

# 扩展关系数据模型下的 Credal 网络近似推理

瞿英<sup>1,2</sup>, 吴祈宗<sup>1</sup>, 李萍<sup>1</sup>

QU Ying<sup>1,2</sup>, WU Qi-zong<sup>1</sup>, LI Ping<sup>1</sup>

1.北京理工大学 管理与经济学院,北京 100081

2.河北科技大学 经济与管理学院,石家庄 050081

1.School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China

2.School of Economics and Management, Hebei University of Science and Technology, Shijiazhuang 050081, China

E-mail: quying1973@126.com

QU Ying, WU Qi-zong, LI Ping. Method of credal network approximation inference based on extended relational data model. *Computer Engineering and Applications*, 2010, 46(1): 39-41.

**Abstract:** The extended relational data model is extended with the two operators of extended product join and extended marginalization defined. A method to make Credal network approximation inference is proposed by computing the joint probability distribution and posterior probability of variables through the above operators. Furthermore its arithmetic is developed by SQL statement. Its validity and efficiency can be showed in the application case. The relational data structure of Credal network is constructed. Thus the query optimization mechanism of relational database can be fully used and the efficiency of Credal network inference can be improved greatly.

**Key words:** Credal network; extended relational data model; extended product join; extended marginalization

**摘要:**对扩展关系数据模型作进一步延伸,构建了扩展乘积连接和扩展边缘连接操作。提出利用上述操作计算 Credal 网络中各结点联合概率分布和后验概率的方法,从而实现了 Credal 网络的近似推理,并给出了该近似推理的 SQL 语言实现算法,算例结果表明该方法是有用的。建立了 Credal 网络的关系数据结构,利用关系数据库的查询优化机制,有效规避了 Credal 网络推理过程中 Credal 集顶点组合爆炸问题,提高了 Credal 网络推理的效率。

**关键词:**Credal 网络;扩展关系数据模型;扩展乘积连接操作;扩展边缘操作

**DOI:**10.3778/j.issn.1002-8331.2010.01.013 **文章编号:**1002-8331(2010)01-0039-03 **文献标识码:**A **中图分类号:**TP311

## 1 引言

Credal 网络是贝叶斯网络的扩展,通过概率分布集合或凸集分布理论在参数概率值上引入不确定性,来表示有关参数概率的知识不完备性和模糊性,符合人类专家认识问题的客观规律,从而为在概率模型下模糊不确定信息对决策结果影响的研究提供理论支持,并为可靠性决策提供理论依据。Credal 网络的本质是放宽参数概率严格要求的贝叶斯网络,不必为每个节点指定在其父节点一定配置下的唯一的概率分布,而是可以用概率分布集合、概率分布区间、不等式约束等不同的概率测度方法来表达专家或数据信息形成的对参数概率值的不确定性表达<sup>[1-2]</sup>。Credal 网络的推理是建立在贝叶斯网络推理的基础之上,计算在一定扩展和证据条件下的某一变量的分类值的最小/最大概率(即区间的上下界)。在强扩展类型下的查询变量

的后验概率计算问题可以看作在 Credal 集约束下的贝叶斯网络推理的全局优化问题<sup>[1]</sup>。解决这种优化问题的困难在于各边缘 Credal 集或条件 Credal 集给 Credal 网络强扩展带来了大量的顶点组合,即使对于小型的网络也容易出现顶点组合爆炸问题。1996 年, S.K.M 提出了用扩展的关系模型表示贝叶斯网络<sup>[3-4]</sup>,从而使贝叶斯网络与关系数据库相结合,充分利用了传统关系模型的优点,使贝叶斯网络推理实现起来更为容易。对文献[4-5]中的扩展关系模型概念及边缘操作和乘积联结作进一步拓展,提出一种利用拓展的扩展关系模型实现 Credal 网络推理的方法,并给出了具体实现过程。

## 2 扩展关系数据模型的推广

传统的关系模型中,一张关系表  $R$  包含  $n$  个属性  $\{A_1, A_2,$

**基金项目:**河北省科技厅科技支撑计划项目(the Scientific and Technology Plan of Department of Science and Technology of Hebei Province under Grant No.09213509D);河北省科技厅软科学基金(the Soft Science Foundation of Department of Science and Technology of Hebei Province under Grant No.074572214);河北省教育厅软科学基金(the Soft Science Foundation of Department of Education of Hebei Province under Grant No.2007205)。

