第19章　用不确定信息进行推理

一个agent关于它的任务和环境常常只有不确定的信息。到目前为止所讨论的技术对不确定知识的表示和推理能力十分有限。一条语句如 P∨Q允许我们表示不确定性——P和Q哪一个是真的，但是还没有描述如何表示关于P或Q不确定的程度。

在一般逻辑中，我们能从P和P　Q中推导出Q，即，如果一个agent知道P　Q，它随后学会了P，那么它也能推导出Q。当信息不确定时有类似的推理过程吗？各种形式化已被用来对不确定信息进行表示和推理。前面已经提到过被MYCIN和PROSPECTOR使用的形式化。开发得最好的形式化（一些人对“最适合的”有争议）是基于概率的。在本章的开始，先回顾一下概率论的基础知识。

19.1　概率论简介

19.1.1　基本思想

假定有随机变量V_1,V_2，…，V_k的一个集合。当我们想谈论V_i的值而没有说这个值是什么时，我们使用符号v_i。在例子中，随机变量代表一个所讨论领域的特征。随机变量的值可以是不同类型。如果变量代表命题，它们的值为True或False(或者，1或0)；如果变量代表物理度量（如高度、密度和速度等），则值是数字；如果变量代表分类（如颜色、字母表中的字母等），则值是范畴。例如，投一个硬币的结果可以用简单变量C表示，它的值c可以是分类值H（头）或T（尾）。如果我们正在谈论把一个硬币投k次的结果值，我们需要k个变量（$C_1,\cdots,C_k$），它们中的每一个有值H或T。

我们用表达式p（$V_1=v_1,V_2=v_2,\cdots,V_k=v_k$）指称一个联合概率，即变量$V_1$，$V_2,\cdots$，$V_k$的值分别是$v_1$，$v_2,\cdots$，$v_k$时的概率。表达式$p$（$V_1$，$V_2,\cdots$，$V_k$）叫做变量$V_1$，$V_2,\cdots$，$V_k$的联合概率函数。它把变量集合映射为一个在0和1之间的实数。把$p(V_1$，$V_2,\cdots$，$V_k)$中的变量替换为特定的值，以给我们一个表达式$p(v_1$，$v_2,\cdots v_k)$——p（$V_1=v_1,V_2=v_2,\cdots,V_k=v_k$）的缩写形式。因此，对一个公平的硬币投掷，我们有$p$（$H$）=1/2。如果我们将一个硬币公平地投五次，就可能有p（H，T，T，H，T）=1/32。p（H，T，T，H，T）是第一投结果为头、第二投为尾、第三投为尾、第四投为头和第五投为尾的一个联合概率。

概率函数必须满足一定的属性，它们是：

(a)　$0 \le p(V_1, V_2, \ldots, V_k) \le 1$

上式适用于变量的任何分配值，且

(b)　$\sum p(V_1, V_2, \ldots, V_k) = 1$

其中的和是建立在变量的所有值的基础之上。因此，在投硬币的例子中，p（H）=1/2是和（a）一致的，然而$p(H)$=1/2和属性（b）一起限制p（T）等于1/2。这里不过多谈论如何给随机变量值分配概率,就像在命题演算中由合式公式指称的各种命题的真假是基于应用领域的专家主观判断（或者传感数据的知觉处理），随机变量的概率值也同样依赖专家判断或知觉处理。相反，

我们主要关心的是怎么执行计算，以让它告诉我们所感兴趣的变量的概率。

在本章的应用例子中，变量对应一个域的命题。这些命题或为真或为假，相应的命题变量将有True或False值。我们可能不确定关于这些命题的一个或多个的事实，这种不确定性能用相应变量的值的概率表示。因此，本章描述的技术可以认为是第13章和第14章讨论的使用谓词逻辑进行表示和推理的概率方案（正在开发的一阶逻辑概率方案是一个前沿的研究问题，可参见[Nilsson 1986,Glesner & Koller 1995]）。

用一个特定的例子将有助于表达重要的概念。用一个和前面演示命题演算中的推理相同的例子。回想命题原子BAT_OK、MOVES和LIFTABLE，它们的意图分别是电池被充满了电、手臂可以移动（当拿积木时）和积木是可以举起的。除此之外，我们加入原子 GAUGE，它指称电池状态的量度，表示电池是被充满电的。为了使图和公式不太繁杂，用简单的字母 B、M、L和G重新命名这些原子。现在，假定我们不能确定这些原子是 *True*还是*False*。在提取任何传感器的读数前，有一个对这些值的各种组合的先验概率。例如，当其他都是 *True*时，我们认为M为*False*是不太可能的。

因为有4个二值变量，故对这些变量的每种形式（B=b, M=m, L=l, G=g），其中b, m, l和g的值是 *True*或*False*，有16种联合概率。一个agent设计者可以指定这16个值，同时要遵守约束：每个值在0和1之间，它们的和是1。作为一个例子，在下面的表中列出了这些联合概率中的一部分：

