Bayesian 网的独立性推广模型*)

彭青松　张佑生　汪荣贵
(合肥工业大学计算机信息学院　合肥230009)

摘　要　本文提出了Bayesian网的独立性推广模型。Bayesian网能够表示变量之间概率影响关系与条件独立性，但不能表示因果独立性。虽然Noisy OR模型能够较好地表示变量之间的因果独立性，但该模型又因只能表示因果独立性而具有很大的局限性。本文提出的独立性推广模型解决了Bayesian网因果独立性表示能力不足的问题，扩展了Bayesian网与Noisy OR模型的表示范围，同时简化了Bayesian网的条件概率表，并且新模型更能够反映变量之间的概率影响关系。实验结果表明了该模型的实用性。

关键词　Bayesian网，Noisy OR模型，因果独立性，条件独立性

The Extension of the Bayesian Network Based on Independence

PENG Qing-Song　ZHANG You-Sheng　WANG Rong-Gui
(Department of Computer Science and Technology, Hefei University of Technology, Hefei 230009)

Abstract　The Bayesian can express the conditionally independence conveniently, but in appliactions it can't handle causally independence easily. The Noisy OR model can express causally independence well, but the limitation of the Noisy OR model blocks the widely use of itself. In this paper, we present the extension of the Bayesian Network based on independence. Our new model generalizes the Bayesian Networks to handle causally independence properly, and simplifies the conditionally probability table. Experimental results show the availability of our new model.

Keywords　Bayesian networks, Noisy OR model, Conditionally independence, Causally independence

1　Bayesian 网介绍

Bayesian网建立在概率论的基础上，是一种复杂联合概率分布的图形表示方式，能够表示变量之间的条件独立性与因果关系，已应用于故障诊断、语音识别和医疗诊断等方面[1]。这是一种表示变量之间概率影响关系的概率图模型，它符合人们求解问题的习惯。

Bayesian网是通过计算变量之间的联合概率分布来实现推理和决策的。通常用大写字母表示Bayesian网中的各个结点，如记Bayesian网的 n 个结点分别为 $X_1, X_2, \cdots, X_n$。结点的具体取值状态用相应的小写字母表示，例如结点 X_i 的第三个取值记为 $x_{i,3}$。Bayesian中的这 n 个结点的联合概率分布为：

$$p(X_1, X_2, \cdots, X_n) = \prod_{i=1}^{n} p(X_i | X_1, X_2, \cdots, X_{i-1}) \qquad (1)$$

若对于任意的 i，记结点 X_i 的父结点为 $parent(X_i)$，则关于该结点的条件概率为：

$$p(X_i | X_1, X_2, \cdots, X_{i-1}) = p(X_i | parent(X_i)) \qquad (2)$$

因此，(1)式简化为

$$p(X_1, X_2, \cdots, X_n) = \prod_{i=1}^{n} p(X_i | parent(X_i)) \qquad (3)$$

(3)式是根据条件独立性进行简化得到的Bayesian网联合概率分布。应用这种简化方法，可以得到Bayesian中任意变量之间的条件概率。这种方法比根据概率原理来直接进行计算要简洁得多。

例如在图1所示关于学生成绩的Bayesian网[2]中，各变量的名称分别为：智商(X_1)、课程难度(X_2)、应试能力(X_3)、理解教材(X_4)、考试成绩(X_5)与作业成绩(X_6)。要计算当 X_6 给定时 X_1 的条件概率 $p(X_1|X_6)$，根据概率原理得到：

$$p(X_1|X_6) = \frac{p(X_1, X_6)}{p(X_6)} = \frac{\sum_{x_2, x_3, x_4, x_5} p(X_1, X_2, X_3, X_4, X_5, X_6)}{\sum_{x_1, x_2, x_3, x_4, x_5} p(X_1, X_2, X_3, X_4, X_5, X_6)} \qquad (4)$$

