

基于因果影响独立模型的贝叶斯网络参数学习

肖蒙^{1,2}, 张友鹏¹

(1. 兰州交通大学 自动化与电气工程学院, 兰州 730070; 2. 五邑大学 轨道交通学院, 广东 江门 529020)

摘要: 基于因果影响独立模型及其中形成的特定上下文独立关系, 提出一种适于样本学习的贝叶斯网络参数学习算法. 该算法在对局部概率模型降维分解的基础上, 通过单父节点条件下的子节点概率分布来合成局部结构的条件概率分布, 参数定义复杂度较低且能较好地处理稀疏结构样本集. 实验结果表明, 该算法与标准最大似然估计算法相比, 能充分利用样本信息, 具有较好的学习精度.

关键词: 贝叶斯网络; 因果影响独立; 样本集; 参数学习

中图分类号: TP181

文献标志码: A

Parameters learning of Bayesian networks based on independence of causal influence model

XIAO Meng^{1,2}, ZHANG You-peng¹

(1. School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China; 2. School of Railway Tracks and Transportation, Wuyi University, Jiangmen 529020, China. Correspondent: XIAO Meng, E-mail: xiaom@mail.lzjtu.cn)

Abstract: Based on the independence of the causal influence model and the context-specific independent relations arising in the model, a parameters learning algorithm of Bayesian networks suiting for sample learning is proposed. Through decomposing and dimension-reducing the local probability model, the algorithm can synthesize the conditional probability distribution of the local structure with the probability distribution of the child nodes under the single parent node. The algorithm has low parameter-defining complexity and can better deal with the sparse structure sample set. Compared with the standard maximum likelihood estimation algorithm, the experimental results show that the proposed algorithm can fully extract the information from sample data and has higher learning accuracy.

Keywords: Bayesian networks; independence of causal influence; sample set; parameters learning

0 引言

贝叶斯网络^[1]是一种用于不确定性推理的概率图模型. 利用条件独立关系, 贝叶斯网络能够将高维联合分布分解成低维局部条件概率分布的乘积, 从而降低了系统整体概率模型的复杂度^[2]. 当节点代表离散随机变量时, 局部条件概率分布以在父节点所有可能状态组合下的子节点条件概率表(CPT)的形式表示. 然而, 贝叶斯网络对局部概率模型中子节点与父节点之间的依赖关系没有做出进一步的说明或约束^[3]. 这意味着当网络中节点众多且连接稠密时, CPT的参数定义复杂度将随父节点数量呈指数级增加, 会使局部概率模型的参数量化和整个网络的推理计算

变得困难. 同时, 相对于呈指数增长的参数规模, 由于样本量往往太少, 亦会增大参数的估计误差^[4]. 在局部结构中是否也存在着某种独立关系, 并可依此降维分解局部条件概率分布, 已成为进一步降低局部结构乃至系统整体复杂度的有效途径.

因果关系是一种常用的思维方式, 人们用此来分析问题和建立模型. 在某种情形下, 当局部结构中父节点指向子节点的有向弧表示因果影响关系, 且各个父节点施加的因果影响之间相互独立时, 称这种局部结构具有因果影响独立(ICI)^[3,5]性质. 利用ICI性质, 可建立该局部概率模型的因子分解形式, 即可通过对各个父节点对子节点的因果影响来刻画父节点整体

收稿日期: 2014-04-01; 修回日期: 2014-07-05.

基金项目: 铁道部科技研究开发计划重点课题(2012X003-B); 甘肃省自然科学基金项目(1112RJZA040).

作者简介: 肖蒙(1974—), 男, 副教授, 博士生, 从事交通信息工程、系统可靠性的研究; 张友鹏(1965—), 男, 教授, 博士生导师, 从事交通运输安全保障与防护技术等研究.

施加的因果影响. 这种基于因果关系的独立假设, 使得CPT定义所需的独立参数个数降为父节点数量的线性形式, 这便于从专家知识获取CPT参数, 能够降低概率模型参数化的复杂度^[6]. 同时, 由于因子分解形式符合贝叶斯网络的建模思想, 局部结构的进一步分解有助于提高推理计算效率和精度^[3,7].

ICI模型已在实践中得到了广泛应用. QMR-DT BN20^[8-9]和CPCS^[10-11]是两个已用于医疗诊断的贝叶斯网络模型, 分别使用ICI模型的Noisy-Or形式和其多值扩展形式Noisy-Max来描述疾病与症状之间的关系. 文献[12-13]讨论了ICI模型参数学习, 其学习方法以“只有单一原因出现而其他原因均缺失”为条件, 这便于专家运用知识的独立性来确定参数, 但对于通过样本学习获得参数显得条件苛刻.