(B，M，L，G)	联合概率
(*True*，*True*，*True*，*True*)	0.5686
(*True*，*True*，*True*，*False*)	0.0299
(*True*，*True*，*False*，*True*)	0.0135
(*True*，*True*，*False*，*False*)	0.0007
…	

（当然，一个设计者不可能用表中给定的精确程度指定概率。这样做是为了使这些值和本章后面给出的这个例子的其他有关概率相一致）。

当我们知道一个随机变量集合的联合概率的所有值时，就能计算这些随机变量之一的边缘概率(*marginal probability*)。例如，边缘概率p（B = b）被定义为是16个联合概率中B= b 的8个概率之和：

$$p(\mathrm{B}=b)=\sum_{\mathrm{B}=b} p(\mathrm{B},\mathrm{M},\mathrm{L},\mathrm{G})$$

用这个公式，边缘概率p（B=*True*）= 0.95,这是B值为*True*的8个联合概率之和。

更低阶数的联合概率也能通过对所有联合概率的合适项相加而计算得到。例如，对于 B = b，M = m，联合概率p（B = b, M = m）是所有联合概率中4项之和。

$$p(\mathrm{B}=b,\mathrm{M}=m)=\sum_{\mathrm{B}=b,\mathrm{M}=m} p(\mathrm{B},\mathrm{M},\mathrm{L},\mathrm{G})$$

当已知更低阶数的联合概率时，我们也能用它们计算边界和其他更低阶的联合概率。因此，例如

$$p(\mathrm{B}=b)=\sum_{\mathrm{B}=b} p(\mathrm{B},\mathrm{M})$$

和

$$p(\mathrm{B}=b,\mathrm{M}=m)=\sum_{\mathrm{B}=b,\mathrm{M}=m} p(\mathrm{B},\mathrm{M},\mathrm{L})$$

当处理命题变量(有*True*或*False* 值)时，常常利用一个简洁符号 ,例如，不必再写为p（B =

True, M = *False*）的形式，而将它记为*p*（B，¬M）——假定没有取反的变量已被实例化为*True*，取反变量被实例化为*False*。只有当上下文清楚地表明正指示一个实例化变量的概率值，而不是那些变量上的概率函数时，才能使用这种缩写符号。

因此，给定一个随机变量集合的完全联合概率函数（如一个表），从理论上讲，就能计算所有的边缘概率和所有的更低阶的联合概率。然而，当我们有一个极大的随机变量集合时，指定所有的联合概率的任务就变得不可处理，更不用说低阶概率了。幸运的是，在大多数应用中，联合概率要满足一定的特殊条件，这些条件使得对它们的说明和计算变得可行。本章的后面将描述这些条件。

19.1.2 条件概率

我们想能够用一些变量值的信息来获得其他变量值的概率。例如，如果搬积木的机器人感知到自己手臂不能移动，它可能想计算（给定那个事实）电池要被充电的概率。和逻辑推理方法相似，这样的计算叫做概率推理。在解释如何执行概率推理前，必须先定义什么是条件概率。

给定V_j，V_i的条件概率函数由$p(V_i|V_j)$表示。对变量V_i和V_j的任何值，可以给出

$$p(V_i|V_j)=\frac{p(V_i, V_j)}{p(V_j)}$$

其中$p(V_i,V_j)$是V_i和V_j的联合概率，$p(V_j)$是V_j的边缘概率。从这个表达式，我们也能按照条件概率表示一个联合概率

$$p(V_i, V_j)=p(V_i|V_j)p(V_j)$$

回到搬积木的例子，给定条件手臂不能移动，我们能计算电池被充电的概率

$$p(\mathrm{B}=True|\mathrm{M}=False)=\frac{p(\mathrm{B}=True, \mathrm{M}=False)}{p(\mathrm{M}=False)}$$

这个表达式的分子分母都能用前面解释的联合概率的求和计算得到。

用一个概率的频率的解释，可以帮助我们比较容易地理解条件概率。在这样一个解释中，例如*p*（M=*False*）是手臂不能移动的次数和总的尝试次数的比率（在某个想像的实验中执行无限次）。因此，给定手臂不能移动，电池被充电的概率就是手臂不动、电池充电的次数除以手臂不动的次数的结果。因此，一个条件概率是对一个联合概率的规范化。

