

# 小数据集条件下的多态系统贝叶斯网络参数学习

肖蒙 张友鹏

(兰州交通大学自动化与电气工程学院 兰州 730070)

**摘要** 针对贝叶斯网络中多父节点条件概率分布参数学习问题,提出了一种适用于多态节点、模型不精确、样本信息不充分情形的参数学习方法。该方法利用因果机制独立假设,分解条件概率分布,使条件概率表的规模表现为父节点个数和状态数的线性形式;利用 *Leaky Noisy-MAX* 模型量化了多态系统模型未含因素对参数学习的影响;从小样本数据集中获取模型参数并合成条件概率表。结果表明,该方法能提高参数学习效率与精度。

**关键词** 贝叶斯网络,多态系统,小数据集,因果机制独立,参数学习

**中图分类号** TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.4.052

## Parameters Learning of Bayesian Networks for Multistate System with Small Sample

XIAO Meng ZHANG You-peng

(School of Automation and Electrical Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

**Abstract** To learn parameters for conditional probability distribution of multi-father nodes, a method was proposed which applies to the multistate nodes in the inaccurate model under the condition of insufficient sample information. Using the assumption of independence of causal interaction, the conditional probability distribution is decomposed and the size of conditional probability table is linear in the numbers of the parent nodes and their states. Using Leaky Noisy-MAX model, the influence of factors not included in the multistate system model can be quantified on the parameters learning. The model parameters extracted from small sample can create conditional probability tables. The results show that the method can improve the efficiency and precision of parameter learning.

**Keywords** Bayesian networks, Multistate system, Small sample, Independence of causal interaction, Parameters learning

## 1 引言

利用概率方法进行不确定性推理是人工智能研究领域的重要课题之一<sup>[1]</sup>,其实质是用一组随机变量来刻画问题,用联合概率分布来表示问题的知识,并按照概率论原则进行推理计算。联合概率分布是建立在所含全部随机变量状态空间的笛卡尔乘积之上的多维函数。当变量的数量很多或变量的值域范围较大时,联合概率分布的复杂程度将呈指数增长,导致直接利用联合概率分布进行不确定推理变得十分困难。

Pearl 提出了贝叶斯网络(Bayesian Networks)模型<sup>[2]</sup>,该模型用有向无环图来定性描述变量间的拓扑关系,即各变量对应图中的节点,变量之间的直接依赖关系对应图中的有向弧;用条件概率分布来定量刻画变量对其父节点的依赖与独立关系。贝叶斯网络模型能够利用变量间的条件独立关系,将联合分布分解成多个复杂程度低的概率分布,从而降低了系统概率模型的复杂度,提高了推理效率和模型实用性。贝叶斯网络有其严格的数学语言,同时也符合人脑进行信息获取和思维推理的基本形式,因此,成为了人工智能领域进行不确定性推理和建模的直观有效工具,已在辅助决策、医学诊

断、自然语言理解、故障诊断、图像解释等不确定推理和预测方面<sup>[3-5]</sup>有了很多成功的应用。

在对实际问题应用贝叶斯网络模型时,主要的难度在于建立模型拓扑结构和确定用于量化模型的概率参数<sup>[1]</sup>。概率参数包括祖先节点的先验边缘概率分布和中间节点、叶子节点的条件概率分布。当网络中某个子节点和其所有父节点都代表离散随机变量时,该子节点的条件概率分布以其所有父节点可能状态组合的条件概率表(Conditional Probability Table, CPT)形式呈现。对一个有  $n$  个二态父节点的二态子节点,其完全 CPT 定义需要  $2^n$  个独立参数。当父节点数量  $n$  增加,节点间关系趋向复杂时,CPT 规模将呈指数级增加。当面对的是多态系统,父节点和子节点代表的随机变量均有  $m$  个逻辑状态时,完全的 CPT 定义需要  $(m-1) \cdot m^n$  个独立参数。此时,CPT 的规模将以节点的状态数为底数,呈指数级增加,这无疑会使 CPT 参数学习和推理计算变得更为复杂且困难。

为高效获得 CPT 参数,研究者通过假设局部条件分布满足某种规律,如因果机制独立(Independence of Causal Interaction, ICI)<sup>[6,7]</sup>。利用 ICI 模型,建立具体问题的 Noisy 模型,

到稿日期:2014-05-18 返修日期:2014-08-13 本文受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资助项目(1112RJZA040)资助。

肖蒙(1974-),男,博士生,副教授,主要研究方向为交通信息工程、复杂系统可靠性,E-mail: xiaom@mail.lzjtu.cn;张友鹏(1965-),男,教授,博士生导师,主要研究方向为交通运输安全保障与防护技术。