本文针对贝叶斯网络参数学习问题, 以局部结构的CPT为对象, 利用ICI模型得到局部结构条件概率分布的因子分解形式, 在此基础上, 提出一种适于样本条件下的CPT参数学习方法. 该参数学习方法能充分利用样本信息, 通过单父节点条件下的概率分布合成CPT参数, 并能够提高参数学习的精度.

1 因果影响独立的性质及ICI模型

1.1 因果影响独立的性质分析

因果影响独立蕴含了一定的独立关系、因果关系和合成关系. 独立关系表现为, 任一原因无需与其他原因协作便可对结果施加影响, 即其单独产生的影响结果独立于其他原因及其他原因的影响结果. 因果关系则表现为原因是结果的必要正因关系, 即当原因缺失时, 其影响结果必然缺失. 这种必要正因关系既存在于单一原因与其单独影响结果之间, 也存在于众原因与同一结果之间. 因果影响独立的实质是一种基于因果关系的条件独立, 这也符合贝叶斯网络的局部马尔科夫性质. 因果影响独立在将众原因对结果的整体影响分解为单个因果影响的同时, 用诸如逻辑或、逻辑与、最大值、最小值等简单形式来合成各个单独因果影响对同一结果的整体作用. 既然整体作用源自于各单独影响, 因此需要在ICI模型中对单独因果影响进行量化, 并以概率形式予以表示. 根据以上分析, 可做如下定义.

定义1(因果影响独立) 以离散随机变量集 $C = \{C_1, C_2, \dots, C_n\}$ 代表众原因, 离散随机变量 E 代表同一结果, 离散随机变量集 $M = \{M_1, M_2, \dots, M_n\}$ 代表各原因的单独影响结果, 且 M 与 E 的状态空间相同(记为 $\Omega_M = \Omega_E$), 则称 C 对 E 是因果影响独立的, 如果下面条件成立:

1) 任意 M_i 仅依赖于 C_i , 即在 C_i 条件下, M_i 独

立于所有其他的 C_j 和 $M_j (j \neq i)$;

2) 记 C, M, E 的缺失状态分别为 $\neg C = \{\neg C_1, \neg C_2, \dots, \neg C_n\}$, $\neg M = \{\neg M_1, \neg M_2, \dots, \neg M_n\}$, $\neg E$, 在 C 与 E, C_i 与 M_i 之间存在必要正因关系, 即 $P(\neg M_i | \neg C_i) = 1$ 和 $P(\neg E | \neg C_1, \neg C_2, \dots, \neg C_n) = 1$;

3) 在 Ω_E 上存在一个完全逻辑函数 $f(M) = f(M_1, M_2, \dots, M_n)$, 且当 $f(M) = E$ 时, $P(E|M) = 1$, 反之 $P(E|M) = 0$.

1.2 ICI模型及其扩展形式

根据因果影响独立定义, 可以建立其贝叶斯网络模型, 如图1所示.

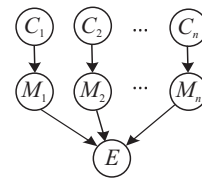


图1 因果影响独立模型

由图1所含条件独立关系, 可以得到条件概率分布 $P(E|C)$ 的因子分解形式.

定理1(ICI模型的CPT定义^[13]) 设原因变量 C_1, C_2, \dots, C_n 独立地影响结果变量 E , 则对于任意 $e \in \Omega_E$, 有

$$P(E = e | C) = \sum_{M|f(M)=e} \prod_{i=1}^n P(M_i | C_i). \quad (1)$$

其中 $P(M_i | C_i)$ 可以理解为在 $f(M) = e$ 约束条件下, 父节点 C_i 对 $E = e$ 的影响贡献概率.

定理1说明, 在ICI模型中, 子节点对多父节点的条件概率分布可由各单父节点的影响贡献概率的乘积形式得到. 若模型中各节点均代表二值变量时, 式(1)左端的概率分布所含独立参数的个数是 2^n , 而右端的是 $2n$. 因此, 因果影响独立大大减少了完全定义CPT所需参数的数目, 从指数级降为线性级. 当父节点数较多且呈现多态时, 这种降幅是十分巨大的.

在对实际问题建模中, 如疾病诊断、故障诊断等, 经常很难将结果的全部可能致因作为可观测变量列入模型, 而未列入模型的未知原因会对已建模型的参数产生影响. 这种影响主要反映在各已知原因与结果之间的必要正因关系上, 即在专家知识提取或样本数据学习中出现 $P(\neg E | \neg C_1, \neg C_2, \dots, \neg C_n) < 1$ 的情况. 这种情况的出现并不与定义1矛盾, 因为此时只有当未知原因和各已知原因全部缺失时, 结果才会必然缺失. 记未知原因为 C_L , 其单独影响结果为 M_L , 则必要正因关系应为 $P(\neg E | \neg C_1, \neg C_2, \dots, \neg C_n, \neg C_L) = 1$. 若考虑未列入原因的影响, 便可得到ICI模型的扩展形式, 即Leaky ICI模型及其CPT定义.