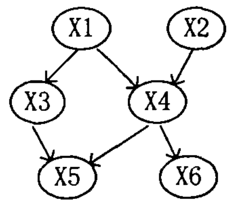


图1　学生成绩的Bayesian网

从理论上来说，通过(4)式可以计算出条件概率 $p(X_1|X_6)$，但其计算量相当大。当结点的个数增加时，需要计算的概率数目呈指数级增长。同时，对Bayesian网中所有变量在各取值下的联合概率进行估计虽然理论上可行，但在具体操作中却是不现实的[3]。根据Bayesian网中的条件独立性，(4)式简化为：

$$p(X_1|X_6) = \frac{\sum_{x_2, x_3, x_4, x_5} p(X_1, X_2, X_3, X_4, X_5, X_6)}{\sum_{x_1, x_2, x_3, x_4, x_5} p(X_1, X_2, X_3, X_4, X_5, X_6)} \qquad (5)$$

$$= \frac{p(X_1) \sum_{x_2} p(X_2) \sum_{x_3} p(X_3|X_1) \sum_{x_4} p(X_4|X_1, X_2) \sum_{x_5} p(X_5|X_3, X_4) p(X_6|X_4)}{\sum_{x_1} p(X_1) \sum_{x_2} p(X_2) \sum_{x_3} p(X_3|X_1) \sum_{x_4} p(X_4|X_1, X_2) \sum_{x_5} p(X_5|X_3, X_4) p(X_6|X_4)}$$

*)基金项目：安徽省自然科学基金(No. 04-03042207)资助课题。

通过对(5)式与(4)式的比较知，根据条件独立性，简化后需要估算的条件概率数目明显减少。

利用 Bayesian 网进行推理、诊断或决策首先要建立合适的 Bayesian 网，而建立 Bayesian 网需要确定 Bayesian 网的结构和概率参数，这是 Bayesian 网应用的前提。而在实际应用中，估算多个原因结点对一个结点的概率影响关系较为复杂，人们更愿意估计两个结点之间的概率影响关系。

本文针对 Bayesian 网的条件独立性和因果独立性，提出了 Bayesian 网的独立性推广模型，该模型扩展了 Bayesian 网对具有因果独立性变量的表示能力，使 Bayesian 网的适用范围更为广泛，并且继承了 Noisy OR 模型的优点。最好通过实例来说明了该扩展模型的实用性。

2 条件独立性与因果独立性

Bayesian 网进行知识表示及推理的优点是它能够表示变量之间的概率影响关系与条件独立性。其中条件独立性的定义[4]为：

定义1(条件独立性) 若 X、Y 和 Z 满足下列条件之一，则称当 Z 给定时，X 和 Y 是条件独立的：

(1)$p(X|Y,Z)=p(X|Z)$，并且 $p(X|Z)\neq 0$，$p(Y|Z)\neq 0$

(2)$p(X|Z)=0$，或 $p(Y|Z)=0$。

例如在图2中，结点 X_1表示有小偷光临，X_2表示有地震发生，X_3表示报警器发出报警声。显然不论 X_1、X_2和 X_3如何取值，都有 $p(X_3|X1)>0$，$p(X_3|X_2)>0$。若报警器是否报警(X_3)是未知的，则有小偷光临(X_1)与有地震发生(X_2)二者相互独立。若已知报警器在报警，则可以推断出有小偷光临或者有地震发生。但如果通过其他途径确认当前有地震发生，则对有小偷光临的信度有所减小，此时 X_1和 X_2是互相条件独立的。即当 X_3给定时，结点 X_1与 X_2条件独立，$p(X_1|X_3)=p(X_1|X_2,X_3)$。

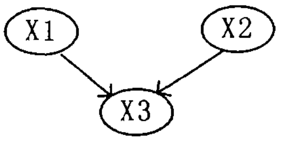


图2 简单 Bayesian 网

将条件独立性用于 Bayesian 网，可以降低其联合概率的复杂程度。但简化后的联合概率仍然需要对每个结点指定其关于所有父结点的条件概率，条件概率的估计仍然非常困难。在实际应用中，而仅估计两两结点之间的条件概率要容易得多。并且一些父结点可以单独影响结果的发生，此时客观上不需要估计当前结点关于所有父结点的条件概率。因此，估计两个结点之间的条件概率也是具体应用的需要。

仍然以图2为例，二值结点 X_1(有小偷光临)与 X_2(有地震发生)均可以单独对二值结点 X_3(报警器发出报警声)产生影响。此时只需要估算条件概率 $p(X_3|X_1)$与 $p(X_3|X_2)$，而不需要估计联合概率分布 $p(X_3|X_1,X_2)$。此时称结点 X_1与 X_2是互为因果独立的。

定义2(因果独立性) 在 Noisy OR 模型中，若原因变量 $X_1,X_2,\cdots,X_n$ 都可以单独对结点 E 发生概率影响关系，则称这 n 个结点是互相因果独立的。