能对局部条件分布作进一步因子分解,降低了CPT定义和参数获取的复杂度。文献[3-5]将Noisy模型用于目标分类、故障诊断及态势估计,并给出了具体模型参数的学习方法,但问题状态均为二态。文献[8,9]讨论了Noisy模型的参数学习及推理计算,但当样本数据不充分时,还不能直接得到模型参数。本文针对多态系统贝叶斯网络建模中参数的学习问题,在分析ICI模型的条件独立和因果机制的基础上,利用Leaky Noisy-MAX模型构造了一种只需小样本数据集进行CPT参数学习的方法,该方法能使条件概率表的规模表现为父节点数和状态数的线性形式,并充分考虑了模型不精确情况,从而有效降低了参数学习的复杂度,提高了参数学习精度。

## 2 因果机制独立分析及其模型

因果与独立关系是人的常用思维习惯,自然也常被用来分析问题和建立系统模型。在实际应用中,研究者往往利用因果关系来建立贝叶斯网络结构,其实质是用因果关系进行条件分布独立假设<sup>[10]</sup>。当通过对问题知识的具体分析,发现多个原因(父节点)共同影响同一结果(子节点)并且是各自独立地对结果造成影响时,这个局部网络的概率关系存在着条件独立和因果关系。此时,我们认为父节点与子节点之间满足因果机制独立假设,并可用ICI模型来表现其条件独立和因果关系。

假设某局部贝叶斯网络中的子节点 $Y$ 有 $n$ 个父节点 $\mathbf{X} = \{X_1, X_2, \dots, X_n\}$ ,如图1(a)所示,且 $Y$ 与 $\mathbf{X}$ 之间满足因果机制独立假设。为直观体现 $\mathbf{X}$ 对 $Y$ 的独立影响,我们引入辅助节点 $\mathbf{Z} = \{Z_1, Z_2, \dots, Z_n\}$ ,分别表示各父节点对 $Y$ 的独自影响结果<sup>[8]</sup>,而 $\mathbf{X}$ 对 $Y$ 的整体影响表现为 $\mathbf{Z}$ 的整体分布,如图1(b)所示。

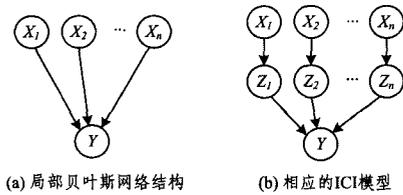


图1 局部贝叶斯网络结构的ICI模型

### 2.1 ICI模型的条件独立分析

由因果机制独立假设及图1(b)所示的 $\mathbf{X}, Y, \mathbf{Z}$ 间的拓扑结构,可以得到以下条件独立关系:

I. 父节点之间相互独立,即 $(X_1 \perp X_2 \perp \dots \perp X_n)$ ;

II. 辅助节点只依赖于其直接父节点,而独立于其它的父节点和辅助节点,即 $(Z_i \perp X_{j \in [1, n], j \neq i} | X_i), (Z_i \perp Z_{j \in [1, n], j \neq i} | X_i)$ ;

III. 根据贝叶斯网络的整体马尔科夫性<sup>[1]</sup>, $\mathbf{Z}$  d-分隔 $\mathbf{X}$ 和 $Y$ , $\mathbf{X}$ 与 $Y$ 在给定 $\mathbf{Z}$ 时条件独立,即 $(\mathbf{X} \perp Y | \mathbf{Z})$ 。

由条件独立关系 I、II 可以得到 $\mathbf{Z}$ 对 $\mathbf{X}$ 的条件分布:

$$P(\mathbf{Z} | \mathbf{X}) = P(Z_1, Z_2, \dots, Z_n | X_1, X_2, \dots, X_n) = \prod_i P(Z_i | X_i) \quad (1)$$

由条件独立关系 III 可以得到图1(b)所示的局部网络的联合概率分布:

$$P(\mathbf{X}, \mathbf{Z}, Y) = P(\mathbf{Z} | \mathbf{X})P(Y | \mathbf{Z})P(\mathbf{X}) \quad (2)$$

根据式(1)、式(2), $Y$ 节点的条件分布 $P(Y | \mathbf{X})$ 可由对 $\mathbf{Z}$ 的边缘概率得到:

$$P(Y | \mathbf{X}) = \frac{\sum_{\mathbf{Z}} P(\mathbf{X}, \mathbf{Z}, Y)}{P(\mathbf{X})} = \sum_{\mathbf{Z}} P(\mathbf{Z} | \mathbf{X})P(Y | \mathbf{Z}) \quad (3)$$

因为 $Z_i$ 代表父节点 $X_i$ 对 $Y$ 的独自影响结果,而 $Y$ 的可能状态取决于 $\mathbf{X}$ 的各自影响结果 $\mathbf{Z}$ 的合成。定义合成函数 $f(\mathbf{Z})$ ,且有

$$P(Y=y | \mathbf{Z}) = \begin{cases} 1, & \text{if } f(\mathbf{Z})=y \\ 0, & \text{if } f(\mathbf{Z}) \neq y \end{cases} \quad (4)$$

