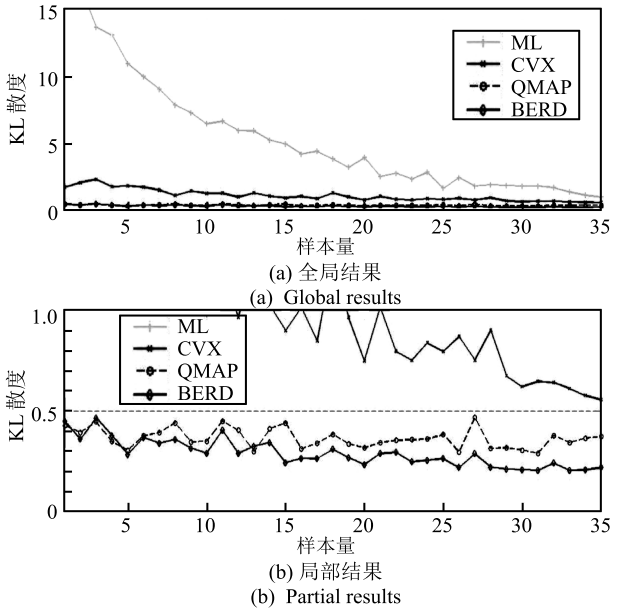



图 8 算法精度分析之样本量渐增实验

Fig. 8 Accuracy analysis for algorithms altering data size

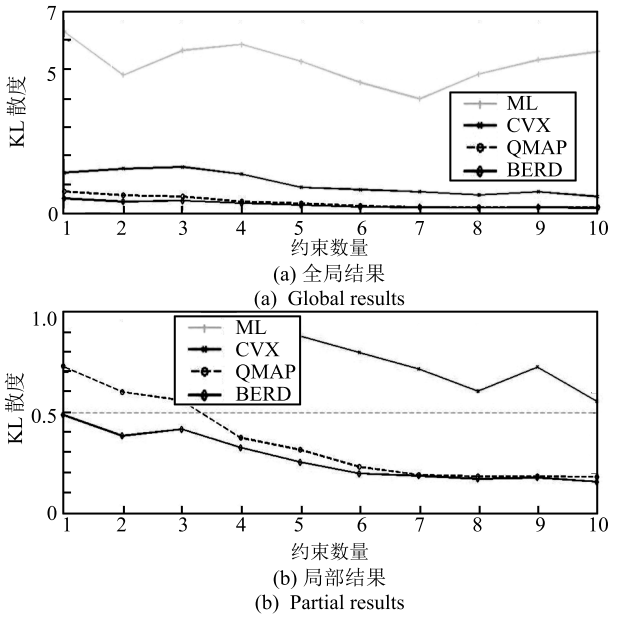
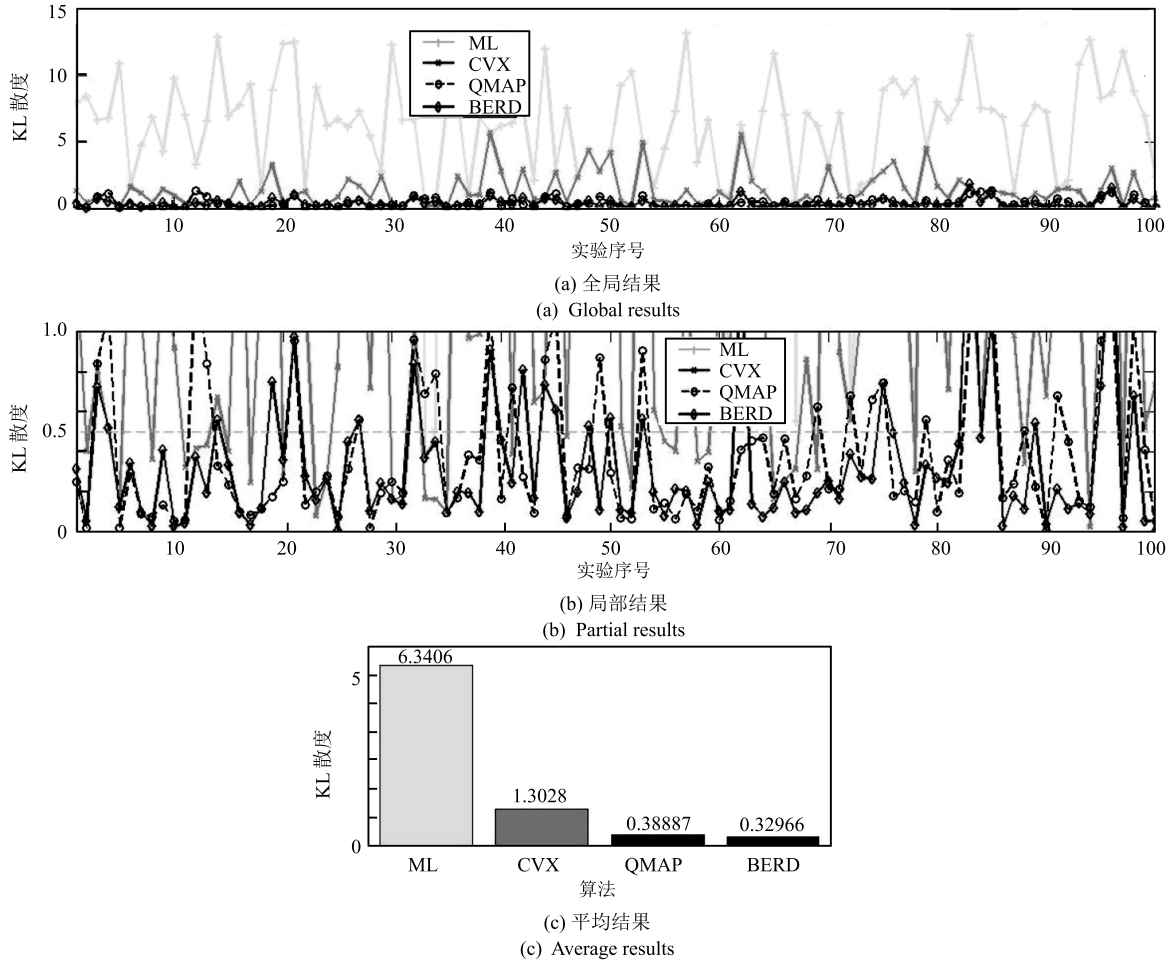