在这种情况下，图2称为 Noisy OR 模型。Noisy OR 模型中结点的各个父结点之间都是因果独立的。在图2的模型中，结点 X_3发生的条件概率为：

$$p(X_3|X_1,X_2)=1-\prod_{i=1,2}(1-p(X_3|X_i)) \tag{6}$$

3 独立性推广模型

Bayesian 网是表示结点之间概率影响关系的图模型，它能够把变量之间的这种概率影响关系以图模型的形式表示出来。但 Bayesian 网的精确推理是 NP 难题[5]，并且其近似推理也是 NP 难题[6]。Noisy OR 模型是经过简化的 Bayesian 网络，其同一结点的各个父结点之间是因果独立的关系。根据 Bayesian 网与 Noisy OR 模型的共同特性，得出独立性推广模型如下：

定义3(独立性推广模型) 在具有 n 个结点 $X=\{X_1,X_2,\cdots,X_n\}$的 Bayesian 网中，若其中任一个结点 X_i 的 m 组相互因果独立的父结点集合为 $parent(X_i)=\{X_{i1},X_{i2},\cdots,X_{im}\}$，$\forall\ 1\leqslant j\leqslant m$，$X_{ij}$包含结点 $X_{ij_1},X_{ij_2},\cdots,X_{ij_k}$，即 $X_{ij}=\{X_{ij1},X_{ij2},\cdots,X_{ijk}\}$。$\forall\ 1\leqslant j,j'\leqslant m,j\neq j'$，有 $X_{ij}\cap X_{ij'}=\Phi$。在这 m 组父结点与子结点之间分别插入中间结点，共需要加入 m 个中间结点，分别为 $I_{i1},I_{i2},\cdots,I_{im}$。新插入的中间结点与子结点 X_i 的取值范围相同，并且互相之间因果独立。对所有结点都进行扩展后得到的概率图模型称为 Bayesian 网的独立性推广模型。

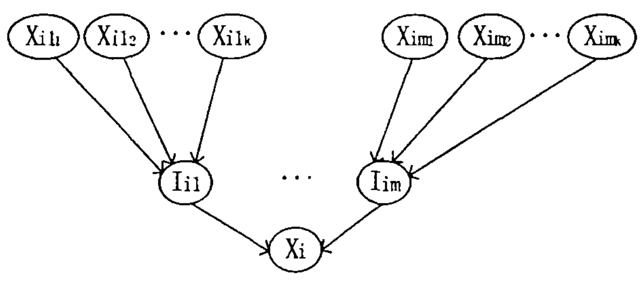


图3 学生成绩的 Bayesian 网推广模型的一部分

图3为 Bayesian 网的独立性推广模型的一部分，为关于结点 X_i 与其 m 组父结点 $parent(X_i)=\{X_{i1},X_{i2},\cdots,X_{im}\}$加入中间结点 $I_{i1},I_{i2},\cdots,I_{im}$以后的情况。并且中间结点 $I_{i1},I_{i2},\cdots,I_{im}$与 X_i 具有相同的取值状态。若结点集合 $X=\{X_1,X_2,\cdots,X_n\}$中的所有结点都是二值结点，则该 Bayesian 网独立性推广模型的联合概率分布为：

$$\begin{aligned}p(X_1,X_2,\cdots,X_n)&=\prod_{i=1}^{n}p(X_i|parent(X_i))=\prod_{i=1}^{n}p(X_i|X_{i1},X_{i2},\cdots,X_{im})=\prod_{i=1}^{n}p(X_i|I_{i1},I_{i2},\cdots,I_{im})=\prod_{i=1}^{n}\\&(1-\prod_{j=1}^{m}(1-p(X_i|I_{ij}))=\prod_{i=1}^{n}(1-\prod_{j=1}^{m}(1-p(X_i|X_{ij1},X_{ij2},\cdots,X_{ijk}))=\prod_{i=1}^{n}(1-\prod_{j=1}^{m}(1-p(X_i|parent(I_{ij}))))\end{aligned} \tag{7}$$

其中，I_{ij}为在扩展模型中结点 X_i 的第 j 个中间结点，$parent(\cdot)$表示当前结点的父结点集合。

若在结点 $X=\{X_1,X_2,\cdots,X_n\}$中，有部分结点是多值结点，则独立性推广模型的联合概率分布为：

$$\begin{aligned}p(X_1,X_2,\cdots,X_n)&=\prod_{i=1}^{n}p(X_i|I_{i1},I_{i2},\cdots,I_{im})=\\&\prod_{i=1}^{n}\sum_{\substack{k1k2,\cdots,km=1;\\ \max(k1,k2,\cdots,km)=k0}}^{k_i}p(x_{i,k1}|I_{i1})p(x_{i,k2}|I_{i2})\cdots p(x_{i,km}|I_{im})=\prod_{i=1}^{n}\sum_{\max(k1,k2,\cdots,km)=k0}^{k_i}\prod_{kj=1}^{k}p(x_{i,kj}|I_{ij})=\end{aligned}$$