推论 1 (Leaky ICI 模型的 CPT 定义^[13]) 设原因变量 $C_1, C_2, \dots, C_n, C_L$ 独立地影响结果变量 E , 其中 C_L 为未知原因, 则对于任意 $e \in \Omega_E$, 有

$$P(E = e|C) = \sum_{\substack{M, M_L \\ f(M, M_L) \leq e}} P(M_L) \prod_{i=1}^n P(M_i|C_i). \quad (2)$$

ICI 模型是 Leaky ICI 模型的一种特例, 当 $P(\neg C_L) = 1$ 时, 即未知原因始终缺失时, 式(2)就等于式(1).

2 CPT 参数学习的数学模型

由式(1)和(2)可见, 通过模型参数 $P(M_i|C_i)$ 可以间接地完全定义 CPT, 而且 $P(M_i|C_i)$ 定义所需的独立参数个数远少于 CPT 定义所需的独立参数个数.

在 ICI 模型中 $P(M_i = e|C_i) = P(E = e|\neg C_1, \dots, \neg C_{i-1}, C_i, \neg C_{i+1}, \dots, \neg C_n)$ ^[13], 即单一原因的独立影响程度可以按照所有其他原因缺失时来度量, 这将便于专家利用知识的独立性来确定 $P(M_i|C_i)$ 分布. 但若通过样本学习来获取 ICI 模型参数, 则需要符合其他原因均处于缺失状态条件的足够样本, 以此来选择样本显得苛刻而不切实际. 由于在 ICI 模型和 Leaky ICI 模型中, $P(E|C)$ 可由各父节点的影响概率贡献 $P(M_i|C_i)$ 按 $f(M)$ 形式来合成, 若 $P(M_i|C_i)$ 与 $P(E|C_i)$ 之间存在某种关系, 而 $P(E|C_i)$ 便于由样本估计得到, 则可为人们利用因果影响独立的优势, 基于样本直接学习 CPT 参数提供一种思路.

Noisy-OR 模型和 Noisy-MAX 模型是 ICI 模型的常见具体形式, 表现为合成函数 $f(M)$ 是 OR/MAX 运算. Noisy-OR 模型的适用条件之一是要求各节点的状态空间只能有两个, 即模型涉及的离散随机变量为二态变量. 在对复杂问题建模时, 系统内部事件或部件的可能状态更多地呈现出多态. 作为 Noisy-OR 模型在多态状态空间上的推广形式, Noisy-MAX 模型的适用场景更为广泛. 不失一般性, 下面以 Leaky Noisy-MAX 模型为例, 讨论在单一父节点条件下与完全父节点集条件下, 子节点的条件概率分布之间的关系, 并建立本文 CPT 参数学习的数学模型.

若对 Leaky Noisy-MAX 模型中子节点的所有可能状态进行排序, 且其缺失状态为最小值^[12], 则可以得到完全父节点集条件下的子节点概率分布的分解形式.

推论 2 在 Leaky Noisy-MAX 模型中, 原因变量 $C_1, C_2, \dots, C_n, C_L$ 独立地影响结果变量 E , 其中 C_L 为未知原因, 则对于任意 $e \in \Omega_E$, 有

$$P(E \leq e|C) = P(M_L \leq e) \prod_{i=1}^n P(M_i \leq e|C_i). \quad (3)$$

证明 若 $P(E \leq e|M, M_L) = 1$, 当且仅当任意

$M_i \leq e$ 和 $M_L \leq e$, 则由式(2)得

$$\begin{aligned} P(E \leq e|C) &= \sum_{\substack{M, M_L \\ \text{Max}(M, M_L) \leq e}} P(M_L) \prod_{i=1}^n P(M_i|C_i) = \\ &= \sum_{M_1 \leq e} \dots \sum_{M_n \leq e} \sum_{M_L \leq e} P(M_L) \prod_{i=1}^n P(M_i|C_i) = \\ &= P(M_L \leq e) \prod_{i=1}^n P(M_i \leq e|C_i). \end{aligned}$$

推论得证. \square

$P(E \leq e|C)$ 称为累加条件概率分布. CPT 参数可由下式进一步得到:

$$P(E = e|C) = \begin{cases} P(E \leq e|C) - P(E \leq e-1|C), & e \neq \neg E; \\ P(E = e|C), & e = \neg E. \end{cases} \quad (4)$$

命题 1 在 Leaky Noisy-MAX 模型中, 在结果变量 $E \leq e$ 条件下, 原因变量 C_1, C_2, \dots, C_n 之间相互独立.