**作者简介:**瞿英(1973-),女,博士生,副教授,主要研究领域为数据挖掘,信息系统理论与方法等;吴祈宗(1947-),男,博士生导师,教授,主要研究领域为信息系统理论与方法,管理决策的方法与技术等;李萍(1978-),女,博士生,主要研究领域为数据挖掘,信息系统理论与方法等。

**收稿日期:**2009-01-13 **修回日期:**2009-02-16

..., A<sub>n</sub>)。在扩展的关系模型中,一张扩展关系表  $\Phi_R$ ,除了包含  $n$  个属性外,还增加了一列  $f_{\Phi_R}$ ,可以看作是元组  $t$  到非负实数域的一个映射,  $\Phi_R: \{t|t \in \Phi_R\} \rightarrow [0, +\infty]^{3-4}$ 。也就是说扩展的关系数据模型是在传统关系数据模型的基础上增加了一个属性  $f_{\Phi_R}$ ,该属性的值由函数  $\phi_R$  决定。若定义  $\phi_R(t_i)=1$ ,则该模型中各元组分量  $t[f_{\Phi_R}]$  的值均为 1,该列可以缺省,此时该模型就是传统的关系数据模型,因此传统的关系数据模型可以看作是扩展关系数据模型的特例。

Credal 网络是贝叶斯网络的扩展,其中各结点的参数可以是概率分布集合、概率分布区间、不等式约束等不同的概率测度方法,这里仅考虑参数为概率区间的形式。为了适应区间概率的形式,这里把扩展关系模型中的  $f_{\Phi_R}$  分解为两个属性  $f_{\Phi_R}^L$  和  $f_{\Phi_R}^U$ ,其中  $t[f_{\Phi_R}^L] \leq t[f_{\Phi_R}^U]$ ,如表 1 所示,分别表示 Credal 网络的各节点的上下概率,这样就形成了扩展关系模型的进一步拓展形式  $\Phi_R^e$ 。

表 1 扩展关系数据模型的延伸  $\Phi_R^e$

A <sub>1</sub>	A <sub>2</sub>	...	A <sub>m</sub>	$f_{\Phi_R}^L$	$f_{\Phi_R}^U$
$t_{11}$	$t_{12}$	...	$t_{1m}$	$\phi_R^L(t_1)$	$\phi_R^U(t_1)$
$t_{21}$	$t_{22}$	...	$t_{2m}$	$\phi_R^L(t_2)$	$\phi_R^U(t_2)$
⋮	⋮	⋮	⋮	⋮	⋮
$t_{i1}$	$t_{i2}$	...	$t_{im}$	$\phi_R^L(t_m)$	$\phi_R^U(t_m)$

扩展关系数据模型支持传统关系数据模型的基本数据操作,如选择、投影、连接等,还增加了两个新的操作,即边缘操作(Marginalization)和乘积连接(Product join)<sup>[4-5]</sup>。这里对这两个操作作如下扩展。

**定义 1 (扩展边缘操作)** 设  $X$  是属性全集  $U$  的一个子集,  $\Phi_R^e(\downarrow X)$  表示在关系  $\Phi_R^e$  下对  $X$  属性的边缘操作,操作结果为属性  $X \cup f_{\Phi_R}$  上的一个关系,从  $\Phi_R^e$  中构造  $\Phi_R^e(\downarrow X)$  如下:

(1) 在  $X \cup f_{\Phi_R}^L \cup f_{\Phi_R}^U$  属性上对  $\Phi_R^e$  进行投影,不删除重复元组,得到  $\Phi_R^e[X \cup f_{\Phi_R}^L \cup f_{\Phi_R}^U]$ ;

(2) 设  $t$  为  $\Phi_R^e[X \cup f_{\Phi_R}^L \cup f_{\Phi_R}^U]$  中的一个元组,对所有  $t[X]$  相同的元组,将它们合并为一个元组  $t'$ ,使对于  $X$  的每一个配置  $X_c$ ,  $t'[X]=t[X]=X_c$ ,  $\Phi_R^e(\downarrow X)\phi_R^L(t')(t') = \sum_{t[X]=X_c} \phi_R^L(t)$ ,  $\Phi_R^e(\downarrow X)\phi_R^U(t')(t') = \sum_{t[X]=X_c} \phi_R^U(t)$ 。

