维普资讯 http://www.cqvip.com

计算机科学1999Vol.26№.10

一类贝叶斯网络的线性推理*)

The Linear Reasoning of Bayesian Networks

覃秋梅
(广东省珠海市第三中学电脑室　珠海519000)
(清华大学智能技术与系统国家重点实验室　北京100084)
张师超
(新加坡国立大学信息系统与计算机科学技术系)

Abstract Uncertain reasoning is one of the key technologies in intelligent systems and diagnosis systems. Probably one of the most popular models for uncertain reasoning is Bayesian networks, but the computational complexity probabilistic inference using belief networks is NP-hard. In view of it, two theorems are given in this paper. We substitute a linear function for the conditional probability matrix to propagate probabilities in Bayesian networks so that the reasoning speed is improved.

Keywords Uncertain reasoning, Bayesian network, Encoding technology, Linear reasoning.

贝叶斯网络提供了表示变量集之间概率依赖性的一个自然有效的方法，而且其推理方法是主观贝叶斯方法的一个扩展，具有坚实的概率理论基础，因此，许多人工智能的研究者都采用贝叶斯网络作为一种知识表示的方法，将其应用到各种问题领域。如：故事理解[1]、规划[2]、电路错误检测[7]和医学诊断[7]等等。但是，贝叶斯网络已遭受到一些人工智能研究者的批评，因为它们需要大量的数值概率值使不确定关系量化，不管这些概率是否直接从数值中估计，还是作为主观概率由领域专家给出，或者是两者的结合。另一方面，贝叶斯网络的计算常常是NP难的[3~5]。

贝叶斯网络有两种计算[7]：信度(Belief)更新和信度修正。前者是根据新证据计算各节点更新后的信度(新知识在网络中的传播)。信度修正是计算出满足新证据的各节点的取值情况(寻找最合理的解释)。对这两种计算，最早的方法是Pearl提出的信息传递方法[7]。但这种方法局限于单联通网络中，即网络中任两个节点之间至多存在一条有向的通路。而且，它的计算复杂度是NP难的[3]。最近，Santos提出了一种用线性规划的思想来处理信度修正[8,9]，对节点的每一种取值都进行编码，运用一种称之为线性位势函数(Linear Potential Functions)的估计函数来压缩概率信息。实践证明，这种方法在信度修正的处理上比Pearl的方法有效，而且对网络的拓扑结构没有任何限制。对于信度更新，我们在文[6]中采用编码技术将统计技术与概率技术联系起来，建立了一种近似函数来传播概率，从而提高了其推理速度。

可是，文[6]中主要采用统计技术，缺乏一定的理论基础，为此，我们在文[6]的基础上，总结归纳出完全可用线性函数来进行推理的一类贝叶斯网络，由此得出两个定理，从而对文[6]给出了部分理论根据。

1. Judea Pearl的信息传递方法

贝叶斯网络由一系列表示因果关系的规则结合而成，对于因果关系X→Y，X的取值范围为$\{X_1, X_2, \cdots, X_m\}$，Y的取值范围为$\{Y_1, Y_2, \cdots, Y_n\}$，贝叶斯网络采用条件概率矩阵M来确定X与Y之间的因果关系。

$$M_{Y|X}=P(y|x)=\begin{bmatrix} p(Y_1|X_1) & p(Y_2|X_1) & \cdots & p(Y_n|X_1) \\ p(Y_1|X_2) & p(Y_2|X_2) & \cdots & p(Y_n|X_2) \\ \vdots & & & \\ p(Y_1|X_m) & p(Y_2|X_m) & \cdots & p(Y_n|X_m) \end{bmatrix}$$

其中$p(Y_j|X_i)$表示X取值X_i时，Y取值Y_j的概率。

为了表示不断获得的新证据和新事实对每一个变量可能取值的影响，对每一个节点A定义一个信度分布，记为BEL(a)，即：

$$BEL(a)\triangleq P(A=a|e)$$

表示在当前所拥有的所有事实和证据e条件下，随机

*)本文获清华大学智能技术与系统国家重点实验室开发基金资助。

维普资讯 http://www.cqvip.com

变量 A 取值 a 的可能程度。

考虑图1的链型网络，假设在链的头部和尾部都获得证据。即 $e=\{e^+Ue^-\}$，则：

$$BEL(x)=p(x|e^+,e^-)=\frac{p(x,e^+,e^-)}{p(e^+,e^-)}$$
$$=\frac{p(e^-|x,e^+)p(x|e^+)p(e^+)}{p(e^+,e^-)}$$
$$=\alpha p(e^-|x)p(x|e^+) \quad (1)$$