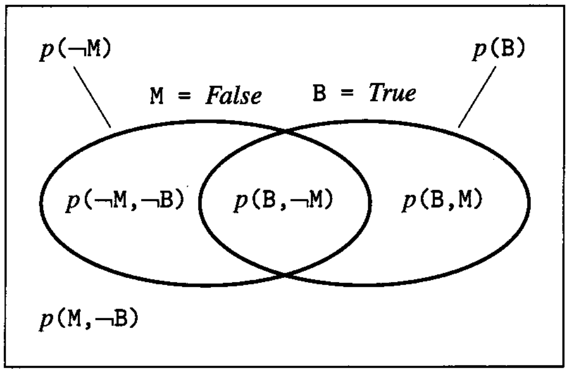
如图19-1所示的Venn[⊖]图有助于说明联合概率和条件概率（对少量的变量）。在那个图中，显示了两个重叠的椭圆区域，一个表示手臂不能移动的几率（M=*False*），一个表示电池被充电的几率（B=*True*）。每一个区域都和相应的（边缘）概率成比例，它们用简化符号标在图中，两个椭圆外部的区域对应手臂能动、电池没充电（*p*(M=*True*, B=*False*)）的几率。

尤其要注意椭圆的三个独立的不相交的部分，它们分别对应手臂不动电池没电、手臂不动电池有电和手臂能动电池有电的组合几率。这些分开的部分的每个区域与图中相应的联合概率成比例。我们从联合概率计算边缘概率的方法显然是源于图中*p*（B）=*p*（B，M）+*p*（B，¬M）的事实。

我们也有几个变量基于另外几个变量的组合条件概率。例如，（用简写符号）

⊖ John Venn 是一个英国逻辑学家[Venn 1880]。

$$p(\neg G, B|\neg M, L) = \frac{p(\neg G, B, \neg M, L)}{p(\neg M, L)}$$



$$p(B|\neg M) = p(B,\neg M)/p(\neg M)$$

图19-1 一个Venn图

在计算任何条件概率时，出现在计算中的联合概率和边缘概率能从前面描述的包含所有必需变量的任何完全联合概率集中计算得到。

我们也能按照一个条件概率链表达一个联合概率。例如

$$p(B, L, G, M) = p(B|L, G, M)p(L|G, M)p(G|M)p(M)$$

这个链规则的一般形式是

$$p(V_1, V_2, \ldots, V_k) = \prod_{i=1}^{k} p(V_i|V_{i-1}, \ldots, V_1)$$

链规则表达式依赖于我们选择对 V_i排序的方式。不同的排序给出不同的表达式，但对变量值的相同集合它们都有相同的值。

由于在一个联合概率函数中变量排序的方式并不重要（只要跟踪谁是谁就行了），我们能写出：

$$p(V_i, V_j) = p(V_i|V_j)p(V_j) = p(V_j|V_i)p(V_i) = p(V_j, V_i)$$

注意到

$$p(V_i|V_j) = \frac{p(V_j|V_i)p(V_i)}{p(V_j)}$$

后面这个等式是非常重要的，叫做贝叶斯法则[⊖]。

下面介绍一个最后的符号约定。当有一个变量集合的联合概率或一个变量集合的条件概率时，使用集合符号将很方便。因此p（$\mathcal{V}$）有时被用作p（$V_1, V_2,\cdots, V_k$）的一个缩写，其中 $\mathcal{V}=\{V_1, V_2,\cdots, V_k\}$。同样地，我们可以用$p$（$\mathcal{V}|\mathcal{V}_j$）,其中 $\mathcal{V}_j$ 也是一个变量集合。如果变量（$V_1, V_2,\cdots, V_k$）分别有值$v_1, v_2,\cdots, v_k$，我们用表达式 $\mathcal{V}=v$ 表示这个事实， $\mathcal{V}$ 和v都是有序列表。

⊖ 贝叶斯法则是Reverend Thomas Bayas [Bayes 1763]首次用公式表示的。

19.2 概率推理

19.2.1 一个一般的方法

概率推理的一般情景置是：我们有命题变量 $V_1, V_2, \cdots, V_k$ 的一个集合 $\mathcal{V}$，并给定了 $\mathcal{V}$ 的子集 ε中的变量的某些值 $\varepsilon=\mathbf{e}$（*True*或*False*）作为证据。在agent应用中，这些“给定”的变量通常有由感知过程决定的值。我们希望计算条件概率 $p(V_i=v_i|\varepsilon=e)$，即给定证据时变量 V_i 的值为 v_i 条件概率。我们把这个过程叫概率推理。

由于 V_i 有值*True*或*False*,故我们对两个条件概率感兴趣，它们是 $p(V_i=True|\varepsilon=e)$ 和 $p(V_i=False|\varepsilon=e)$。当然，我们只要计算它们中的一个就行了，因为 $p(V_i=True|\varepsilon=e)+p(V_i=False|\varepsilon=e)=1$，不管ε为何值。用“笨”方法说明 $p(V_i=True|\varepsilon=e)$ 的计算。用条件概率的定义，我们有

$$p(V_i=True|\mathcal{E}=\mathbf{e})=\frac{p(V_i=True,\mathcal{E}=\mathbf{e})}{p(\mathcal{E}=\mathbf{e})}$$