图 9 算法精度分析之约束渐增实验

Fig. 9 Accuracy analysis for algorithms altering number of constraints

图 8～10 都分有 (a) 和 (b) 两个子图, 图 (a) 为全局图, 图 (b) 为图 (a) 纵坐标 [0, 1] 之间的局部图, 后者能让读者更清楚地看到 CVX、QMAP 和 BERD 算法的精度特征; 图 10 还有一个子图 (c), 是全程实验的平均值展示.




(a) 全局结果
(a) Global results

(b) 局部结果
(b) Partial results

(c) 平均结果
(c) Average results

图 10　算法精度分析之随机实验

Fig. 10　Accuracy analysis for algorithms altering data randomly

由图 8 可知: 1) 随着样本量增加, BERD 算法的精度收敛最快且很快稳定下来, QMAP 的精度收敛速度也比较快但中间有较大波动, CVX 和 ML 的收敛速度依次变慢; 2) 随着样本量的增加, 本文算法相对 QMAP 的优势越来越明显.

由图 9 可知: 1) 随着约束数量的增加, BERD 算法精度收敛最快, QMAP 算法次之, CVX 算法再次, 而 ML 算法精度没有随约束数量而收敛 (因为 MLE 算法只用数据学习参数); 2) 当约束较少时, 本文算法相对 QMAP 的优势较大.

由图 10 可知: 在随机条件下, BERD 算法的学习结果保持了相对稳定, 平均精度在 QMAP 算法的基础上提高了 28.3 %, 这说明本文算法有较好的适用性.

因此, 本文提出的 BERD 算法不仅在精度上有较大提升, 而且随样本与约束数量的增加收敛很快. 这说明: 1) 本文算法对数据和约束的利用效率较高. 2) BERD 算法的三种新思想 — 数据信息分类; 用数据构造 QCSP 和数据再利用是合理的、有效的.