$e^+\rightarrow U\rightarrow X\rightarrow Y\rightarrow e^-$

图1

令 $\lambda(X)=p(e^-|x)$ 代表由症状推断病因；$\pi(X)=p(x|e^+)$ 代表由病因推断症状；那么，(1)式可写为：

$$BEL(X)=\alpha\,\lambda(X)\pi(X)$$

其中 α 为标准化常数(也称归一化常数)。其目的是使 BEL(X)满足概率意义。

对于图(1)，证据 e^+ 直接作用于 U 产生 $\pi(u)$，然后通过 U→X，将 π 信息传给 X，使得：

$$\pi(x)=p(x|e^+)$$
$$=\sum_u p(x|u,e^+)p(u|e^+)$$
$$=\sum_u p(x|u)p(u|e^+)=\pi(u)\cdot M_{y|x}$$

同理，证据 e^- 直接作用于 Y，产生 $\lambda(y)$，然后通过 X→Y 将 λ 信息传给 X，使得

$$\lambda(X)=p(e^-|x)$$
$$=\sum_y p(e^-|y,x)p(y|x)$$
$$=\sum_y p(e^-|y)p(y|x)=M_{y|x}\cdot\lambda(y)$$

对于一般的网络，$\pi-\lambda$ 信息的传播与上类似。所以，无论新到达的证据或者信息对网络中的哪一个节点直接产生影响，改变了该节点的信度分布，都可以通过网络中有向边的联系传到网络中的其它节点，每一个节点在更新其信度分布时，应同时考虑由其双亲节点传来的 π 信息以及由其子女节点传来的 λ 信息。然后再将更新后的 $\pi-\lambda$ 信息传给其子女节点和双亲节点。它是一种双向传播。

2. 编码

贝叶斯网络是一种模拟人类推理过程中因果关系的有向无环图，它的每一个节点是一个可取多个值的变量，节点间的关系由其间连线表示，若节点 M 的值对节点 N 之值有直接影响($M\rightarrow N$)，则两者之间直接连线，连线强度用条件概率 P(n|m)表示，在贝叶斯网络中，信息的传递是基于一个向量与一个矩阵相乘的模型，对于一个规则 X→Y，令 X 的所有点值为 $(X_1,X_2,\cdots,X_m)$，Y 的所有点值为 $(Y_1,Y_2,\cdots,Y_n)$，且 $(x_1,x_2,\cdots,x_m)\in S(X)$。(S(X)为 X 的状态空间)是 X 的观察值，则 Y 的状态 $(y_1,y_2,\cdots,y_n)$ 可通过观察值 $(x_1,x_2,\cdots x_m)$ 和 $M_{y|x}$ 决定，即

$$(x_1,x_2,\cdots,x_m)\cdot M_{y|x}=(y_1,y_2,\cdots,y_n)$$

或者 $$x\cdot M_{y|x}=y \quad (2)$$

对于(1)式，我们的目标是构造一个形如 $F(x)=k_1E_x(x)+k_0$ 的函数，使得

$$k_2E_x(x)+k_0=E_Y(y)$$

即 $$k_1E_x(x)+k_0=E_Y(xM_{y|x}) \quad (3)$$

对于(3)式，需要确定常数 k_1 和 k_0 以及映射(编码)$E_x(x)$ 和 $E_y(y)$。首先定义编码 E_x 为

$$E_x(x_1,x_2,\cdots,x_m)=10^dx_1+10^{2d}x_2+\cdots+10^{md}x_m$$

这里，d 是大于或等于 X 和 Y 中点值概率值小数点后保留的位数的一个整数。例如，令 $m=n=2$，$p(X_1)=0.23$，$p(X_2)=0.77$，$p(Y_1)=0.234$，$p(Y_2)=0.766$，则 d 大于等于3。