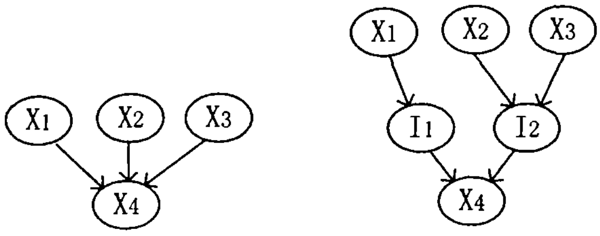
$$\prod_{i=1}^{n} \sum_{\max(k1,k2,\cdots,km)=k0}^{k_i} \prod_{j=1}^{k} p(x_{i,k_j} | parent(I_{ij})) \tag{8}$$

其中，结点 X_i 共有 k_i 个取值状态，分别为 $x_{i,1}$、$x_{i,2}$、…、$x_{i,k}$，并且用 x_{k0} 表示结点 X_i 的当前取值状态，I_{ij} 为扩展模型中 X_i 的第 j 个中间结点，$parent(\cdot)$ 表示当前结点的父结点。

4 应用实例

4.1 示例实验一

图4为一关于考试成绩的 Bayesian 网，结点 X_1 表示突然发高烧，X_2 表示理解教材，X_3 表示应试能力，X_4 表示考试通过，则当 X_4 给定时，X_2 与 X_3 互相条件独立，而 X_1 与 $\{X_2,X_3\}$ 之间互相因果独立，因此该 Bayesian 网由图4(a)扩展为图4(b)更为合理。图4(b)中加入了两个中间结点 I_1 与 I_2，并且这两个中间结点与 X_4 的取值状态完全相同。



(a) (b)

图5 简单 Bayesian 网及其独立性扩展模型

模型扩展前，图4(a)的 Bayesian 网需要的计算条件概率为 $p(X_4|X_1,X_2,X_3)$。而扩展后图4(b)需要的条件概率为 $p(X_4|X_1)$ 与 $p(X_4|X_2,X_3)$。条件概率的复杂性得到了简化，并且需要计算的条件概率的数目由2×2×2=8减少到2+2×2=6。根据式(7)知，结点 X_4 发生的条件概率为：

$$p(X_4|X_1,X_2,X_3)=1-\prod_{i=1,2}(1-p(X_4|I_i))=1-\prod_{i=1,2}(1-p(X_4|parent(I_i)))=1-(1-p(X_4|X_1))\times(1-p(X_4|X_2,X_3)) \tag{9}$$

图4(b)的扩展模型与扩展前图4(a)的模型相比，减少了需要计算的条件概率的数目，并且降低了条件概率本身的复杂性。该扩展模型符合人们的求解习惯。

4.2 示例实验二

在图5(a)中，各结点与图1中结点的意义相同，条件概率表分别如表1至表6所示。

表1 结点 X1的概率表

X1	X1=智商高	X1=智商一般
P(x1)	0.3	0.7

表2 结点 X2的概率表

X2	X2=非常难	X2=难	X3=一般
P(x2)	0.3	0.4	0.3

表3 结点 X3关于 X1的条件概率表

	X1=智商高	X1=智商一般
X3=应能力强	P(x3|x1)=0.7	P(x3|x1)=0.3
X3=应试图弱	P(x3|x1)=0.35	P(x3|x1)=0.65

表4 结点 X4关于 X1与 X2的条件概率表

	X1=智商高			X1=智商一般		
	X2=非常难	X2=难	X3=一般	X2=非常难	X2=难	X3=一般
X4=理解教材	P(x4|x1,x2)=0.55	P(x4|x1,x2)=0.75	P(x4|x1,x2)=0.95	P(x4|x1,x2)=0.1	P(x4|x1,x2)=0.4	P(x4|x1,x2)=0.5
X4=不理解教材	P(x4|x1,x2)=0.45	P(x4|x1,x2)=0.25	P(x4|x1,x2)=0.05	P(x4|x1,x2)=0.9	P(x4|x1,x2)=0.7	P(x4|x1,x2)=0.5