### 3.2　实用性分析

本文用联合树推理算法进行推理实验来分析算法实用性. 推理过程中, 以洒水S和下雨R两个节点的状态为证据, 在后验 BN 网络中推理草坪湿润节点W的条件概率表, 并用 KL 散度计算其与真实 BN 推理结果的差异, 差异越小说明估计参数的推理结果越好. 具体的计算公式为

$$KL_{\text{Inference}} = \sum_{w=0}^{1} p_{\text{Estimation}}(w|s,r) \ln \frac{p_{\text{Estimation}}(w|s,r)}{p_{\text{True}}(w|s,r)} \tag{25}$$

式中, $KL_{\text{Inference}}$ 为后验 BN 与真实 BN 的推理差异; $w$, $s$ 和 $r$ 分别为 $W$, $S$ 和 $R$ 节点的取值; $p_{\text{Estimation}}(w|s,r)$ 和 $p_{\text{True}}(w|s,r)$ 分别为估计参数与真实参数推理得到的节点的后验概率密度.

**实验 7.** 推理实验之一: 1) 训练样本量 (控制变

量) 在 1～35 之间逐渐增加, 每次迭代均用真实参数对应的 BN 随机产生; 2) 约束数量在 1～9 之间随机产生, 并且约束从约束备选集合中随机抽取; 3) 真实参数位置未知, 每次实验都从由所有备选约束确定的可行域中随机抽取; 4) 测试样本量为 100, 用真实参数对应的 BN 随机产生; 5) 在每个条件 1) 之下, 条件 2) 和条件 3) 构造 50 个后验 BN, 推理后用式 (25) 计算推理精度并求平均值. 实验结果如图 11 (a) 所示.

**实验 8.** 推理实验之二: 1) 约束数量 (控制变量) 由 1～9 渐增, 并且约束从约束备选集合中随机抽取; 2) 训练样本量在 1～35 之间随机产生, 每次迭代均用真实参数对应的 BN 随机产生; 3) 真实参数位置未知, 每次实验都从由所有备选约束确定的可行域中随机抽取; 4) 测试样本量为 100, 用真实参数对应的 BN 随机产生; 5) 在每个条件 1) 之下, 按条件 2) 和条件 3) 构造 50 个后验 BN, 推理后用式 (25) 计算推理精度并求平均值. 实验结果如图 11 (b) 所示.

由图 11 可知, 1) 本文算法估计参数的推理精度最高; 2) 随着样本量及约束数量的增加, 本文算法估计参数的推理精度优势加大; 3) 在约束较少时, 本文算法估计参数的推理精度优势十分明显. 证明本文算法在实用性方面拥有较大优势.

### 3.3 时间复杂度分析

已经证明, 凸规划问题可以在多项式时间内完成[20], 则可设 CVX 算法的时间复杂度为

$$T_{\rm CVX} = {\rm O}(f(n)) \tag{26}$$

其中, $f(n)$ 为 $n$ 的多项式, ${\rm O}(f(n))$ 表示多项式时间复杂度. 设 Monte Carlo 抽样的次数为 $n$, 可行域的超体积为 $V$, 参数空间的超体积为单位 1, 将抽样一次的代码看成一个整体, 则 Monte Carlo 抽样的代码执行次数为

$$f_{\rm MC}(n) = \frac{n}{V} + C$$

上式为一次多项式, 其中 $C$ 为常数. 所以, $n$ 次 Monte Carlo 抽样的时间复杂度为

$$T_{\rm MC} = {\rm O}(n)$$

而 QMAP 和本文的 BERD 算法的时间主要消耗在 Monte Carlo 抽样中, 即

$$T_{\rm QMAP} \approx T_{\rm BERD} \approx T_{\rm MC}$$

上式中, $T_{\rm QMAP}$ 和 $T_{\rm BERD}$ 分别为 QMAP 和 BERD 的时间复杂度. 故 QMAP 和 BERD 算法的时间复杂度为一次多项式时间, 即

$$\begin{aligned} T_{\rm QMAP} &= {\rm O}(n) \\ T_{\rm BERD} &= {\rm O}(n) \end{aligned} \tag{27}$$