根据式(1)、式(3)、式(4),若记 $c_{z_i}^{x_i} = P(Z_i = z_i | X_i = x_i)$ ,可得条件概率分布 $P(Y | \mathbf{X})$ 的因子分解形式,即

$$P(Y=y | \mathbf{X}) = \sum_{\mathbf{Z}} P(Y=y | \mathbf{Z}) \prod_i P(Z_i | X_i) = \sum_{\mathbf{z} | f(\mathbf{z})=y} \prod_i c_{z_i}^{x_i} \quad (5)$$

式(5)即为条件概率分布 $P(Y | \mathbf{X})$ 的ICI模型定义。分析式(5)可知,我们利用因果机制独立假设,将多维局部条件概率分布 $P(Y | \mathbf{X})$ 因子分解为单维条件概率分布 $P(Z_i | X_i)$ 的乘积形式,从而降低了CPT获取的复杂度,也利于在样本数据不完整时有效获取参数。当原因节点和结果节点均代表二态随机变量时,完全定义CPT所需独立参数数量由 $2^n$ 降为 $2n$ 。这种CPT因子分解形式对多态系统的贝叶斯网络参数学习而言更为有效。

### 2.2 ICI模型的扩展形式

在图1(a)中,我们假设 $\mathbf{X}$ 包含了所有影响 $Y$ 的原因,即当所有原因均缺失时,结果 $Y$ 必然缺失。若记节点 $\mathbf{X}, Y$ 的缺失状态分别为 $\neg \mathbf{X} = \{\neg X_1, \neg X_2, \dots, \neg X_n\}, \neg Y$ 时, $P(\neg Y | \neg X_1, \neg X_2, \dots, \neg X_n) = 1$ 成立。但在建立具体复杂系统的贝叶斯网络时,有时不可能将全部原因都列入模型。而未列入的原因将隐含在专家知识或样本数据中,并会对已建立的贝叶斯网络模型的参数学习产生影响。这从因果关系上表现为当所有已列入模型的原因节点均处于缺失状态时,结果节点仍可能不呈现缺失状态,即出现 $0 < P(Y | \neg X_1, \neg X_2, \dots, \neg X_n) < 1$ 现象。

若考虑未列入原因的影响时,就可得到ICI模型的扩展形式,即Leaky ICI模型。如图2所示,虚线框内节点 $\mathbf{X}' = \{X'_1, X'_2, \dots, X'_n\}$ 为未列入原因,其对 $Y$ 的整体影响用辅助节点 $Z_L$ 表示。我们添加辅助节点 $\mathbf{Z}' = \{Z'_1, Z'_2, \dots, Z'_n\}$ ,表示 $\mathbf{X}'$ 对 $Z_L$ 的各自独立影响,且 $\mathbf{X}'$ 与 $Z_L$ 之间满足因果机制独立假设,从而 $\mathbf{X}', \mathbf{Z}', Z_L$ 构成局部ICI模型结构。由式(5)得到 $Z_L$ 对 $\mathbf{X}'$ 的条件概率分布为:

$$P(Z_L | \mathbf{X}') = \sum_{\mathbf{z}' | f(\mathbf{z}')=Z_L} \prod_i P(Z'_i | X'_i) \quad (6)$$

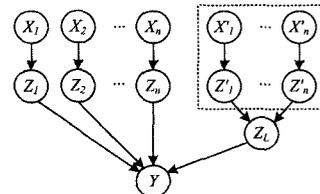


图2 扩展形式的ICI模型结构

在Leaky ICI模型中, $\mathbf{X}, \mathbf{X}'$ 各自独立影响 $Y$ ,即 $\mathbf{X} \perp \mathbf{X}'$ 。若记 $c_{z_L}^{z_i} = P(Z_L = z_L)$ ,根据式(5)、式(6), $Y$ 对 $\mathbf{X}$ 的条件概率分布为:

$$\begin{aligned} P(Y=y | \mathbf{X}) &= \sum_{\mathbf{X}'} P(Y=y | \mathbf{X}, \mathbf{X}')P(\mathbf{X}' | \mathbf{X}) \\ &= \sum_{\mathbf{X}'} \sum_{\mathbf{z}, z_L | f(\mathbf{z}, z_L)=y} P(\mathbf{Z} | \mathbf{X})P(Z_L | \mathbf{X}')P(\mathbf{X}') \\ &= \sum_{\mathbf{z}, z_L | f(\mathbf{z}, z_L)=y} \prod_i P(Z_i | X_i) \prod_i P(Z'_i | X'_i)P(X'_i) \end{aligned}$$

$$\begin{aligned}
&= \sum_{\mathbf{Z}} \prod_i P(Z_i | X_i) \sum_{Z_L | f(\mathbf{Z})=Z_L} \sum_{Z_L | f(\mathbf{Z}, Z_L)=y} \sum_{\mathbf{X}'} \prod_i P(Z_i', X_i') \\
&= \sum_{\mathbf{Z}} \prod_i P(Z_i | X_i) \sum_{Z_L | f(\mathbf{Z}, Z_L)=y} P(Z_L) \\
&= \sum_{\mathbf{Z}} \prod_i c_{Z_i}^{X_i} \sum_{Z_L | f(\mathbf{Z}, Z_L)=y} c_{Z_L}^y \quad (7)
\end{aligned}$$