E_Y 的编码方法与 E_X 的编码一样。常量 k_1 和 k_0 可采用统计技术求出[6]。从而我们可用下面函数来表示 X 与 Y 之间的因果关系：

$$F(x)=K_1E_X(x)+k_0$$

每当获得 X 的一个状态值 x，则用上面的近似函数求得结果 F(x)后，可以根据它求解 $y=(y_1,y_2,\cdots y_n)$，即

$$y_i=(INT(F(x)/10^{(i-1)d}-INT(F(x)/10^{id})10^d)/10^d(i=1,2,\cdots,n)$$

这里，INT()为取整函数。为了保证概率意义，最终的结果应是

$$y_i=Max(0.1-\sum_{t=i+1}^{n}y_t)(i=1,2,\cdots,n-1)$$

3. 两值节点中的线性推理

在贝叶斯网络中，考虑的节点变量通常是只取两个值(即只取"是"或"非")，我们称这样的节点为两值节点。对于规则 X→Y，当 X 与 Y 均为两值节点时，它们之间的条件概率矩阵完全可以用一个线性函数来代替。设 $a=(a_1,a_2)$ 为 X 的任一个状态，由 Pearl 的方法得出 Y 的状态 $b=a\cdot M_{y|x}$ 则采用文[6]中的编码方法，一定存在：

$$k_1E_X(a)+k_0=E_Y(b) \quad (4)$$

其中，$k_1=m_{11}-m_{21}$　$k_0=m_{21}\cdot10^d+m_{12}\cdot10^{2d}$。

证明：设 X 与 Y 之间的条件概率矩阵为：

$$M_{y|x}=\begin{bmatrix}m_{11} & m_{12}\\ m_{21} & m_{22}\end{bmatrix}$$

则 Y 在 X 处于 (a_1,a_2) 状态下的状态为 (b_1,b_2)，且

$$(b_1,b_2)=(a_1,a_2)\cdot\begin{bmatrix}m_{11} & m_{12}\\ m_{21} & m_{22}\end{bmatrix}$$
$$=[(a_1\cdot m_{11}+a_2\cdot m_{21}),(a_1\cdot m_{12}+a_2\cdot m_{22})]$$

即 $b_1=a_1\cdot m_{11}+a_2\cdot m_{21}$，

维普资讯 http://www.cqvip.com

$b_2=a_1\cdot m_{12}+a_2\cdot m_{22}$

因为 $a_1+a_2=1$，所以

$b_2=a_1\cdot m_{11}+(1-a_1)\cdot m_{21}$

$=a_1\cdot(m_{11}-m_{21})+m_{21}$

$b_2=a_1\cdot m_{12}+a_2\cdot m_{22}$

$=(1-a_2)\cdot m_{21}+a_2\cdot m_{22}$

$=m_{12}+a_2\cdot(m_{22}-m_{12})$

而 $m_{11}=P(y=Y_1|x=X_1)$

$m_{12}=P(y=Y_2|x=X_1)$

$m_{21}=P(y=Y_1|x=X_2)$

$m_{22}=P(y=Y_2|x=X_2)$

故有 $m_{11}-m_{12}=1, m_{21}+m_{22}=1$

从而 $b_2=m_{12}+a\cdot(m_{22}-m_{12})$

$=m_{12}+a_2\cdot(1-m_{21}-1+m_{11})$

$=m_{12}+a_2\cdot(m_{11}-m_{21})$

运用2.3节中的编码方法，有：

$E_X(a)=a_1\cdot 10^d+a_2\cdot 10^{2d}$

$E_Y(b)=b_1\cdot 10^d+b_2\cdot 10^{2d}$

$=[a_2\cdot(m_{11}-m_{21})+m_{21}]\cdot 10^d+[m_{12}+a_2\cdot(m_{11}-m_{21})]\cdot 10^{2d}$

$=a_1\cdot(m_{11}-m_{21})\cdot 10^d+m_{21}\cdot 10^d+m_{12}\cdot 10^{2d}+a_2\cdot(m_{11}-m_{21})\cdot 10^{2d}$

$=(m_{11}-m_{21})\cdot(a_1\cdot 10^d+a_2\cdot 10^{2d})+m_{21}\cdot 10^d+m_{12}\cdot 10^{2d}$