证明 根据式(2), 有

$$\begin{aligned} P(C|E \leq e) &= \frac{\sum_{C, M} P(C, M, E \leq e)}{\sum_{C, M} P(C, M, E \leq e)} = \\ &= \frac{\sum_{C, M} P(M_L) \prod_{i=1}^n P(M_i|C_i) P(C_i) P(E \leq e|M, M_L)}{\sum_{C, M} P(M_L) \prod_{i=1}^n P(M_i|C_i) P(C_i) P(E \leq e|M, M_L)} = \\ &= \frac{P(M_L \leq e) \prod_{i=1}^n P(M_i \leq e|C_i) P(C_i)}{\sum_C P(M_L \leq e) \prod_{i=1}^n P(M_i \leq e|C_i) P(C_i)} = \\ &= \frac{\prod_{i=1}^n P(M_i \leq e, C_i)}{\sum_C \prod_{i=1}^n P(M_i \leq e, C_i)} = \prod_{i=1}^n P(C_i|M_i \leq e). \quad (5) \end{aligned}$$

而对于任意 $C_i \in C$ 及其对应的 $M_i \in M$, 设 $C' = C \setminus C_i$, $M' = M \setminus M_i$, 有

$$\begin{aligned} P(C_i|E \leq e) &= \frac{\sum_{C', M} P(C', C_i, M, E \leq e)}{\sum_{C', C_i, M} P(C', C_i, M, E \leq e)} = \\ &= \frac{\sum_{C'} P(M_L \leq e) P(M_i \leq e|C_i) P(C_i) P(M' \leq e|C') P(C')}{\sum_{C'; C_i} P(M_L \leq e) P(M_i \leq e|C_i) P(C_i) P(M' \leq e|C') P(C')} = \end{aligned}$$

$$\frac{P(M_L \leq e)P(M_i \leq e|C_i)P(C_i) \sum_{C'} P(M' \leq e|C')P(C')}{P(M_L \leq e) \sum_{C_i} P(M_i \leq e|C_i)P(C_i) \sum_{C'} P(M' \leq e|C')P(C')} = \frac{P(M_i \leq e, C_i)}{\sum_{C_i} P(M_i \leq e, C_i)} = P(C_i|M_i \leq e). \quad (6)$$

由式(5)和(6),得

$$P(C|E \leq e) = \prod_{i=1}^n P(C_i|E \leq e). \quad (7)$$

命题成立. \square

命题 1 实际是一种在因果独立关系中特定环境下 ($E \leq e$) 存在的条件独立, 常被称为特定上下文独立^[14]. 对于其他合成形式的 ICI 模型, 也有相似的特定上下文独立关系, 如在 Noisy-OR 形式中, 当子节点处于缺失状态时, 父节点之间亦相互独立.

引理 1 在 Leaky Noisy-MAX 模型中, 对于任意 $C_i \in \mathbf{C}, e \in \Omega_E$, 有

$$P(E \leq e|C_i) = P(M_L \leq e)P(M_i \leq e|C_i) \prod_{j=1|j \neq i}^n P(M_j \leq e), \quad (8)$$

$$P(E \leq e) = P(M_L \leq e) \prod_{i=1}^n P(M_i \leq e). \quad (9)$$

证明 由式(3)及命题 1, 有

$$\begin{aligned} P(E \leq e|C_i) &= \sum_{C'} P(E \leq e|C', C_i)P(C'|C_i) = \\ &= \sum_{C'} P(M_L \leq e)P(M_i \leq e|C_i) \times \\ &= \prod_{j=1|j \neq i}^n P(M_j \leq e|C_j)P(C_j) = \\ &= P(M_L \leq e)P(M_i \leq e|C_i) \prod_{j=1|j \neq i}^n P(M_j \leq e). \end{aligned}$$

在结果变量 $E \leq e$ 条件下, 原因变量 C_1, C_2, \dots, C_n 之间相互独立, 因此对应的 M_1, M_2, \dots, M_n 之间亦相互独立, 有

$$\begin{aligned} P(E \leq e) &= \sum_{M, M_L} P(E \leq e|M, M_L)P(M, M_L) = \\ &= \sum_{\substack{M, M_L \\ \max(M, M_L) \leq e}} P(M_L) \prod_{i=1}^n P(M_i) = \\ &= P(M_L \leq e) \prod_{i=1}^n P(M_i \leq e). \quad \square \end{aligned}$$

推论 3 在 Leaky Noisy-MAX 模型中, 对于任

意 $e \in \Omega_E$, 有

$$P(E \leq e|C) = \frac{\prod_{i=1}^n P(E \leq e|C_i)}{[P(E \leq e)]^{n-1}}. \quad (10)$$

证明 由式(8)可得

$$\begin{aligned} \prod_{i=1}^n P(E \leq e|C_i) &= [P(M_L \leq e)]^n \left[\prod_{j=1}^n P(M_j \leq e) \right]^{n-1} \prod_{i=1}^n P(M_i \leq e|C_i). \end{aligned} \quad (11)$$