表5 结点 X5关于 X3与 X4的条件概率表

	X3=应能力强		X3=应能力弱	
	X4=理解教材	X4=不理解教材	X4=理解教材	X4=不理解教材
X5=A	P(x5|x3,x4)=0.5	P(x5|x3,x4)=0.2	P(x5|x3,x4)=0.2	P(x5|x3,x4)=0.05
X5=B	P(x5|x3,x4)=0.3	P(x5|x3,x4)=0.3	P(x5|x3,x4)=0.4	P(x5|x3,x4)=0.1
X5=C	P(x5|x3,x4)=0.1	P(x5|x3,x4)=0.3	P(x5|x3,x4)=0.3	P(x5|x3,x4)=0.15
X5=D	P(x5|x3,x4)=0.05	P(x5|x3,x4)=0.15	P(x5|x3,x4)=0.05	P(x5|x3,x4)=0.35
X5=E	P(x5|x3,x4)=0.05	P(x5|x3,x4)=0.05	P(x5|x3,x4)=0.05	P(x5|x3,x4)=0.35

表6 结点 X6关于 X4的条件概率表

	X4=理解教材	X4=不理解教材
X6=A	P(x6|x4)=0.4	P(x6|x4)=0.05
X6=B	P(x6|x4)=0.3	P(x6|x4)=0.05
X6=C	P(x6|x4)=0.2	P(x6|x4)=0.2
X6=D	P(x6|x4)=0.05	P(x6|x4)=0.3
X6=E	P(x6|x4)=0.05	P(x6|x4)=0.4

在图5(b)中加入结点 X_7(临考发高烧)以后，结点 X_7 关于 X_5 的条件概率如表7所示。

表7 结点 X7关于 X5的条件概率表

	X7=临考发高烧	X7=临考不发烧
X5=A	P(x5|x7)=0.025	P(x5|x7)=0.2
X5=B	P(x5|x7)=0.025	P(x5|x7)=0.2
X5=C	P(x5|x7)=0.025	P(x5|x7)=0.2
X5=D	P(x5|x7)=0.025	P(x5|x7)=0.2
X5=E	P(x5|x7)=0.9	P(x5|x7)=0.2

在图5(b)中新加入的中间结点 I_1 与 I_2，分别对应结点集合 $\{X_1,X_2,\cdots,X_6\}$ 与 $\{X_7\}$，二者之间是互相因果独立的关系。在其概率表中可以看出，当 $\{X_7$=临考发高烧$\}$ 时，考试成绩为 E 的概率变得非常大。而当 $\{X_7$=临考不发烧$\}$ 时，考试成绩的各个条件概率值取值相同，即不发生影响。

表8 结点 X5关于 X1与 X2的条件概率表

	X1=智商高 X2=课程难度一般	X1=智商一般 X2=课程非常难	X1=智商高 X2=课程非常难
P(X5=A|X1,X2)	0.3972	0.1228	0.2952
P(X5=B|X1,X2)	0.3255	0.1895	0.2895
P(X5=C|X1,X2)	0.1648	0.2052	0.2027
P(X5=D|X1,X2)	0.0580	0.2570	0.1220
P(X5=E|X1,X2)	0.0545	0.2255	0.0905

(下转第223页)

2000, 9(3): 390～404

8 Choi H, Baraniuk R G. Multiscale image segmentation using wavelet-domain hidden Markov models. IEEE Trans Image Processing, 2001, 10(9): 1309～1321

9 Luettgen M R, Karl W C, Willsky A S, Tenney R R. Multiscale representations of Markov random fields. IEEE Trans Signal Processing, 1993, 41(12): 3377～3396

10 Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm. J R Statist Soc B, 1977, 39(1): 1～38

11 Rissanen J. A universal prior for integers and estimation by minimum description length. Ann Statist, 1983, 11(9): 417～431

（上接第184页）

图5(b)中的两个中间结点 I_1 与 I_2 只是辅助结点，起推理的作用，并无实际意义。假定它的取值范围都同结点 X_5。在图5(a)中，在一些关于结点 X_1 与 X_2 取值的证据下，结点 X_5 的条件概率如表8所示。

在图5(b)中加入了新的结点 X_7 和两个中间结点 I_1，I_2，并且有 $P(I_2|X_1,X_2)=P(X_5|X_1,X_2)$。当 $\{X_7=$临考不发烧$\}$ 时，由于它单独对考试成绩不发生影响，因此本实验中只考虑 $\{X_7=$临考发高烧$\}$ 的情况，由此可以得到条件概率表分别如表9、表10所示。