其中，$p(V_i=True, \varepsilon=e)$ 通过使用从高阶联合概率计算低阶联合概率的方法获得：

$$p(V_i=True,\mathcal{E}=\mathbf{e})=\sum_{V_i=True,\mathcal{E}=\mathbf{e}} p(V_1,\ldots,V_k)$$

其中 $V_i, i=1,\cdots,k$ 构成了命题变量集合。即，对 $V_i=True$，证据变量有它们的给定值。对所有的联合概率值求和。$p(\varepsilon=e)$ 的计算能用同样的方式进行，然而像下一个例子演示的一样，它不需要明确地计算。

作为一个例子，假如我们有下面给出的联合概率：

$p(P, Q, R) = 0.3$

$p(P, Q, \neg R) = 0.2$

$p(P, \neg Q, R) = 0.2$

$p(P, \neg Q, \neg R) = 0.1$

$p(\neg P, Q, R) = 0.05$

$p(\neg P, Q, \neg R) = 0.1$

$p(\neg P, Q, \neg R) = 0.05$

$p(\neg P, \neg Q, \neg R) = 0.0$

$\neg R$ 作为证据，我们希望计算 $p(Q|\neg R)$。用刚刚给定的过程，我们计算

$$p(Q|\neg R)=\frac{p(Q,\neg R)}{p(\neg R)}=\frac{[p(P,Q,\neg R)+p(\neg P,Q,\neg R)]}{p(\neg R)}$$

$$=\frac{(0.2+0.1)}{p(\neg R)}=\frac{0.3}{p(\neg R)}$$

现在我们既可直接计算边缘 $p(\neg R)$，也可（像通常所做的一样）用刚刚使用的相同方法计算 $p(\neg Q|\neg R)$——通过利用 $p(Q|\neg R)+p(\neg Q|\neg R)=1$ 来避免计算 $p(\neg R)$ 。

下面用后一种方式进行：

$$p(\neg Q|\neg R)=\frac{p(\neg Q,\neg R)}{p(\neg R)}=\frac{[p(P,\neg Q,\neg R)+p(\neg P,\neg Q,\neg R)]}{p(\neg R)}$$

$$=\frac{(0.1+0.0)}{p(\neg R)}=\frac{0.1}{p(\neg R)}$$

由于这两个量的和为1，我们得到$p(Q \mid \neg R)=0.75$。

一般地讲，使用这个方法的概率推理是难处理的，因为在有 k个变量的情况下执行它，我们需要联合概率$p(V_1, V_2,\cdots, V_k)$的 2^k个值的一个显式列表。对很多问题，即使我们知道这样一个列表也不能写出它（一般不这样做）。

考虑到这种难处理性，我们可能要问："人是如何对不确定信息有效地推理的？" Pearl[Pearl1986,Pearl 1988,Pearl 1990]推测人类通过一个特殊的方式把一个领域的知识公式化来做推理。这种方式能极大地简化一定变量在给定证据下的条件概率的计算。这些有效的知识公式化涉及到各种变量中的条件独立性——马上要讲的一个主题。

19.2.2 条件独立

给定变量集合 $\mathcal{V}_j$，如果$p(V|\mathcal{V}_i, \mathcal{V}_j) = p(V|\mathcal{V}_j)$，那么我们就说变量$V$条件独立于变量集 $\mathcal{V}_i$，用符号$I(V, \mathcal{V}_i|\mathcal{V}_j)$阐述这个事实。条件独立后面的直觉知识是如果$I(V, \mathcal{V}_i \mid \mathcal{V})$，那么 $\mathcal{V}_i$不会告诉我们比我们通过 $\mathcal{V}_j$ 知道的任何更多的东西。对V 而言，如果我们知道 $\mathcal{V}_j$，可以忽略 $\mathcal{V}_i$ 。在举积木的例子中，这应该是合理的：如果我们已知（通过其他的一些方式）电池有电（B=*True*），那么在手臂移动的范围内，我们不需要有关 G的（量规指示电池有电）明确知识。即 p（M|B，G）= p（M|B）。

给定一个集合 $\mathcal{V}$，如果一个变量V_i 是条件独立于另一个变量V_j，则有（按照定义）$p(V_i|V_j, \mathcal{V}) = p(V_i|\mathcal{V})$。根据条件概率的定义，有$p(V_i|V_j, \mathcal{V})p(V_j|\mathcal{V}) = p(T_i,V_j|\mathcal{V})$。对$I(V_i,V_j|\mathcal{V})$。组合这两个结果产生：

$$p(V_i, V_j|\mathcal{V}) = p(V_i|\mathcal{V})p(V_j|\mathcal{V})$$

注意到V_i和V_j对称地出现。因此，给定 $\mathcal{V}$，说V_i 条件独立于V_j，也就是说V_j条件独立于V_i。它足以说明给定 $\mathcal{V}$，V_i和V_j 是条件独立的。相同的结果可用于集合，即给定 $\mathcal{V}$，如果 $\mathcal{V}_i$ 和 $\mathcal{V}_j$ 是条件独立的，那么$p(\mathcal{V}_i, \mathcal{V}_j|\mathcal{V}) = p(T_i|\mathcal{V})p(V_j|\mathcal{V})$。如果 $\mathcal{V}$ 是空集，我们简单地说 $\mathcal{V}_i$ 和 $\mathcal{V}_j$是独立的。