式(7)即为 *Leaky* ICI 模型的定义。已列入模型中的原因节点  $\mathbf{X}$  对结果节点  $Y$  的影响与式(5)一致,而未列入模型中的原因即隐藏原因节点  $\mathbf{X}'$  是通过  $Z_L$  对结果节点施加影响的。*Leaky* ICI 模型由于考虑了已建模型不精确情况下其他因素对模型参数的影响,因此能有效提高模型参数学习的精度。

在 ICI 模型的实际应用中,根据  $\mathbf{X}$  对  $Y$  的整体影响方式,合成函数  $f(\mathbf{Z})$  可能有不同形式,常见的有逻辑形式 OR/AND 和代数形式 MAX/MIN。当系统内部事件或属性关系复杂时,其贝叶斯网络模型中可能同时出现多种形式。*Noisy-OR* 模型是一种常见的 ICI 模型形式,其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,这对于实际系统而言过于理想。对于复杂系统,系统内部事件或部件的可能状态更多地呈现出多态。*Noisy-MAX/MIN* 模型是 *Noisy-OR/AND* 模型在多态状态空间上的扩展,对建立复杂多态系统的贝叶斯网络模型更为适用。文献[11]以 3 种经典贝叶斯网络模型(ALARM、HAILFINDER、HEPARII)为例,对采用 MAX 合成形式的 ICI 模型的通用性进行了比较研究,研究表明 *Noisy-MAX* 对复杂条件概率分布的参数学习具有较好的适用性。文献[12]对 *Noisy-MAX* 模型中有向图的双向性进行了扩展,提高了推理计算效率。下面针对 MAX 合成形式的具体 ICI 模型进行讨论,以得到模型参数的获取方式并通过模型参数构造 CPT。

### 3 MAX 合成形式的 ICI 模型

当合成函数  $f(\mathbf{Z})$  以 MAX 形式出现时,我们称这种 ICI 模型为 *Noisy-MAX* 模型,其结果节点的状态取决于各原因节点的最大影响,即:

$$Y = f(\mathbf{Z}) = \text{Max}(Z_1, Z_2, \dots, Z_n) \quad (8)$$

#### 3.1 Noisy-MAX 模型

对于 *Noisy-MAX* 模型,辅助节点  $Z_i$  代表在其直接原因节点  $X_i$  的影响下结果节点  $Y$  的可能状态取值,所有原因节点  $\mathbf{X}$  对结果节点  $Y$  的整体影响由式(8)合成得到,因此辅助节点  $Y$  与  $\mathbf{Z}$  所对应的随机变量有一样的状态空间。应用 *Noisy-MAX* 模型时,需对结果节点的状态取值进行排序,且其缺失状态为最小值<sup>[8]</sup>。当  $Y \leq y$  时,

$$f(\mathbf{Z}) \leq y \Rightarrow \forall Z_i \leq y \quad (9)$$

定义模型累加参数  $C_y^{X_i} = \sum_{Z_i \leq y} c_{Z_i}^{X_i}$ , 由式(5)可得到 *Noisy-MAX* 模型的概率分解形式:

$$P(Y \leq y | \mathbf{X}) = \sum_{\mathbf{Z} | \text{Max}(\mathbf{Z}) \leq y} \prod_i c_{Z_i}^{X_i} = \sum_{Z_1 \leq y} \dots \sum_{Z_n \leq y} \prod_i c_{Z_i}^{X_i} = \prod_i C_y^{X_i} \quad (10)$$

则  $Y$  对  $\mathbf{X}$  的条件概率分布为:

$$P(Y=y | \mathbf{X}) = \begin{cases} P(Y \leq y | \mathbf{X}) - P(Y \leq y-1 | \mathbf{X}) \\ = \prod_i C_y^{X_i} - \prod_i C_{y-1}^{X_i}, & \text{if } y \neq y_{\min} \\ P(Y \leq y | \mathbf{X}) = \prod_i C_y^{X_i}, & \text{if } y = y_{\min} \end{cases} \quad (11)$$