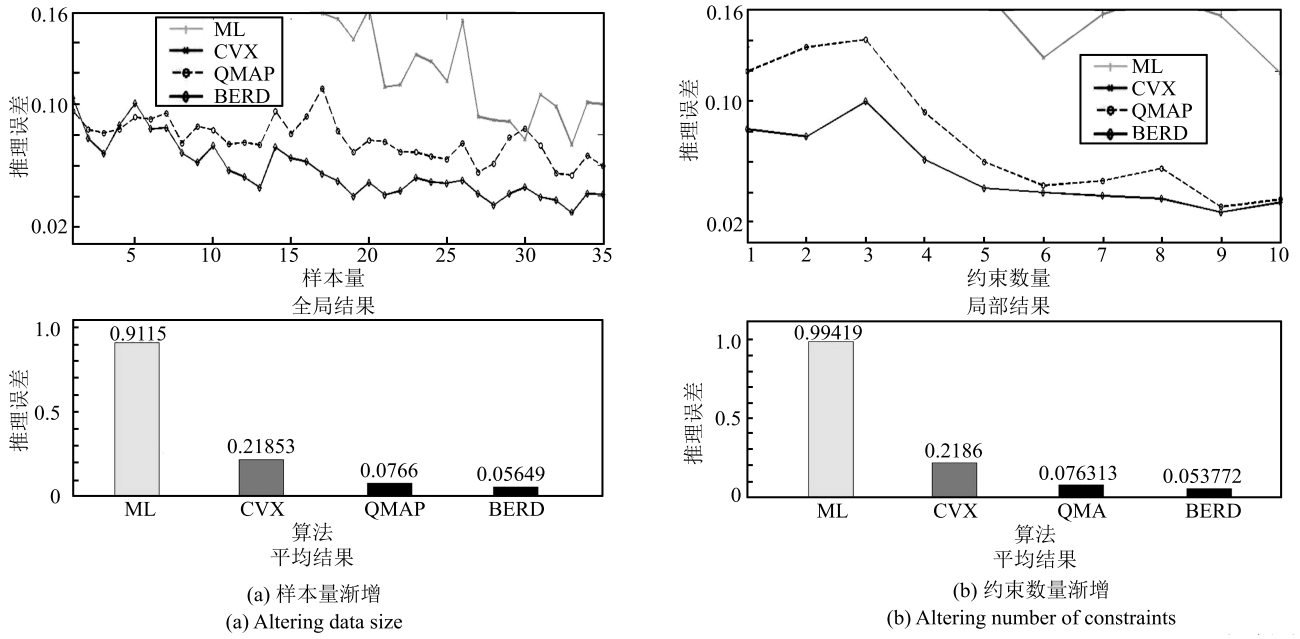



图 11 算法推理精度分析

Fig. 11 Accuracy analysis for algorithms inference

虽然 QMAP、BRED 以及 CVX 算法的时间复杂度都为多项式时间, 但前面两种算法多项式次数是一次, 故理论上时间消耗低于 CVX 算法. 但为了保证精度, Monte Carlo 抽样的次数一般较大 (本文中为 2 000), 导致有时 QMAP 和 BRED 的时间消耗会大于 CVX 算法. 本文分别通过实时性实验和时效性验证算法复杂度和效率, 实验设置如下:

**实验 9.** 实时性及时效性实验: 1) 样本量 (控制变量) 由 1～35 递增; 2) 约束数量在 1～9 之间随机产生, 并且约束从约束备选集合中随机抽取; 3) 真实参数位置未知, 每次实验都从由所有备选约束确定的可行域中随机抽取; 4) 在每个条件 1) 之下, 按条件 2) 和条件 3) 构造 BN, 对 BN 抽样 10 次进行参数学习, 记录每次的 KL 散度和时间消耗, 并对 50 次结果求平均值, 最后用式 (28) 计算时效性. 实验结果如图 12 所示.

$$E = \mathrm{e}^{\frac{1}{\overline{KL}\times\overline{T}}} \tag{28}$$

式中, $E$ 是时效性, $\overline{KL}$ 是平均 KL 散度, $\overline{T}$ 是平均时间消耗.