不失一般性，我们说给定集合 $\mathcal{V}$，如果每个变量条件独立于所有其他的变量，那么变量$V_1,\ldots,V_k$是相互条件独立的。由于

$$p(V_1, V_2, \ldots, V_k|\mathcal{V}) = \prod_{i=1}^{k} p(V_i|V_{i-1}, \ldots, V_1, \mathcal{V})$$

且每个V_i是条件独立于其他给定的 ，于是我们有

$$p(V_1, V_2, \ldots, V_k|\mathcal{V}) = \prod_{i=1}^{k} p(V_i|\mathcal{V})$$

当 $\mathcal{V}$ 为空时，我们有

$$p(V_1, V_2, \ldots, V_k) = p(V_1)p(V_2)\cdots p(V_k)$$

故我们说变量是元条件独立的。

条件独立性能用贝叶斯网（也叫信念网）结构方便地表示。这些结构对概率推理是非常有

用的。用贝叶斯网表示的条件独立能大量地节约概率推理计算。

19.3 贝叶斯网

一个贝叶斯网是一个有向无环图（DAG)，它的节点用随机变量标识。一个贝叶斯网规定图中的每个节点V_i条件独立于由V_i的父节点给定的V_i的非后代节点构成的任何节点子集。也就是说，假设$\mathcal{A}$（V_i）是图中非 V_i 后代节点的任何节点集合，设$\mathcal{P}$（V_i）是图中V_i的直接双亲。图仅仅是陈述对图中的所有V_i, I（V_i, $\mathcal{A}$ $(V_i)|$ $\mathcal{P}$ (V_i)）的一种方式，I（V_i, $\mathcal{A}$ $(V_i)|$ $\mathcal{P}$ (V_i)）的意思是p（$V_i|\mathcal{A}$ (V_i), $\mathcal{P}$ $(V_i)) = p(V_i|\mathcal{P}$ $(V_i))$。

假设V_1, V_2,⋯, V_k 是贝叶斯网中的节点，给定由网络假设的条件独立性，我们能写出网中所有节点的联合概率如下：

$$p(V_1, V_2, \ldots, V_k) = \prod_{i=1}^{k} p(V_i|\mathcal{P}(V_i))$$

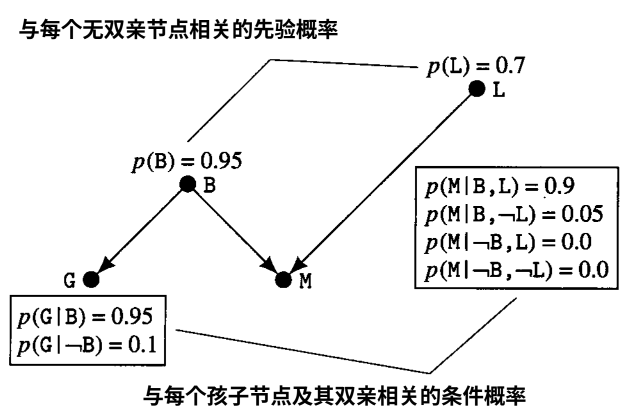
这个表达式能用一个直接的方式推导出，利用与贝叶斯网 DAG蕴含的部分序一致的链规则序，将条件独立性应用于链规则表达式，可计算所有变量的联合概率。

贝叶斯网有时叫做因果网，因为可以将连接节点的弧认为是表达了直接的因果关系。人类专家常常能把原因和结果用一种方式联系起来，这种方式显示了继承的条件独立性，这也能用得到的贝叶斯网描绘。使用直觉因果概念来构建贝叶斯网，通常可以使其能实现内含的条件独立性假设更为合适。一位研究者说过 [Heckerman 1996,p.14] ：“……为了构造一个给定变量集的贝叶斯网，我们从原因变量向直接结果画弧。在几乎所有的情况下，这样做会生成一个贝叶斯网[它的条件独立性蕴含是精确的]。”

用举积木的例子演示一个贝叶斯网的构造。我们首先想像这个例子的“第一个原因”，即和命题“电池被充电”（B）和“积木是可举起来的”（L）相对应的变量。B和L对M（“手臂移动”）有一个因果影响，B对G（“量规指示电池被充电了”）也有因果影响。因此，我们将为这个问题画一个如图 19-2所示的贝叶斯网。注意，除了其他的，网络陈述了 p（M|G，B，L）$=p$（M|B，L)。如果在网络中有另外的节点 U（意指积木被举起)，将不会有p（M|G，B，L，U）=p（M|B，L)，因为U是M的一个后继（积木被拿起影响了手臂移动的概率，除此之外，还有其他能影响吗?)。网中所有节点的联合概率函数的表达式都在图中给出。