若设节点  $Y$  的状态数为  $m_y$ ,  $\mathbf{X}$  中各节点的状态数为  $\{m_1, m_2, \dots, m_n\}$ 。通过分析式(10)可知,完全定义节点  $Y$  的 CPT 所需独立参数的数量,由关于父节点个数和状态数的指数形式降为线性形式,即由  $(m_y - 1) \prod_i m_i$  减少为  $(m_y - 1) \sum_i m_i$ 。

#### 3.2 Leaky Noisy-MAX 模型

*Noisy-MAX* 模型未考虑未列入原因对模型参数的影响,易导致在实际建模中出现置信过度现象。若将所有未列入模型的其他未知原因对  $Y$  的影响仍表示为  $Z_L$ , 并定义模型参数  $c_y^L = P(Z_L = y)$  和模型累加参数  $C_y^L = \sum_{Z_L \leq y} c_{Z_L}^L$ , 由式(7)、式(10)可以得到 *Leaky Noisy-MAX* 的因子分解形式:

$$P(Y \leq y | \mathbf{X}) = \prod_i C_y^{X_i} \sum_{Z_L \leq y} P(Z_L) = C_y^L \prod_i C_y^{X_i} \quad (12)$$

同样,结合式(11),可以得到节点  $Y$  的 CPT。

### 4 Leaky Noisy-MAX 模型参数的获取

在应用 *Leaky Noisy-MAX* 模型来分析、构造多态系统贝叶斯网络模型的局部条件分布时,模型参数  $c_y^{X_i}$ 、 $c_y^L$  的学习是关键。如同贝叶斯网络参数学习,模型参数一般也是通过专家知识提取或样本数据学习得到。

#### 4.1 模型参数学习的 3 种形式

根据样本数据集规模或专家知识提取规则不同,*Leaky Noisy-MAX* 模型参数的学习方式可有以下 3 种形式。

I. 样本数据集丰富完整,能够对列入模型的已知原因直接得到其参数  $c_y^{X_i}$ ,也能够得到未列入模型的其他未知原因的其参数  $c_y^L$ ,进而通过式(11)、式(12)得到 CPT 参数。此时,  $c_y^{X_i}$  表示“当所有其他原因缺失时,若  $X_i = x_i$ ,则  $Y = y_i$  的概率是  $c_y^{X_i}$ ”<sup>[9]</sup>,即模型参数的概率形式为  $c_y^{X_i} = P(Y = y | X_i = x_i, \neg X_{j[\forall j \neq i]}), \neg Z_L)$ ,  $c_y^L = P(Y = y | \neg X_1, \neg X_2, \dots, \neg X_n)$ 。

II. 样本数据集部分完整,但缺失未列入模型的其他未知原因的相关数据,只能得到“当已列入模型的其他已知原因缺失时,若  $X_i = x_i$ ,则  $Y = y$  的概率是  $p_y^{X_i}$ ”,即  $p_y^{X_i} = P(Y = y | X_i = x_i, \neg X_{j[\forall j \neq i]})$ 。此类形式可视为是第 I 类形式对  $Z_L$  的边缘分布,即  $c_y^{X_i} = \sum_{Z_L} p_y^{X_i}$ 。而参数  $c_y^L$  隐含在  $p_y^{X_i}$  中,不能直接得到。

III. 样本缺失较多,只能从小数据集中得到“若  $X_i = x_i$ ,则  $Y = y$  的概率是  $q_y^{X_i}$ ”,即  $q_y^{X_i} = P(Y = y | X_i = x_i)$ 。此类形式可视为是第 II 类形式对其它已知原因的边缘分布,即  $q_y^{X_i} = \sum_{X_{X_i}} p_y^{X_i}$ 。同样,参数  $c_y^L$  隐含在  $q_y^{X_i}$  中,不能直接得到。

在多态系统的实际建模中,经常面对样本数据集不完整或缺失较多的情形。下面主要对第 III 类形式的模型参数学习进行讨论。

#### 4.2 小数据集的模型参数学习

定义累计概率  $Q_y^{X_i} = P(Y \leq y | X_i = x_i)$ , 并将  $Q_y^{X_i}$  在整个  $\mathbf{X}$  空间分解。由式(12)得:

$$\begin{aligned}
Q_y^{X_i} &= P(Y \leq y | X_i = x_i) \\
&= \frac{P(Y \leq y, X_i = x_i)}{P(X_i = x_i)} = \frac{\sum_{\mathbf{X} | X_i = x_i} P(Y \leq y, \mathbf{X})}{P(X_i = x_i)} \\
&= \frac{\sum_{\mathbf{X} | X_i = x_i} P(Y \leq y | \mathbf{X}) P(\mathbf{X})}{P(X_i = x_i)} = \frac{\sum_{\mathbf{X} | X_i = x_i} C_y^L \prod_i C_y^{X_i} P(\mathbf{X})}{P(X_i = x_i)}
\end{aligned}$$

$$= \frac{C_y^x C_y^i P(X_i = x_i) \prod_{X_j \in \mathcal{X}(X_i)} C_y^{X_j} P(X_j)}{P(X_i = x_i)}$$

$$= C_y^x C_y^i \prod_j \sum_{X_j \in \mathcal{X}(X_i)} C_y^{X_j} P(X_j) \quad (13)$$