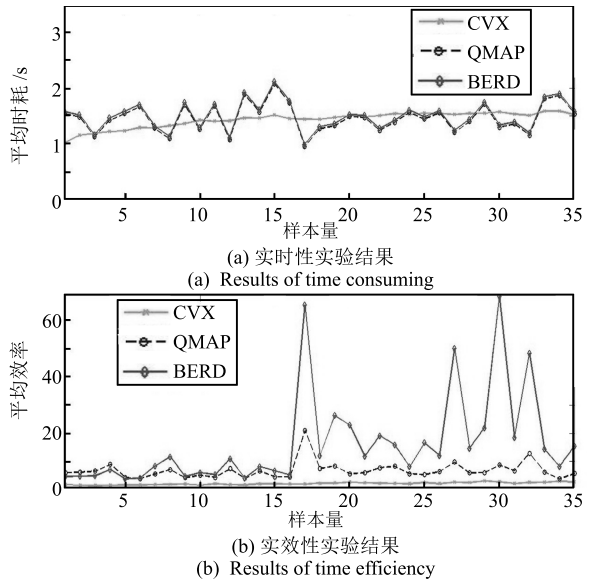



图 12　算法实时性及时效性分析

Fig. 12　Analyses of time consuming and efficiency

图 12 (a) 为实时性实验结果, 由图可知 (不考虑 ML 算法): 1) CVX 算法的时间消耗比较稳定, 随样本量增加略微增大; 2) QMAP 和 BERD 时间消耗几乎相同, 在 CVX 水平上下波动. 该实验证明本文算法为多项式时间复杂度 (因为 CVX 为多项式时间复杂度). 图 12 (b) 为时效性实验结果, 由图可知: 1) 本文算法随样本增加效率增加最明显; 2) 多数情况下本文算法的时效性最高, 最差的为 CVX 算法.

## 4　总结与讨论

已有参数学习算法认为数据和先验知识是独立的, 在参数学习过程中将二者机械结合. 但本文经深入研究后发现数据与先验知识并不独立, 并据此提出了一种新算法, 提高了凸约束条件下的 BN 参数学习精度. 现将本文成果总结如下:

1) 通过长期的实验、观察和归纳, 发现数据和先验知识并不独立, 不能将二者截然分开, 否者会浪费它们之间的约束信息. 这是本文的基础.

2) 提出了数据信息分类思想, 将样本数据携带的信息分为两类: a) 刻画先验参数的后验概率, 转换成约束后可过滤先验知识中的干扰信息; b) 刻画后验参数的概率分布, 作为 MAP 思想求解参数的基础. 这是本文的第一个创新点.

3) 基于成果 1) 和 2), 提出了用数据构造 QCSP, 这是一种新的数据利用方法的思想. 该思想从新的角度挖掘了数据的信息, 有根据地增加了约束, 过滤掉先验知识中的干扰信息, 为参数学习算法精度的提高指明了新方向. 这是本文的第二个创新点.

4) 在成果 2) 的基础上, 提出了数据再利用思想, 通过对不同数据利用方法 (这些方法所用的数据信息必须有所差异) 的有序综合使用, 在充分挖掘数据潜力的同时, 让数据和先验知识的结合更加合理, 这对探寻其他新的 BN 参数学习方法有较大的启示作用. 这是本文的第三个创新点.

5) 提出了新算法 — 凸约束条件下基于数据再利用的贝叶斯估计 (BERD). 该算法首先利用数据构造 QCSP (第一次使用数据), 过滤掉先验知识中的干扰信息, 得到有效区域; 再用 Monte Carlo 法基于均匀分布规律抽样有效区域得到有效先验参数; 接着用 Beta 分布拟合有效先验参数以充分提取先验参数信息; 最后用贝叶斯公式求取参数 (再次用到数据). 经实验证明, 新算法估计参数的精度和推理精度均明显高于已有算法, 且为一次多项式时间复杂度.

尽管本文算法在很多方面都有较好的表现, 但是它仍然存在进一步提升的空间, 具体可以总结为以下几点: 1) 当数据极少时, 精度相较 QMAP 算法没有优势, 故通过改进新算法的参数学习结果, 提高极少数据条件下的参数学习精度是一个继续研究的方向; 2) 用式 (15) 所构造的约束并不是最优的, 改进之以进一步提高新算法参数学习精度将是另一个继续研究的方向.