由式(9)可得

$$\begin{aligned} [P(E \leq e)]^{n-1} &= [P(M_L \leq e)]^{n-1} \left[\prod_{i=1}^n P(M_i \leq e) \right]^{n-1}. \end{aligned} \quad (12)$$

将式(11)除以(12), 并由(3)可得

$$\begin{aligned} \frac{\prod_{i=1}^n P(E \leq e|C_i)}{[P(E \leq e)]^{n-1}} &= P(M_L \leq e) \prod_{i=1}^n P(M_i \leq e|C_i) = P(E \leq e|C). \quad \square \end{aligned}$$

式(10)就是基于 Leaky Noisy-MAX 形式的 CPT 参数学习模型. 式(10)表明, 若贝叶斯网络局部结构中存在因果影响独立关系, 则可利用样本直接得到单父节点条件下的子节点概率分布 $P(E|C_i)$, 进而得到父节点集条件下的子节点概率分布 $P(E \leq e|C)$, 即 CPT 参数. 由于在其他合成形式的 ICI 模型中也存在相似的特定上下文独立关系, 本文算法也能扩展至其他合成形式的 ICI 模型中.

若设局部结构中各节点的可能状态均为 m 个, 则式(10)左端的概率分布所含独立参数的个数是 $(m-1)m^n$, 而右端的个数是 $(m-1)(mn-1)$, 显然本文算法能够有效降低 CPT 参数定义的复杂度. 这种所需独立参数的减少, 也使人们能够从相同规模样本中获取更多的统计信息, 有助于提高参数学习的精度.

3 算法描述

设 D 是一组包含 m 条完整样本且符合独立同分布假设的样本集. 利用 Leaky Noisy-MAX 形式的 ICI 模型, 学习 CPT 参数的算法步骤如下.

Step 1: 用最大似然估计方法^[2], 从样本集 D 中得到各单父节点条件下子节点的概率分布 $P(E = e|C_i = j) = N_{ije}/N_{ij}$. 和子节点的边缘分布 $P(E = e) = N_{..e}/m$, 其中 $N_{ije}, N_{ij}, N_{..e}$ 分别表示样本集 D 中符合 $C_i = j$ 和 $E = e, C_i = j, E = e$ 的样本个数.

Step 2: 得到各单父节点条件下子节点的累加条

件概率分布为

$$P(E \leq e|C_i = j) = \sum_{k=1}^e P(E = k|C_i = j),$$

以及子节点的累加边缘概率分布为

$$P(E \leq e) = \sum_{k=1}^e P(E = k).$$

Step 3: 通过式 (10), 利用累加概率分布 $P(E \leq e|C_i)$ 和 $P(E \leq e)$, 得到完全父节点集条件下的累加概率分布 $P(E \leq e|C)$.

Step 4: 通过式 (4) 对累加概率分布 $P(E \leq e|C)$ 依序递减后, 可得到 $P(E = e|C)$, 即 CPT 参数.

4 算例分析

4.1 仿真模型

本文以 Hepar II 模型中的 TB (总胆红素) 节点及其全部父节点的局部结构为仿真实验模型. Hepar II 模型 (<http://genie.sis.pitt.edu/>) 是以大量实际病例信息和专家经验知识为建模基础, 用于肝脏疾病诊断的贝叶斯网络模型. 文献 [15] 指出, 该模型在量化各节点的 CPT 参数中未使用 Noisy-OR/MAX 形式, 这能够保证本文算法仿真实验的有效性.

在该局部结构中, 子节点 TB 代表检测指标值, 父节点为: FH (功能性高胆红素血)、PBC (原发性胆汁性肝硬化)、CiH (肝硬化)、GS (胆结石)、CH (慢性肝炎), 分别代表 5 种疾病. 该 5 种疾病均能对同一检测指标 TB 产生独立影响. 各节点变量的状态空间为 $FH = \{\text{present, absent}\}$, $PBC = \{\text{present, absent}\}$, $CiH = \{\text{decompensate, compensate, absent}\}$, $CH = \{\text{active, persistent, absent}\}$, $GS = \{\text{present, absent}\}$, $TB = \{a1_0, a6_2, a19_7, a88_{20}\}$. 因该局部结构所含父节点较多, 且部分节点呈现多态, 故其 CPT 规模较大, 共包含 216 个独立参数.