**例 1** 图 1 中,  $\Phi_R^e$  对  $A_1, A_2$  扩展边缘操作得到扩展边缘关系  $\Phi_R^e(\downarrow A_1 A_2)$

$\Phi_R^e =$	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>	$f_{\Phi_R}^L$	$f_{\Phi_R}^U$
	0	0	0	1	$d_{11}$	$d_{12}$
	0	0	1	1	$d_{21}$	$d_{22}$
	0	1	0	0	$d_{31}$	$d_{32}$
	0	1	1	0	$d_{31}$	$d_{32}$
	1	0	0	1	$d_{41}$	$d_{42}$
$\Phi_R^e(\downarrow A_1 A_2) =$	A <sub>1</sub>	A <sub>2</sub>	$f_{\Phi_R}^L$	$f_{\Phi_R}^U$		
	0	0	$d_{11}+d_{21}$	$d_{12}+d_{22}$		
	0	1	$d_{31}+d_{31}$	$d_{32}+d_{32}$		
	1	0	$d_{41}$	$d_{42}$		

图 1 扩展边缘操作示例图

**定义 2 (扩展乘积连接操作)** 设  $\Phi_X^e$  和  $\Phi_Y^e$  是两个拓展的扩

展关系模型,其关系模式分别为  $\Phi_X^e(X \cup f_{\Phi_X}^L \cup f_{\Phi_X}^U)$  和  $\Phi_Y^e(Y \cup f_{\Phi_Y}^L \cup f_{\Phi_Y}^U)$ ,  $\Phi_X^e$  和  $\Phi_Y^e$  的扩展乘积连接  $\Phi_X^e \times \Phi_Y^e$  可构造如下:

- (1) 计算  $\Phi_X^e$  和  $\Phi_Y^e$  的自然联结  $\Phi_X^e M \Phi_Y^e$ 。
- (2) 增加新列  $f_{\Phi_X \Phi_Y}^L$  和  $f_{\Phi_X \Phi_Y}^U$  到  $\Phi_X^e M \Phi_Y^e$  中,  $f_{\Phi_X \Phi_Y}^L$  定义为  $\phi_X^L(t[X]) \cdot \phi_Y^L(t[Y])$ ,  $f_{\Phi_X \Phi_Y}^U$  定义为  $\phi_X^U(t[X]) \cdot \phi_Y^U(t[Y])$ ,  $t$  为  $\Phi_X^e M \Phi_Y^e$  中的一个元组。
- (3) 将  $\Phi_X^e M \Phi_Y^e$  投影到属性集  $X \cup Y \cup f_{\Phi_X \Phi_Y}^L$  上,即得到  $\Phi_X^e \times \Phi_Y^e$ 。

**例 2** 图 2 中  $\Phi_X^e$  与  $\Phi_Y^e$  进行扩展乘积连接操作得到关系  $\Phi_X^e \times \Phi_Y^e$ 。

A <sub>1</sub>	A <sub>2</sub>	$f_{\Phi_X}^L$	$f_{\Phi_X}^U$	
0	0	$a_{11}$	$a_{12}$	
0	1	$a_{21}$	$a_{22}$	
1	0	$a_{31}$	$a_{32}$	
1	1	$a_{41}$	$a_{42}$	
A <sub>2</sub>	A <sub>3</sub>	$f_{\Phi_Y}^L$	$f_{\Phi_Y}^U$	
0	0	$b_{11}$	$b_{12}$	
0	1	$b_{21}$	$b_{22}$	
1	0	$b_{31}$	$b_{32}$	
1	1	$b_{41}$	$b_{42}$	
A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	$f_{\Phi_X \Phi_Y}^L$	$f_{\Phi_X \Phi_Y}^U$
0	0	0	$a_{11} \cdot b_{11}$	$a_{12} \cdot b_{12}$
0	0	1	$a_{11} \cdot b_{21}$	$a_{12} \cdot b_{22}$
0	1	0	$a_{21} \cdot b_{31}$	$a_{22} \cdot b_{32}$
0	1	1	$a_{21} \cdot b_{41}$	$a_{22} \cdot b_{42}$
1	0	0	$a_{31} \cdot b_{11}$	$a_{32} \cdot b_{12}$
1	0	1	$a_{31} \cdot b_{21}$	$a_{32} \cdot b_{22}$
1	1	0	$a_{41} \cdot b_{31}$	$a_{42} \cdot b_{32}$
1	1	1	$a_{41} \cdot b_{41}$	$a_{42} \cdot b_{42}$