## References

1 Pearl J. *Probabilistic Reasoning in Intelligent Systems*. Massachusetts: Morgan Kaufmann, 1988.

2 Jin Nai-Gao, Yin Fu-Liang, Chen Zhe. Audio-visual speaker tracking based on dynamic Bayesian network. *Acta Automatica 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)

3 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)

4 Du You-Tian, Chen Feng, Xu Wen-Li. Approach to human activity multi-scale analysis and recognition based on multilayer dynamic Bayesian network. *Acta Automatica Sinica*, 2009, **35**(3): 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, **35**(3): 225−232)

5 Wan Jiu-Qing, Liu Qing-Yun. Data association in visual sensor networks based on high-order spatio-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)

6 Tamda Y, Imoto S, Araki H, Nagasaki M, Printe C, Charnock-Jones D S, Miyano S. Estimating genome-wide gene networks using nonparametric Bayesian network models on massively parallel computers. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2011, **8**(3): 683−697

7 Chang R, Shoemaker R, Wang W. A novel knowledge-driven systems biology approach for phenotype prediction upon genetic intervention. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2011, **8**(5): 1170−1182

8 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

9 Jin Can-Can, Zuo Hong-Fu, Zhang Ying, Bai Fang. Research on risk evaluation for airlines based on BBN. *Acta Aeronautica ET Astronautica Sinica*, 2013, **34**(3): 588−596
(金灿灿, 左洪福, 张营, 白芳. 基于 BBN 的航空公司风险评估技术研究. 航空学报, 2013, **34**(3): 588−596)

10 Yin Xiao-Wei, Qian Wen-Xue, Xie Li-Yang. A method for system reliabiltiy assessment based on Bayesian networks. *Acta Aeronautica ET Astronautica Sinica*, 2008, **29**(6): 1482−1489
(尹晓伟, 钱文学, 谢里阳. 系统可靠性的贝叶斯网络评估方法. 航空学报, 2008, **29**(6): 1482−1489)

11 Druzdzel M J. Probabilistic Reasoning in Decision Support Systems: From Computation to Common Sense [Ph. D. dissertation], Carnegie Mellon University, Pennsylvania, 1993.

12 Druzdzel M, van der Gaag L C. Building probabilistic networks: "where do the numbers come from?" guest editors′ introduction. *IEEE Transactions on Knowledge and Data Engineering*, 2000, **12**(4): 481−486

13 Sijbers J, den Dekker A J, Scheunders P, Van Dyck D. Maximum-likelihood estimation of Rician distribution parameters. *IEEE Transactions on Medical Imaging*, 1998, **17**(3): 357−361

14 Helsper E M, van der Gaag L C, Groenendaal 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. Whittlebury Hall, UK: Springer, 2004. 280−292

15 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
(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)

16 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, 2000. 644−652

17 Altendorf E E, Restificar A C, Dietterich T G. 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

18 Liao W H, Ji Q. Learning Bayesian network parameters under incomplete data with domain knowledge. *Pattern Recognition*, 2009, **42**(11): 3046−3056

19 de Campos C P, Tong Y, Ji Q. Constrained maximum likelihood learning of Bayesian networks for facial action recognition. In: Proceedings of the 10th European Conference on Computer Vision. Marseille, France: Springer, 2008. 168−181

20 de Campos C P, Ji Q. Improving Bayesian network parameter learning using constraints. In: Proceedings of the 19th International Conference on Pattern Recognition. Tampa, Florida: IEEE: 2008. 1−4

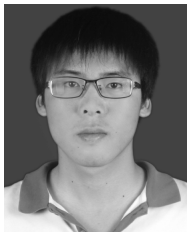
21 Feelders A, van der Gaag L C. Learning Bayesian network parameters under order constraints. *International Journal of Approximate Reasoning*, 2006, **42**(1−2): 37−53

22 Chang R, Wang W. Novel algorithm for Bayesian network parameter learning with informative prior constraints. In: Proceedings of the 2010 International Joint Conference on Neural Networks. Barcelona: IEEE, 2010. 1−8

23 Zhang Lian-Wen, Guo Hai-Peng. *Introduction to Bayesian Networks*. Beijing: Science Press, 2006.
(张连文, 郭海鹏. 贝叶斯网引论. 北京: 科学出版社, 2006.)

24 Zhou Y, Fenton N, Neil M. Bayesian network approach to multinomial parameter learning using data and expert judgments. *International Journal of Approximate Reasoning*, 2014, **55**(5): 1252−1268

25 Murphy K. A brief introduction to graphical models and Bayesian networks [Online], available: http://www.cs.ubc.ca/murphyk/Bayes/bnintro.html, July 1, 2015



**杨 宇** 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: youngiv@126.com
(**YANG Yu** 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 sets.)

**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.)

**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的贝叶斯网络参数学习.
E-mail: shangsl@gmail.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 sets.)