我们看到，为了计算给定贝叶斯网的联合概率值，如图19-2所示，我们需要知道恰恰以它的父节点为条件的每个节点的条件概率函数。对没有父节点的节点，概率不以其他节点为条件，这些叫做这些变量的先验概率。因此，一个随机变量集合的概率的一个完整说明涉及到这些变量的一个贝叶斯网及网中每个变量的条件概率表（CPT)。

举积木例子中联合概率的贝叶斯网公式应该与由使用链规则：



p(G, B, M，L)$=p$(G|B) p (M|B,L) p(B) p (L)

图19-2 一个贝叶斯网

$$p(\mathrm{G, M, B, L}) = p(\mathrm{G|B, M, L})p(\mathrm{M|B, L})p(\mathrm{B|L})p(\mathrm{L})$$

获得的一个相似公式（假定没有任何条件独立）进行比较。注意到贝叶斯网公式更简单。没有贝叶斯网规定的条件独立性，对这个例子的所有 4个变量的一个联合概率的规范涉及到指定 16个独立的联合概率（实际仅需要 15个，因为它们的和必须为 1）。从图19-2可以明显看到，由贝叶斯网所做的假设要求我们仅指定 8个概率。当在领域中的变量中有几个条件独立时，从贝叶斯网计算的联合概率表达式要求的概率规范比没有这些独立时的还要少。这种减少有时会使难处理的问题变得可处理。

19.4 贝叶斯网的推理模式

在贝叶斯网中有三种重要的推理模式。为了解释它们，继续我们的例子。

• 因果推理或由上向下推理。给定积木是可举起的，假如我们想计算手臂能移动的概率 p(M|L)。由于积木可举起是手臂能移动的原因之一。我们说这个计算是因果推理的一个例子。L称做用于推理的证据，M叫询问节点。下面说明我们在这种情况下如何执行推理：首先，我们把p（M|L）（一个边缘概率）扩展成两个联合概率的和（因为我们想提及 M的另一个双亲B）：

$$p(\mathrm{M|L}) = p(\mathrm{M, B|L}) + p(\mathrm{M, \neg B|L})$$

接下来，我们希望M以另一双亲为条件，因此我们用一个链规则形式得到

$$p(\mathrm{M|L}) = p(\mathrm{M|B, L})p(\mathrm{B|L}) + p(\mathrm{M|\neg B, L})p(\mathrm{\neg B|L})$$

但是p（B|L）= p（B）（从网结构而来，注意B没有任何双亲）。同样地，p（¬B|L）= p（¬B）

因此，p（M|L）= p（M|B，L）p（B）+ p（M|¬B，L）p（¬B）。由于所有的这些数值与网络一同给出，我们能计算

$$p(\mathrm{M|L}) = 0.855$$

我们在这个例子中执行的操作值得注意，因为它们能被一般化为如我们后面看到的更复杂的因果推理。主要操作如下：

a) 按照给定证据的V和它的所有双亲（它们不是证据）的联合概率，重新表达给定证据的询问节点V的所求条件概率。

b) 回到以所有双亲为条件的 V的概率，重新表达这个联合概率。

• 诊断推理或自底向上推理。现在我们计算 p（¬L|¬M），即给定手臂未移动时积木不可举起的概率。其中询问和证据的角色刚好和它们在上一个例子中的相反。由于我们用一个结果（或症状）来推理一个起因，我们称这类推理叫诊断推理。

$$p(\mathrm{\neg L|\neg M}) = \frac{p(\mathrm{\neg M|\neg L})p(\mathrm{\neg L})}{p(\mathrm{\neg M})} \text{(贝叶斯规则)}$$

现在我们计算p（¬M|¬L）= 0.9525（用因果推理），并计算

$$p(\mathrm{\neg L|\neg M}) = \frac{0.9525 \times 0.3}{p(\mathrm{\neg M})} = \frac{0.28575}{p(\mathrm{\neg M})}$$

同样地，$p(\mathrm{L|\neg M}) = \frac{p(\mathrm{\neg M|L})p(\mathrm{L})}{p(\mathrm{\neg M})} = \frac{0.0595 \times 0.7}{p(\mathrm{\neg M})} = \frac{0.03665}{p(\mathrm{\neg M})}$。