4.2 实验方法

本文以 Hepar II 模型中该局部结构的已建 CPT 作为标准 CPT, 在 Matlab 中使用贝叶斯网络工具箱 BNT 建立该局部结构的贝叶斯网络模型. 实验中, 根据标准 CPT, 以逻辑抽样法随机生成不同规模完整样本集后, 利用本文算法进行样本学习以获得 CPT 参数. 为评价本文算法学习精度和运行效率, 将标准最大似然估计 (MLE) 算法作为对比, 以学习得到的 CPT 与标准 CPT 之间的距离为学习精度衡量, 以算法运行所需机时为运行效率衡量. 其中, 距离衡量以父节点集的联合概率分布为权值, 取各父节点状态组合下 CPT 参数之间的欧氏距离和 KL 距离^[15-16]之和, 即按下式计算 CPT 之间的距离:

$$D_{\text{Eucl}}(P_A(E|C), P_B(\hat{E}|C)) =$$

$$\sum_{i=1}^n P(C_i) \sqrt{\sum_{j=1}^m (P_A(E_j|C_i) - P_B(\hat{E}_j|C_i))^2}, \quad (13)$$

$$D_{KL}(P_A(E|C), P_B(\hat{E}|C)) =$$

$$\sum_{i=1}^n P(C_i) \sum_{j=1}^m P_A(E_j|C_i) \log \frac{P_A(E_j|C_i)}{P_B(\hat{E}_j|C_i)}. \quad (14)$$

其中: n 为父节点状态组合数, m 为子节点状态数. 为使式 (14) 有效, 当学习得到的某一 CPT 参数等于零时, 用一个极小常数代替 (10^{-8}). 为避免随机干扰, 实验中对不同规模的样本集分别生成 10 组, 以该 10 组样本集学习结果的平均值作为实验观测值.

4.3 对比分析

对于样本组数分别为 100, 200, \dots , 2000 的不同规模完整样本集, 用本文算法和 MLE 算法学习后, 学习精度和算法运行效率实验结果的比较情况分别如图 2、图 3 所示.

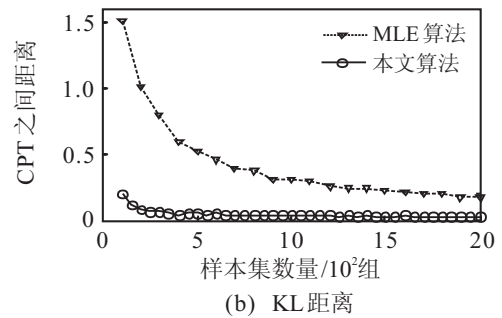
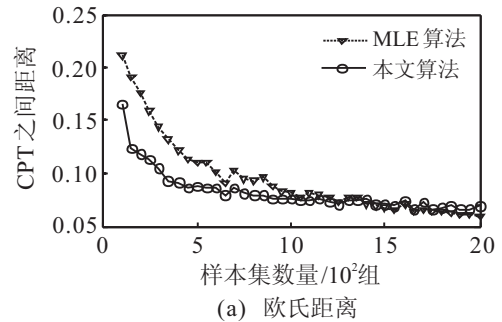


图 2 两种算法的参数学习精度比较

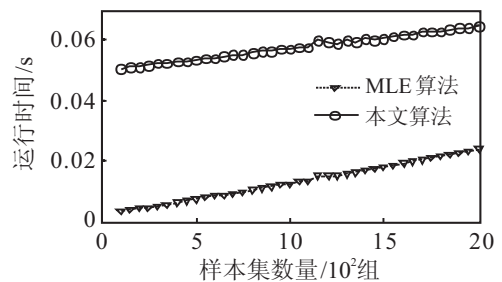


图 3 两种算法的运行效率比较

图 2 表明, 本文算法的参数学习精度整体优于 MLE 算法. 当样本集规模较小时, 本文算法的学习精度优势明显; 当样本集规模增大时, 两种算法的学习精度趋于接近.

图3表明,本文算法的运行效率低于MLE算法,运行时间最大差值为0.048 s.这是因为本文算法除了以最大似然法估计各 $P(E|C_i)$ 和 $P(E)$ 参数之外,还需进行各概率分布的累加、合成、递减计算过程,从而使得本文算法的计算复杂度高于MLE算法.

本文算法和MLE算法都用最大似然法来估计所需独立参数值.在该仿真模型中,本文算法所需独立参数个数为52个,而CPT定义所需独立参数的个数为216个.根据估计值的大样本性质,当样本量趋于无穷大时,最大似然估计值必然趋近于参数的原生分

布^[2].因此,当样本规模一定时,具有越多样本信息的统计量,其估计值就越逼近真实值.由于本文选用的仿真模型中父节点数较多且部分节点呈现多态,当样本集为稀疏结构时,由于某些状态组合没有统计信息,MLE算法学习得到的部分参数为0,这会使后续推理过程出现问题.例如,对应父节点集某一状态组合的标准CPT参数为 $P(TB|FH = \text{absent}, PBC = \text{absent}, CiH = \text{absent}, CH = \text{persistent}, GS = \text{absent}) = (0.9375, 0.0568, 0.0028, 0.0028)$.对于不同规模完整样本集,本文算法与MLE算法所得该CPT参数如表1所示.