由式(13)可得到当  $X_i = x_i$  时的模型参数  $C_y^L$ , 记为  $C_y^{L_i}$ :

$$C_y^{L_i} = \frac{Q_y^{x_i}}{C_y^x \prod_j \sum_{X_j \in \mathcal{X}(X_i)} C_y^{X_j} P(X_j)} \quad (14)$$

模型参数  $C_y^L$  可由各原因节点对应的参数  $C_y^{L_i}$  的均值得到, 即

$$C_y^L = \sum_i P(X_i) C_y^{L_i} \quad (15)$$

当  $X_i = \neg x_i$  时, 因  $C_y^{\neg x_i} = 1$ , 由式(13)得

$$\frac{Q_y^{x_i}}{Q_y^{\neg x_i}} = \frac{C_y^L C_y^{x_i} \prod_j \sum_{X_j \in \mathcal{X}(X_i)} C_y^{X_j} P(X_j)}{C_y^L C_y^{\neg x_i} \prod_j \sum_{X_j \in \mathcal{X}(X_i)} C_y^{X_j} P(X_j)} = \frac{C_y^{x_i}}{C_y^{\neg x_i}} = C_y^{x_i} \quad (16)$$

由式(14)一式(16)可得到各模型参数  $c_y^{x_i}$ 、 $c_y^L$ , 进而通过式(11)、式(12)计算得到结果节点  $Y$  的 CPT 参数。

## 5 算例分析

某贝叶斯网络局部有 2 个父节点  $X_1$ 、 $X_2$  和 1 个子节点  $Y$ 。 $X_1$ 、 $X_2$  各自独立地影响  $Y$ , 即满足因果机制独立假设。各节点变量的状态空间为:  $X_1 = \{\text{decreased, normal, increased}\}$ ,  $X_2 = \{\text{absent, present}\}$ ,  $Y = \{\text{absent, mild, moderate, severe}\}$ , 它们的缺失状态分别为:  $\neg X_1 = \text{normal}$ ,  $\neg X_2 = \text{absent}$ ,  $\neg Y = \text{absent}$ , 且父节点  $X_1$ 、 $X_2$  的先验概率  $P(X)$  如表 1 所列。

表 1 父节点的先验概率表

P(X <sub>1</sub> )			P(X <sub>2</sub> )	
decreased	normal	increased	absent	present
0.1	0.8	0.1	0.7	0.3

由于样本数据不完备, 我们只能获得子节点对父节点的条件概率  $q_y^{x_i}$ , 如表 2 所列。表 2 中所含独立参数个数为  $[(4-1) \times (3+2)] = 15$ 。

表 2 子节点对父节点的边缘条件概率表

Y	P(Y X <sub>1</sub> )			P(Y X <sub>2</sub> )	
	decreased	normal	increased	absent	present
severe	0.1	0.05	0.1	0.05	0.1
moderate	0.3	0.1	0.4	0.05	0.2
mild	0.4	0.15	0.3	0.1	0.2
absent	0.2	0.7	0.2	0.8	0.5

根据表 2 中条件概率  $q_y^{x_i}$ , 由式(14)一式(16)可得到各模型参数  $c_y^{x_i}$ 、 $c_y^L$ , 如表 3 所列。

表 3 模型参数表

Y	$c_y^{x_1}$			$c_y^{x_2}$		$c_y^L$
	decreased	normal	increased	absent	present	
severe	0.0526	0	0.0526	0	0.0526	0.0275
moderate	0.3592	0	0.2415	0	0.1696	0.0348
mild	0.3025	0	0.4202	0	0.1528	0.0815
absent	0.2857	1	0.2857	1	0.6250	0.8562

表 3 中参数  $c_{y=\text{normal}}^{x_1} = 1$  和  $c_{y=\text{absent}}^{x_2} = 1$ , 与 ICI 模型中因果条件独立假设一致。由于样本信息不充分, 模型未包括因素的概率影响  $c_y^L$  无法直接获得, 但隐含在条件概率  $q_y^{x_i}$  中并可由式(14)析出。根据表 3 中的各模型参数, 由式(11)、式(12)可得到节点  $Y$  的 CPT, 如表 4 所列。