令 $k_1=m_{11}-m_{21}$ $k_0=m_{21}\cdot 10^d+m_{12}\cdot 10^{2d}$

则有 $k_1E_X(a)+k_0=E_Y(b)$

由此，我们得到如下定理：

定理1 对于贝叶斯网络中的两个节点 X，Y，若 X 是 Y 的唯一双亲节点，且 X，Y 是两值节点，则 X 与 Y 之间的条件概率矩阵可用一个线性函数来代替用于传播概率。

4. 其它贝叶斯网络的线性推理

一般地，对于一个规则 X→Y，令 X 的所有点值为 $(X_1,X_2,\cdots,X_m)$，Y 的所有点值为 $(Y_1,Y_2,\cdots,Y_n)$，若 m=n 且 X 与 Y 的条件概率矩阵：

$$M_{Y|X}=\begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ & \vdots & & \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix}$$

满足条件：(1)$m_{11}=m_{22}=\cdots=m_{nn}$；(2)除对角线上的数外，其余的 $m_{ij}(i\neq j)$ 均相等；则对于 X 的任一个状态 a，Y 的状态 $b=a\cdot M_{Y|X}$，采用文[6]的编码方法，a 与 b 之间有如下关系：

$$k_1E_X(a)+k_0=E_Y(b) \qquad (5)$$

其中 $k_1=(m_{11}-m_{21})$，$k_0=m_{21}\cdot(10^d+10^{2d}+\cdots+10^{nd})$，

证明：对于 X 的任一个状态 $a=(a_1,a_2,\cdots,a_n)$，Y 在 X 处于 a 状态下的状态为 $b=(b_1,b_2,\cdots,b_n)$，且有：$(b_1,b_2,\cdots,b_n)=$

$$(a_1,a_2,\cdots,a_n)\cdot\begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ & \vdots & & \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix}$$

即 $b_i=\sum_{j=1}^{n}a_j\cdot m_{ji}$

对 b 进行编码，得

$E_Y(b)=10^db_1+10^{2d}b_2+\cdots+10^{nd}b_n$

由于 $b_i=\sum_{j=1}^{n}a_j\cdot m_{ji}(i=1,\cdots,n)$，所以上式变为

$E_Y(b)=10^d(a_1\cdot m_{11}+a_2\cdot m_{21}+\cdots+a_n\cdot m_{n1})+10^{2d}(a_1\cdot m_{12}+a_2\cdot m_{22}+\cdots+a_n\cdot m_{n2})+\cdots+10^{nd}(a_1\cdot m_{1n}+a_2\cdot m_{2n}+\cdots+a_n\cdot m_{nn})$

因为 $m_{11}=m_{22}=\cdots=m_{nn}$，所以有

$E_Y(b)=(10^d+10^{2d}+\cdots+10^{nd})\cdot m_{11}+10^d\cdot(a_2\cdot m_{21}+\cdots+a_n\cdot m_{n1})+10^{2d}(a_1\cdot m_{12}+a_2\cdot m_{22}+\cdots+a_n\cdot m_{n2})+\cdots+10^{nd}(a_1\cdot m_{1n}+a_2\cdot m_{2n}+\cdots+a_{n-1}\cdot m_{(n-1)n})$

其余的 m_{ij} 均相等，所以有

$E_Y(b)=(10^d+10^{2d}+\cdots+10^{nd})\cdot m_{12}+10^d\cdot m_{21}\cdot(a_2+\cdots+a_n)+10^{2d}\cdot m_{21}\cdot(a_1+a_2+\cdots+a_n)+\cdots+10^{nd}\cdot m_{1n}\cdot(a_1+a_2+\cdots+a_{n-1})$

因为 $\sum_{i=1}^{n}a_i=1$，所以

$E_Y(b)=(10^d+10^{2d}+\cdots+10^{nd})\cdot m_{11}+10^d\cdot m_{21}\cdot(1-a_1)+10^{2d}\cdot m_{21}\cdot(1-a_2)+\cdots+10^{nd}\cdot m_{1n}\cdot(1-a_n)$

所以

$E_Y(b)=(m_{11}-m_{21})(10^da_1+10^{2d}a_2+\cdots+10^{nd}a_n)+m_{21}\cdot(10^d+10^{2d}+\cdots+10^{nd})$