表1 不同规模完整样本集下学习结果对比

样本数	本文算法				MLE算法			
	a_{10}	a_{62}	a_{197}	a_{8820}	a_{10}	a_{62}	a_{197}	a_{8820}
100	0.8713	0.0601	0.0686	0	1	0	0	0
200	0.8024	0.1125	0.0732	0.0119	1	0	0	0
300	0.8889	0.0894	0.0096	0.0120	1	0	0	0
400	0.8143	0.1296	0.0387	0.0175	0.8750	0.1250	0	0
500	0.8435	0.0952	0.0338	0.0275	0.9091	0.0909	0	0
600	0.8368	0.1366	0.0205	0.0060	0.8000	0.2000	0	0
700	0.8025	0.0859	0.0897	0.0219	0.9412	0.0588	0	0
800	0.7621	0.2203	0.0131	0.0046	0.8696	0.1304	0	0
900	0.8419	0.1200	0.0070	0.0312	0.9524	0.0476	0	0
1000	0.6885	0.2382	0.0669	0.0064	0.8947	0.1053	0	0
1100	0.8975	0.0528	0.0436	0.0061	0.9714	0.0286	0	0
1200	0.8377	0.1286	0.0099	0.0238	0.9000	0.1000	0	0
1300	0.8435	0.1133	0.0228	0.0204	0.8846	0.0769	0.0385	0
1400	0.9271	0.0272	0.0293	0.0165	1	0	0	0
1500	0.9678	0.0171	0.0013	0.0139	1	0	0	0
1600	0.8985	0.0765	0.0243	0.0007	0.9111	0.0889	0	0
1700	0.8474	0.1131	0.0248	0.0147	0.9574	0.0426	0	0
1800	0.9064	0.0619	0.0174	0.0144	0.9348	0.0435	0	0.0217
1900	0.7770	0.0979	0.0333	0.0919	0.8974	0.1026	0	0
2000	0.9126	0.0780	0.0061	0.0032	0.9767	0.0233	0	0

表1显示,MLE算法对于子节点TB处于 a_{197} 和 a_{8820} 状态的参数学习结果普遍为0,这是由于样本集中没有符合该父节点集与子节点状态组合的样本,而本文算法能得到相应参数的有效值且逐步逼近真实值.因KL距离对极小值敏感,故在图2(b)中,本文算法的CPT距离明显优于MLE算法.为进一步验证原因,选取各包含1000组和10000组样本的两个样本集,以各独立参数估计所利用的样本个数所占样本集比例(样本利用率)为观测值,对不同样本利用比例(坐标横轴)对应的独立参数个数所占参数总数比例(坐标纵轴)进行统计分析,以比较两种算法对样本的利用情况.比较结果如图4所示.

当样本集规模为1000组样本时,相比CPT定义所需的216个独立参数而言,样本信息显得不足.图4(a)表明,在MLE算法中,样本利用率为0(即样本集中没有符合相应节点状态组合的样本)的参数占72%,其余参数的样本利用率均不足10%;而在本文

算法中,样本利用率为0的参数只占4%,样本利用率大于10%的参数达23%.在图4(b)中,样本集规模增加为10000组样本,但MLE算法中样本利用率为0的参数仍有43%,样本利用率大于10%的参数只有3%;而本文算法中样本利用率为0的参数没有,样本利用率大于10%的参数有22%.

样本利用情况对比表明,本文算法的样本利用率明显优于MLE算法,而且在样本集规模较小、样本信息稀疏时,优势更为明显.这是因为,本文算法以符合单一父节点与子节点状态组合的样本为参数估计依据,而MLE算法以符合父节点集与子节点状态组合的样本为估计依据.当面对稀疏结构样本集时,部分父节点集与子节点状态组合对应的样本可能很少或没有;而单一父节点与子节点之间的统计信息则更为广泛地存在于样本集之中.因此本文算法比MLE算法更能充分挖掘样本中的统计信息,提高参数的学习精度.