图 2 扩展乘积关系示例图

### 3 利用扩展关系模型进行 Credal 网络近似推理

#### 3.1 Credal 网络近似推理原理

所谓推理就是在给定证据  $X_E=e$  发生条件下,推断查询变量  $X_q$  各种状态发生的概率,即求解  $K(X_q|X_E=e)$ 。由条件概率计算方法

$$K(X_q|X_E=e) = \frac{K(X_q, X_E=e)}{K(X_E=e)} \tag{1}$$

可知,如果能求出变量的联合概率分布,就可以利用边缘化求出条件概率。这里利用拓展的扩展关系数据模型来求解所有变量的联合概率分布。首先,用拓展的扩展关系表存储贝叶斯网各个结点上的条件概率区间,  $n$  个结点对应  $n$  张扩展关系表,即 Credal 网的每个 CPT 转化为一个扩展关系表,表中的  $f_{\Phi_R}^L$  为结点配置的下概率,  $f_{\Phi_R}^U$  为结点配置的上概率,对所有表进行扩展乘积连接即可得到所有变量的联合概率分布,进而利用该联合概率分布进行扩展边缘操作得到  $K(X_q, X_E=e)$ ,  $K(X_E=e)$ ,从而求出  $K(X_q|X_E=e)$ 。由于概率区间的特殊性,在求得每一变量各个配置下的条件概率区间后,需将其作进一步处理,使其变为可行概率区间,这里参照文献[6]的方法。

**引理 1** ( $n$  维概率合理区间)<sup>[6]</sup> 一个  $n$ -PRI( $L, U$ ) 是合理的,

当且仅当  $\sum_{i=1}^n L_i \leq 1, \sum_{i=1}^n U_i \geq 1$ 。

**引理 2** (可行概率区间的唯一性)<sup>[6]</sup> 给定一个合理的  $n$ -PRI( $L, U$ ),通过下述方法

$$\begin{cases} \bar{U}_j = \min(U_j, 1 - \sum_{\substack{i=1 \\ i \neq j}}^n L_i) \\ \bar{L}_j = \max(L_j, 1 - \sum_{\substack{i=1 \\ i \neq j}}^n U_i) \end{cases} \tag{2}$$

$K(a_1)=[0.10, 0.20]$	$K(b_1)=[0.10, 0.20]$	$K(C A)$	$a_1$	$a_2$	$K(F D)$	$a_1$	$a_2$
$K(a_2)=[0.80, 0.90]$	$K(b_2)=[0.80, 0.90]$	$c_1$	[0.10, 0.30]	[0.16, 0.21]	$f_1$	[0.10, 0.30]	[0.16, 0.21]
		$c_2$	[0.20, 0.45]	[0.22, 0.42]	$f_2$	[0.20, 0.45]	[0.22, 0.42]
		$c_3$	[0.40, 0.60]	[0.40, 0.60]			

$K(D AB)$	$a_1 b_1$	$a_1 b_2$	$a_2 b_1$	$a_2 b_2$
$d_1$	[0.10, 0.20]	[0.80, 0.90]	[0.70, 0.80]	[0.30, 0.30]
$d_2$	[0.80, 0.90]	[0.10, 0.20]	[0.20, 0.30]	[0.70, 0.70]

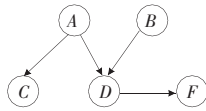


图3 一个 Credal 网络

可导出一个唯一可行的  $(\bar{L}, \bar{U})$ 。

### 3.2 Credal 网络推理实现过程

(1) 将 Credal 网络的每个 CPT 转化为拓展的扩展关系表, 其中的  $f_{\phi R}^U, f_{\phi R}^L$  分别为结点参数的上下概率, 若某结点参数为点概率, 可利用  $a=[a, a]$  转化为区间数。

(2) 对所有表进行扩展乘积连接, 得到所有变量的联合概率分布关系。

(3) 根据查询变量和证据变量, 利用扩展边缘操作求  $K(X_q, X_E=e), K(X_E=e)$ 。

(4) 计算  $\max\{K(X_q, X_E=e)/K(X_E=e)\}$  和  $\min\{K(X_q, X_E=e)/K(X_E=e)\}$ , 从而求得  $K(X_q|X_E=e)$ 。

(5) 对求得的  $K(X_q|X_E=e)$  利用式(2)进行可行性处理。

### 4 算例

利用上述方法对文献[7]中的网络(如图3)进行推理, 假定在  $C=c_1, F=f_1$  的证据下, 求  $K(B|c_1 f_1)$  的后验概率。

首先将5个结点的概率分布表转化为5个扩展关系表, 并将这5个表进行扩展乘积连接, 得到变量的联合概率表, 并进行扩展边缘操作, 得到BCF和CF的边缘关系表, 这里只选择了  $C=c_1, F=f_1$  下的边缘关系表, 利用  $K(B|c_1 f_1)=K(B|c_1 f_1)=K(Bc_1 f_1)/K(c_1 f_1)$  得到所求概率。计算结果如表2所示。