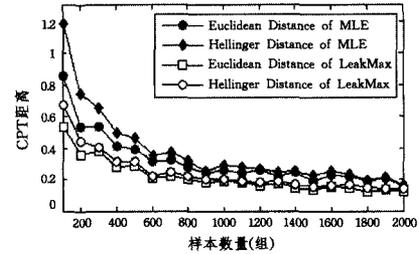
表 4 节点  $Y$  的 CPT 参数表

Y	P(Y X <sub>1</sub> , X <sub>2</sub> )					
	increased present	increased absent	normal present	normal absent	decreased present	decreased absent
severe	0.12712	0.07865	0.07865	0.0275	0.12712	0.07865
moderate	0.44388	0.36979	0.192	0.0348	0.35804	0.25942
mild	0.27611	0.30694	0.19422	0.0815	0.36196	0.41731
absent	0.15289	0.24462	0.53512	0.8562	0.15289	0.24462

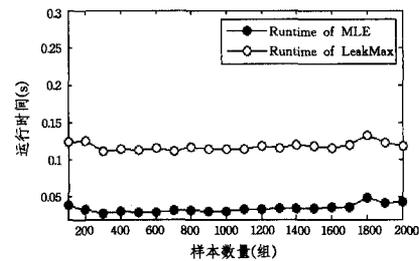
表 4 中所含独立参数个数为  $[(4-1) \times (3 \times 2)] = 18$ 。与表 2 对比可知, 应用 *Leaky Noisy-MAX* 模型, 在量化未列入因素影响的同时, 可将 CPT 完全定义所需的独立参数个数减少为父节点数和状态数的线性形式, 从而降低了完全 CPT 定义的复杂度。

## 6 对比实验分析

以样本数据完整下的最大似然估计 (MLE) 算法和样本数据随机缺失下的期望优化 (EM) 算法为对比, 以学习得到的 CPT 与标准 CPT 之间的 Euclidean 距离和 Hillinger 距离<sup>[11,13]</sup> 为算法学习精度衡量, 以算法运行所需机时为算法效率衡量, 来测试本文提出的参数学习算法的性能。在 Matlab 中建立前文所示的局部贝叶斯网络, 以表 4 中的 CPT 为标准 CPT, 并随机生成不同规模和不同数据缺失率的样本数据集。针对不同规模的完整样本集 (所含样本组数分别为 100, 200, 300, ..., 2000), 利用 MLE 算法和本文算法分别进行 10 次 CPT 参数学习, 实验统计数据如图 3 所示。



(a) 学习精度比较

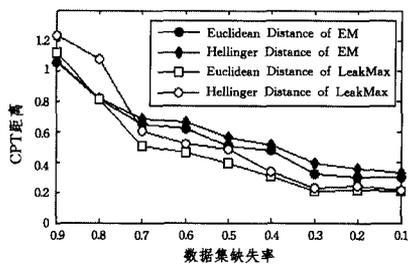


(b) 学习效率比较

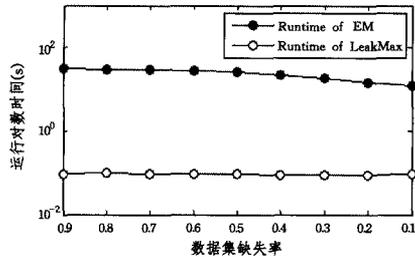
图 3 完整样本集下参数学习性能的比较

由图 3(a) 可知, 完整样本组数越小, 即呈现为小规模数据集时, 本文所提算法的参数学习精度越优于 MLE 算法。当数据集规模增加时, 两种算法的学习精度趋于接近。由图 3(b) 可知, 当样本数据完整时, 本文所提算法的学习效率略低于 MLE 算法, 运行时间最大差值为 0.092s。

针对数据随机缺失的不完整样本集 (所含样本组数均为 1000, 数据缺失率分别为 0.9, 0.8, 0.7, ..., 0.1), 利用 EM 算法和本文算法分别进行 10 次 CPT 参数学习, 实验统计数据如图 4 所示。



(a)学习精度比较



(b)学习效率比较

图4 不完整样本集下参数学习性能的比较

由图4(a)可见,在样本集的数据缺失率达到0.9和0.8时,本文所提算法的学习精度略低于EM算法,但随着数据缺失率逐渐减少,其学习精度逐步优于EM算法,这也与图3(a)所示的统计结果一致。图4(b)显示出,在样本中数据随机缺失时,EM算法的学习效率远低于本文所提算法,这是由于EM算法通过一定的迭代次数来保证学习精度,且数据缺失越多,其计算收敛速度越慢<sup>[1]</sup>。在本实验中,EM算法的迭代次数设为10,其与本文所提算法的运行时间最大差值出现在数据缺失率为0.9处,达到29.95s。综合上述对比实验分析可知,在样本信息不充分的小数据集条件下的CPT参数学习中,本文所提算法能有效提高参数学习的精度与效率。