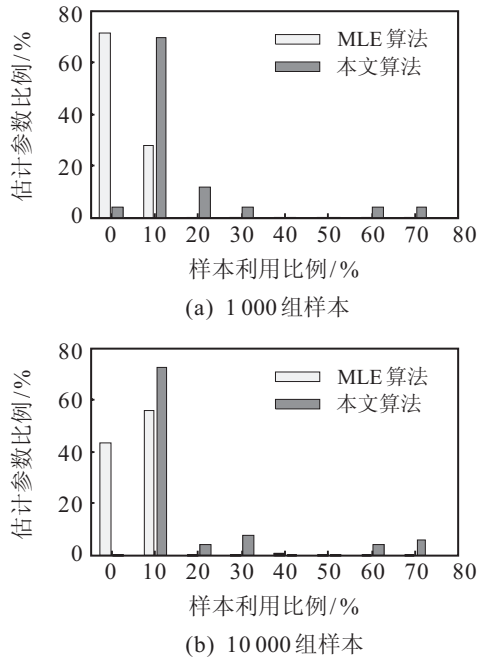


图 4 两种算法的样本利用情况比较

5 结 论

贝叶斯网络中局部结构的 CPT 规模和复杂度对模型参数量化和推理计算的精度与效率影响明显. ICI 模型能够对局部结构的条件概率分布做进一步因子分解, 并能降低 CPT 参数定义的复杂度. 本文以 Leaky ICI 模型的 Noisy-MAX 形式为例, 利用其中形成的特定上下文独立关系, 提出了一种适于样本学习、能由子节点对单父节点的条件概率分布直接合成 CPT 参数的算法. 实验结果表明, 该算法学习精度优于标准 MLE 算法, 且能较好地处理稀疏结构样本集, 但计算复杂度高于 MLE 算法. 本文方法要求局部结构满足因果影响独立性质. 进一步的研究在于如何降低计算复杂度和提高算法的适用性.

参考文献(References)

[1] Pearl J. Probabilistic reasoning in intelligent systems: Networks of plausible inference[C]. Networks of Plausible Inference. San Francisco: Morgan Kaufmann, 1988: 383-408.

[2] 张连文, 郭海鹏. 贝叶斯网引论[M]. 北京: 科学出版社, 2006: 31-74.
(Zhang L W, Guo H P. An introduction to Bayesian networks[M]. BeiJing: Science Press, 2006: 31-74.)

[3] Zhang N L, Poole D. Exploiting causal independence in Bayesian network inference[J]. J of Artificial Intelligence Research, 1996, 5(7): 301-328.

[4] 张宏毅, 王立威, 陈瑜希. 概率图模型研究进展综述[J]. 软件学报, 2013, 24(11): 2476-2497.
(Zhang H Y, Wang L W, Chen Y X. Research progress of probabilistic graphical models: A survey[J]. J of Software, 2013, 24(11): 2476-2497.)

[5] Heckerman D. Causal independence for knowledge acquisition and inference[C]. Proc of the 9th Conf on Uncertainty in Artificial Intelligence. San Mateo: Morgan Kaufmann Publishers, 1993: 122-127.

[6] Yang S, Natarajan S. Knowledge intensive learning: Combining qualitative constraints with causal independence for parameter learning in probabilistic models[C]. Machine Learning and Knowledge Discovery in Databases. Berlin Heidelberg: Springer, 2013: 580-595.

[7] Vomlel J, Tichavský P. Computationally efficient probabilistic inference with noisy threshold models based on a CP tensor decomposition[C]. Proc of the 6th European Workshop on Probabilistic Graphical Models(PGM 2012). Granada, 2012: 355-362.

[8] D'Ambrosio B. Symbolic probabilistic inference in large BN20 networks[C]. Proc of the 10th Conf on Uncertainty in Artificial Intelligence. San Mateo: Morgan Kaufmann Publishers, 1994: 128-135.

[9] Li W, Poupart P, van Beek P. Exploiting structure in weighted model counting approaches to probabilistic inference[J]. J of Artificial Intelligence Research, 2011, 40(1): 729-765.

[10] Pradhan M, Provan G, Middleton B, et al. Knowledge engineering for large belief networks[C]. Proc of the 10th Conf on Uncertainty in Artificial Intelligence. San Mateo: Morgan Kaufmann Publishers, 1994: 484-490.

[11] Luque M, Díez F J. Variable elimination for influence diagrams with super-value nodes[J]. Int J of Approximate Reasoning, 2010, 51(6): 615-631.

[12] Díez F J, Galán S F. Efficient computation for the Noisy-MAX[J]. Int J of Intelligent Systems, 2004, 18(2): 165-177.

[13] Díez F J, Druzdzel M J. Canonical probabilistic models for knowledge engineering[R]. Madrid: UNED, 2007.

[14] Boutilier C, Friedman N, Goldszmidt M, et al. Context-specific independence in Bayesian networks[C]. Proc of the 12th Conf on Uncertainty in Artificial Intelligence. San Mateo: Morgan Kaufmann Publishers, 1996: 115-123.

[15] Zagorecki A, Druzdzel M J. Knowledge engineering for Bayesian networks: How common are Noisy-MAX distributions in practice?[J]. IEEE Trans on Systems, Man, and Cybernetics: Systems, 2013, 43(1): 186-195.

[16] Zagorecki A, Voortman M, Druzdzel M J. Decomposing local probability distributions in Bayesian networks for improved inference and parameter learning[C]. Proc of the 19th Int Florida Artificial Intelligence Research Society Conf. Menlo Park: AAAI Press, 2006: 860-865.

(责任编辑: 李君玲)