由于这两个表达的和必须为1，得到p（¬ L | ¬ M）=0.88632。

用在这个简单例子中的诊断推理计算也能被一般化。主要步骤是使用贝叶斯规则把问题转化成因果推理。

• 辩解。如果我们的证据仅仅是¬ M（手臂不能移动），像刚做的那样，我们能计算积木不能举起的概率。但是如果也给定¬ B（电池没被充电），那么¬ L应该变得不确定。在这种情况下，我们说¬ B解释¬ M，使¬ L不确定。这类推理使用嵌入在一个自底而上或诊断推理中的自顶而下或因果推理。

$$p(\neg L|\neg B, \neg M) = \frac{p(\neg M, \neg B|\neg L)p(\neg L)}{p(\neg B, \neg M)} \text{（贝叶斯规则）}$$

$$= \frac{p(\neg M|\neg B, \neg L)p(\neg B|\neg L)p(\neg L)}{p(\neg B, \neg M)} \text{（条件概率定义）}$$

$$= \frac{p(\neg M|\neg B, \neg L)p(\neg B)p(\neg L)}{p(\neg B, \neg M)} \text{（贝叶斯网结构）}$$

从这个表达式，使用网中给定的概率，用普通方式求解 p（¬B，¬M），我们可以计算p（¬ L | ¬ B，¬ M）=0.030。像预期的一样，它比前面计算的p（¬L | ¬M）更小。再次注意贝叶斯法则的使用，它是辩解过程中的一个重要步骤。

19.5 不确定证据

当证据ε本身不确定时，表达式$p(V|\varepsilon)$，V是一个询问节点，不会给出正确的概率。在贝叶斯网计算中，为了“给定”证据节点，必须确定它们表示的命题的真假。我们能通过让每个证据节点（我们不能确定的）有一个确定的子节点来获得那个要求。因此，在上一个例子中（辩解），假定机器人对它的手臂不能移动不确定，它可能有一个有点不可靠的关节传感器。 在这种情况下，证据能被一个节点M'提供，M'代表命题“手臂传感器说手臂可以移动”。依赖它的读数，我们能确定命题是真还是假。然后用贝叶斯网计算 p(¬L | ¬B，¬M')，而不是p (¬L | ¬B，¬M)。当然，网络将需要p(M' | M)和p(M' | ¬ M)的值，它们描述了传感器的可靠性。

注意图19-2中的网络已提供了事实——关于电池是否被充电我们是不确定的。节点 B有一个子节点G，我们通过概率p(G | B)和p(G | ¬B)表示量规的可信度，留给你（也许在进一步阅读后）去计算p(¬L | ¬G，¬M')。

即使通过一个贝叶斯网给出的简化，用于从一个联合概率计算各种条件概率的强制方法对大的网络来讲，一般仍是难处理的。它的最坏时间复杂度是命题变量的指数。幸运的是，有几个简便方法可用于计算特殊网络的条件概率。下面，先表达贝叶斯网中关于条件独立的另一个结果，然后再考虑上述简便方法。

19.6 D分离

一个贝叶斯网比那些仅涉及个节点双亲的网蕴含了更多的条件独立，例如，在图 19-2中，p（M | G，B）=p（M | B），即，给定B，M条件独立于G（即使没有给出M的双亲）。从直觉上讲，在图19-2的网中，G的知识（结果）能影响B的知识（起因），B会影响M的知识（另一个结果）。但是如果给定原因 B，G并不能告诉我们有关 M更多的事情。在这种情况下，我们说 B *d*分离（依赖方向的分离） G和M。

其他的这种条件独立存在于贝叶斯网中。在这里仅介绍它们，你可以参考 [Pearl 1988, pp, 117~122]寻找证明。

如果对贝叶斯网中的节点 V_i 和 V_j 之间的每个无向路径，在路径上有某个节点 V_b，它有如下的三个属性之一（见图19-3），就说节点 V_i 和 V_j 条件独立于给定的节点集ε (即 $I(V_i, V_j|\varepsilon)$)。三个属性是：

1) V_b 在ε中，且路径上的两条弧都从 V_b 开始。

2) V_b 在ε中，路径上的一条弧以 V_b 为头，另一个以 V_b 为尾。

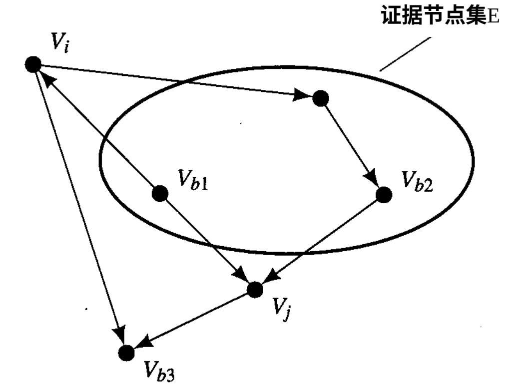
3) V_b 和它的任何后继都不在ε中，路径上的两条弧都以 V_b 为头。

给定ε，当这些条件中的任何一个占据一条路径时，我们说 V_b 阻塞那条路径。注意，在这个结果中引用的路径是无向路径，即路径忽略了弧方向。如果 V_i 和 V_j 之间的所有路径被阻塞，我们说 ε d分离 V_i 和 V_j（依赖方向的分离）且得出结论：V_i 和 V_j 条件独立于给定的ε。图19-2中d分离产生的其他条件独立的例子如下：