所以

$E_Y(b)=(m_{11}-m_{21})\cdot E_X(a)+m_{21}\cdot(10^d+10^{2d}+\cdots+10^{nd})$

令 $K_1=(m_{11}-m_{21})$，$K_0=m_{21}\cdot(10^d+10^{2d}+\cdots+10^{nd})$，则有：

$k_1E_X(a)+k_0=E_Y(b)$

定理2 对于贝叶斯网络中的两个节点 X，Y，若 X 是 Y 的唯一双亲节点，且 X，Y 是 n 值节点，如果 X 与 Y 的条件概率矩阵：

维普资讯 http://www.cqvip.com

$$M_{Y|X}=\begin{bmatrix} m_{11} & m_{12} & \cdots & m_{1n} \\ m_{21} & m_{22} & \cdots & m_{2n} \\ & \vdots & & \\ m_{n1} & m_{n2} & \cdots & m_{nn} \end{bmatrix}$$

满足条件：(1)$m_{11}=m_{22}=\cdots=m_{nn}$；(2)除对角线上的数外，其余的$m_{ij}(i\neq j)$均相等；则X与Y之间的条件概率矩阵可用线性函数式(5)来代替用于传播概率。

应用实例与结论 下面，我们通过一个实例来进一步说明我们的方法与Pearl的方法的一致性。

例1 对于规则X→Y，X与Y之间的条件概率矩阵为：

$$M_{Y|X}=\begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

因为$M_{Y|X}$中对角线元素都是0.8，其余的元素也都等于0.1，满足定理2的条件。我们取d=2，则由(5)式得X与Y之间的线性函数为

$$E_Y(b)=0.7\cdot E_X(a)+101010 \qquad (6)$$

已知一个观察值为a=(0.3,0.5,0.2)，运用Pearl的方法得

$$b=(0.3,0.5,0.2)\cdot\begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$$

$$=(0.31,0.45,0.24)$$

将a的编码$E_X(a)$代入(6)式，得

$$E_Y(b)=244531$$

所以b=(0.31,0.45,0.24)，与Pearl的方法得出同样的结果。

由上面的计算，我们看到，Pearl的方法是一个向量与一个矩阵相乘，乘法计算次数是m·n次(其中m是矩阵的行数，n是矩阵的列数)。而我们的方法却只需要做一次乘法即可，且其计算结果是完全一致的，显然，用我们的方法简单，推理速度快。

结论 在贝叶斯网络中，有一部分网络是可以运用我们的线性推理方法的，通过编码技术，还可将由此算出的结果用在贝叶斯网络中进一步进行推理，这样就可提高贝叶斯网络的推理速度。但是，我们也看到，它只适用于一小部分网络，当然，这仅仅是我们工作的开始，在今后的研究中，我们将致力于研究一种新的编码方法，从而使我们的方法能运用到大部分网络中，或者建立一种较理想的方法来构造变量之间因果关系的函数，从而建立一个新的模型完全替代贝叶斯网络。

参考文献

1 Charniak E, Goldman R. A semantics for probabilistic quantifier-free first-order languages, with particular application to story understanding. In: Proc. of the IJCAI conf.,1989

2 Charniak E, Goldman R. A probabilistic model of plan recognition. In:Proc. of the AAAI Conf.,1991

3 Cooper G. *Probabilistic inference using belief networks is* NP-hard:[Technical Report KSL877-27]. Medical Computer Science Group, Stanford University, Stanford,1987

4 Cooper G. The computational complexity probabilistic inference using belief networks. Artificial Intelligence, 1990(42)

5 Dagum P, Luby M. Approximating probabilistic inference in Bayesian belief networks is NP-hard. Artificial Intelligence,1993(60)

6 Zhang S, Zhang C. A class of tractable problems in Bayesian networks. networks is NP-hard. Artificial Intelligence,1993(60)

7 Zhang S, Zhang C. A class of tractable problems in Bayesian networks. Journal of the ACM,1997(12)

8 Pearl J. *Probabilistic* reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann Publishers,1988

9 Santos J. A Fast Hill-climbing Approach without an Energy Function for Probabilistic Reasoning. In: Proc. of the 5th IEEE Interl. Conf. on Tools with Artificial Intelligence. 1993. 170～179

10 Santos J. On linear potential functions for approximating Bayesian computations. Journal of the ACM, 1996, 43: 399～430

欢迎订阅《计算机科学》

明年

2000年之举 全面提高质量 增加版面容量