表9 中间结点I2关于X1与X2，中间结点I1关于X7的条件概率表

X1＝智商高 X2＝课程难度一般	X1＝智商一般 X2＝课程非常难	X1＝智商高 X2＝课程非常难	X7＝临场发高烧
P(I2＝A|X1,X2)＝ 0.3972	P(I2＝A|X1,X2)＝ 0.1228	P(I2＝A|X1,X2)＝ 0.2952	P(I1＝A|x7)＝0.025
P(I2＝B|X1,X2)＝ 0.3255	P(I2＝B|X1,X2)＝ 0.1895	P(I2＝B|X1,X2)＝ 0.2895	P(I1＝B|x7)＝0.025
P(I2＝C|X1,X2)＝ 0.1648	P(I2＝C|X1,X2)＝ 0.2052	P(I2＝C|X1,X2)＝ 0.2027	P(I1＝C|x7)＝0.025
P(I2＝D|X1,X2)＝ 0.0580	P(I2＝D|X1,X2)＝ 0.2570	P(I2＝D|X1,X2)＝ 0.1220	P(I1＝D|x7)＝0.025
P(I2＝E|X1,X2)＝0.0545	P(I2＝E|X1,X2)＝ 0.2255	P(I2＝E|X1,X2)＝ 0.0905	P(I1＝E|x7)＝0.9

表10 扩展模型中结点X5的条件概率表

	X1＝智商高 X2＝课程难度一般 X7＝临场发高烧	X1＝智商一般 X2＝课程非常难 X7＝临场发高烧	X1＝智商高 X2＝课程非常难 X7＝临场发高烧
P(X5＝A|I1,I2)	0.0099	0.0031	0.0074
P(X5＝B|I1,I2)	0.0262	0.0125	0.0219
P(X5＝C|I1,I2)	0.0304	0.0232	0.0298
P(X5＝D|I1,I2)	0.0280	0.0386	0.0319
P(X5＝E|I1,I2)	0.9236	0.9194	0.9227

表10为根据式(8)进行计算得到的关于 X_5 的条件概率，由该表可以看出，当增加的新结点取值为 $\{X_7=$临考发高烧$\}$ 以后，考试成绩为E的条件概率由0.0545，0.2255，0.0905分别上升到了0.9236，0.9194，0.9227。

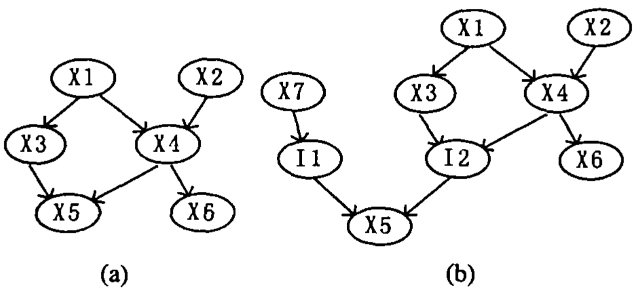


图5 学生成绩的Bayesian网及其扩展

结论 本文中，根据Bayesian网中变量之间的条件独立性与Noisy OR模型中的因果独立性，提出了同时包含条件独立性与因果独立性的推广模型。该推广模型把二者的推理方法予以融合，并且扩展了Bayesian网的应用范围，同时弥补了Noisy OR模型实用性不足的缺点，并且新模型所需的条件概率参数小于Bayesian网中的参数，增加了Bayesian网络构建的效率，降低了推理的复杂度。实验结果表明，该推广模型具有较好的应用前景。

参考文献

1 Santos E, Shimony S E. Deterministic Approximation of Marginal Probabilities in Bayes Nets. IEEE Transactions on Systems, Man, and Cybernetics-Part A: System and Humans, 1998, 28(4): 377～393

2 Friedman N, Getoor L, Koller D, et al. Learning Probabilistic Relational Models. In: Proc. of the 16th Intl. Joint Conf. on Artificial Intelligence (IJCAI), Stockholm, Sweden, 1999

3 Heckerman D, Breese J. Causal Independence for Probability Assessment and Inference Using Bayesian Networks: [Technical Report MSR-TR-94-08]. Microsoft Research, 1994

4 Neapolitan R E. Learning Bayesian Networks. Prentice Hall, Prentice Hall, Upper Saddle River, NJ, 2003

5 Cooper G. Probabilistic inference using belief networks is NP-hard. Artificial Intelligence, 1990, 42: 393～405

6 Dagum P, Luby M. Aproximate probabilistic inference in Bayesian networks in NP hard. Artificial Intelligence, 1993, 60: 141～153