表2 利用扩展关系模型对图3网络推理的结果

B	C	F	$f_{\phi R}^L$	$f_{\phi R}^U$
0	0	0	0.008 022	0.039 396
1	0	0	0.062 224	0.157 626
$c_1 f_1$ 的联合概率	0	0	0.070 246	0.197 022
$K(B c_1 f_1)$	$b_1$		0.114 20	0.199 98
	$b_2$		0.800 02	0.885 80

(上接 36 页)

### 4 结论

主要关注意会知识在机器人体系结构中的应用。从意会知识的表示方式和获取方法两个方面来研究, 给出了基于程序记忆和情节记忆的机器人仿生体系结构。旨在给出一种新的体系结构, 但只是作为现有体系结构的补充, 而非替代原有的体系结构。希望该文提到的思想可以起到抛砖引玉的作用, 使意会知识在机器人领域中的研究能够更加深入。

### 参考文献:

[1] 蒋新松. 机器人学导论[M]. 沈阳: 辽宁科学技术出版社, 1994.  
 [2] Arkin R C. Behavior-based robotics [M]. Cambridge, Massachusetts: MIT Press, 1998.

### 5 结论

文献[7]中利用 Gibbs 采样算法得到的近似后验概率区间为  $p(b_1)=[0.09, 0.24], p(b_2)=[0.76, 0.93]$ , 方法得到的结果概率区间包含在其中, 可见方法与 Gibbs 采样算法的一致性, 方法对上述区间进行了压缩, 信息更准确。采用扩展关系模型进行推理的所有操作全部可用 SQL 语句实现, 简便易行, 同时可利用关系数据库强大的查询优化机制, 提高推理效率。尽管如此, 利用方法得到的结果仍然比较粗糙, 设计更精确的高效推理算法是下一步继续研究的课题。

### 参考文献:

[1] Fagioli E, Zaffalon M. 2U: An exact interval propagation algorithm for polytrees with binary variables[J]. Artificial Intelligence, 1998, 106(1): 77-107.  
 [2] 李玉玲, 吴祈宗, 郑恒, 等. Credal 网络在港口生产安全评价中的应用[J]. 计算机工程与应用, 2007, 43(29): 214-216.  
 [3] Wong S K M. An extended relational model for probabilistic reasoning[J]. Journal of Intelligent Information Systems, 1997, 9(1): 181-202.  
 [4] Wong S K M, Butz C J, Xiang Y. A method for implementing a probabilistic model as a relational database[C]//11th Conference on Uncertainty in Artificial Intelligence (UAI95). Morgan Kaufmann Publishers, 1995: 556-564.  
 [5] 何盈捷, 刘惟一. SQL 实现 Bayesian 网的不确定性推理[J]. 云南大学学报: 自然科学版, 2001, 23(2): 100-103.  
 [6] 王明文. 基于概率区间的决策方法[J]. 系统工程理论与实践, 1997(11): 79-82.  
 [7] 夏元铃, 张自立, 李维华, 等. 带区间参数的贝叶斯网的近似推理[J]. 云南大学学报: 自然科学版, 2007, 30(2): 135-141.

[3] 李恒威, 黄华新. 表征与认知发展[J]. 中国社会科学, 2006(2): 34-44.  
 [4] 刘景钊. 内隐认知与意会知识的深层机制[J]. 自然辩证法研究, 1999, 15: 11-14.  
 [5] Tulving E. Episodic memory: From mind to brain[J]. Annual Reviews of Psychology, 2002, 53: 1-25.  
 [6] Tolman E C. Cognitive maps in rats and men[J]. The Psychological Review, 1948, 55: 189-208.  
 [7] Mataric M, Brooks R. Learning a distributed map representation based on navigation behaviors[M]//Brooks R. Cambrian Intelligence: The Early History of the New AI. Cambridge, Massachusetts: MIT Press, 1990.  
 [8] Konidaris G, Hayes G. An architecture for behavior-based reinforcement learning[J]. Adaptive Behavior, 2005, 13(1): 5-32.  
 [9] Sternberg R J. 认知心理学[M]. 杨炳钧, 陈燕, 译. 北京: 中国轻工业出版社, 2006.