**结束语** 贝叶斯网络利用条件独立关系,将联合概率分布分解为局部条件概率分布的乘积,降低了系统概率模型的复杂度。对于复杂多态系统,当变量的数量很多或变量的值域范围较大时,局部条件概率表的规模和复杂度将按指数级增长,这使得条件概率参数学习变得困难,从而影响了整个网络模型的实用性。本文利用因果机制独立假设,将条件概率分布分解成以单一原因的独立影响概率分布为因子的连乘形式,降低了条件概率分布的复杂度。在模型不精确和样本信息不充分情况下,利用Leaky Noisy-MAX模型,提出了一种条件概率表参数学习方法,其能有效降低参数学习的复杂度,

提高参数学习性能,增强了模型的实用性。

## 参考文献

- [1] 张连文,郭海鹏. 贝叶斯网引论[M]. 北京:科学出版社,2006
- [2] Pearl J. Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference[M]. San Francisco, CA: Morgan Kaufmann Publishers Inc., 1988:383-408
- [3] 王华伟,周经伦,何祖玉,等. 基于贝叶斯网络的复杂系统故障诊断[J]. 计算机集成制造系统,2004,10(2):230-234
- [4] 康长青,方磊,华丽,等. 基于贝叶斯 Noisy Or Gate 网络的多传感器目标分类识别[J]. 计算机测量与控制,2011,19(6):1387-1389
- [5] 柴慧敏,王宝树. 用于态势估计的一种构造贝叶斯网络参数的方法[J]. 计算机科学,2006,33(9):140-142
- [6] Heckerman D. Causal Independence for Knowledge Acquisition and Inference[C]//Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence (UAI'93). San Mateo: Morgan Kaufmann Publishers Inc., 1993:122-127
- [7] Zhang N L, Poole D. Exploiting Causal Independence in Bayesian Network Inference [J]. Journal of Artificial Intelligence Research, 1996, 5:301-328
- [8] Díez F J, Galán S F. Efficient Computation for the Noisy-Max [J]. International Journal of Intelligent Systems, 2004, 18(2): 165-177
- [9] Díez F J, Druzdzel M J. Canonical Probabilistic Models for Knowledge Engineering[R]. Technical Report CISIAD-06-01. UNED Madrid, 2007
- [10] Spirtes P, Glymour C N, Scheines R. Causation, Prediction, and Search[M]. Cambridge: The MIT Press, 2000
- [11] Zagorecki A, Druzdzel M J. Knowledge Engineering for Bayesian Networks: How Common Are Noisy-MAX Distributions in Practice? [J]. IEEE Transactions on Systems, Man, and Cybernetics; Systems, 2013, 43(1): 186-195
- [12] Shi D. Extending Noisy-Max Gates to Bidirectional Models[J]. Journal of Information & Computational Science, 2013, 10(13): 4085-4096
- [13] Zagorecki A, Voortman M, Druzdzel M J. Decomposing Local Probability Distributions in Bayesian Networks for Improved Inference and Parameter Learning[C]//Proceedings of the 19th International Florida Artificial Intelligence Research Society Conference (FLAIRS-2006). Menlo Park, CA: AAAI Press, 2006:860-865

(上接第220页)

- [4] Cimatti A, Pistore M, Roveri M, et al. Weak, strong, and strong cyclic planning via symbolic model checking[J]. Artificial Intelligence, 2003, 147(1/2): 35-84
- [5] 文中华,黄巍,刘任任,等. 模型检测规划中的状态分层方法[J]. 软件学报, 2009, 20(4): 858-869
- [6] Oh J, Meneguzzi F, Sycara K. Probabilistic plan recognition for intelligent information agents[C]//Proc. ICAART. 2011
- [7] Martín Y E, Moreno M D R, Smith D E. Probabilistic Plan Graph Heuristic for Probabilistic Planning[C]//Proc. of the Twenty-Fifth AAAI Conference on Artificial Intelligence. 2011: 1774-1745

- [8] Fu J, Ng V, Bastani F B, et al. Simple and fast strong cyclic planning for fully-observable nondeterministic planning problems [C]//Proc. of the 22nd Int Joint Conf on Artificial Intelligence. Menlo Park, CA: AAAI Press, 2011: 1949-1954
- [9] 周俊萍,殷明浩,谷文祥,等. 部分可观察强规划中约减观察变量的研究[J]. 软件学报, 2009, 20(2): 290-304
- [10] 唐杰,文中华,汪泉,等. 不确定可逆规划的强循环规划解[J]. 计算机研究与发展, 2013, 50(9): 1970-1980
- [11] 刘铭,徐杨,陈崢,等. 基于 Multi-agent 系统的多飞行器协同路径规划方法的研究[J]. 计算机科学, 2012, 39(1): 219-233
- [12] 陈建林,文中华,马丽丽,等. 一种求解最小权值强规划的方法[J]. 计算机工程, 2011, 37(17): 167-171