• I(G,L|B)，因为根据规则1，给定B阻塞了G和L之间的惟一的路径。根据规则3，给定B，M也阻塞了这个路径，因为 M不是证据集的一个成员。

• I（G，L）和 I（B，L），因为按规则了3，给定空的证据集合，M阻塞了G和L之间、B和L之间的（惟一）路径（M不是空的证据集的一个成员）。

然而，注意，B和L不是条件独立于给定的M的。因为虽然M在B和L的路径上，但这个路径上的两条弧都指向M，且M在证据集合中 —— 因此在这种情况下，M没有阻塞路径。



由于 V_i 到 V_j 之间的所有三条路径都被阻塞，给定证据节点，Vi 独立于 V_j。阻塞的节点为：

a) V_{b1} 是一证据节点，两条弧都用 V_{b1} 开始。

b) V_{b2} 是一证据节点，一条弧以 V_{b2} 为头，另一条弧以 V_{b2} 为尾。

c) V_{b3} 及其任一后代都不是证据节点，两条弧都以 V_{b3} 为头。

图19-3 通过阻塞节点的条件独立

*d*分离的概念也能应用到集合。给定ε，如果两个节点集 V_i 和 V_j 被 ε d分离，则它们是条件独立的。给定ε，如果 V_i 中的所有节点和 V_j 中的所有节点之间的每条无向路径被阻塞，则 V_i 和 V_j 被ε d分离。

即使使用了*d*分离，一般地讲，在贝叶斯网中，概率推理仍是NP难题[Copper 1990]。然而，有些简化能在一个叫polytree的重要网络分类中使用。一个polytree网是一个DAG，在该DAG的任何两个节点之间，顺着弧的每一个方向只有一条路径。例如，图 19-2的网络就是一个polytree。用一个扩展的符号例子来说明在 polytree中，是如何执行概率推理的（说明的例子基于由[Russell & Norvig 1995, pp. 447以后]提出的一个算法）。

19.7 在polytree中的概率推理

图19-4的网络是 polytree的一个典型例子。在这个网络中，给定一些其他的节点，我们想计算Q的概率。注意，有些节点仅通过 Q的双亲连到Q，我们说这些节点在Q的上方。其他的节

点仅通过Q的直接后继（它的孩子）连到Q，我们说这些节点在Q的下方。我们也注意到没有任何路径（除了通过Q连接Q上的一个节点和Q下的一个节点外——因为不这样的话，网络将不是一个polytree）将这些定义和连接属性应用到polytree中的每一个节点！我们的例子将有三种类型：

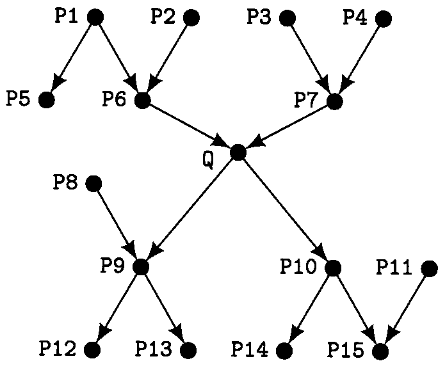


图19-4　一个典型的Polytree

1) 所有的证据节点在Q的上方。作为这种类型的一个典型例子，我们将计算 p（Q | P5，P4）。

2) 所有的证据节点在Q的下方。作为这种类型的一个典型例子，我们将计算 p（Q | P12，P13，P14，P11）。

3) 在Q的上方和下方均有证据节点。

19.7.1　证据在上方

我们先计算 p（Q | P5，P4），其中所有的证据节点均在Q的上方。我们的计算是沿着一个“自底向上”的递归算法执行的。这个算法在给定证据的情况下，计算Q的每个祖先的概率，直到我们到达证据节点或证据节点在那个祖先的下面。算法过程如下：

首先，我们“包括双亲”(Q的)：

$$p(\mathrm{Q}|\mathrm{P5},\mathrm{P4})=\sum_{\mathrm{P6},\mathrm{P7}}p(\mathrm{Q},\mathrm{P6},\mathrm{P7}|\mathrm{P5},\mathrm{P4})$$

(这个求和特殊符号的意思是相加p（Q，P6，P7 | P5，P4）的4个形式——原始的、用¬P6代替P6、用¬P7代替P7以及两个都代替。)

下面，我们用条件独立的定义产生Q的双亲部分的证据，记为

$$p(\mathrm{Q},\mathrm{P6},\mathrm{P7}|\mathrm{P5},\mathrm{P4})=p(\mathrm{Q}|\mathrm{P6},\mathrm{P7},\mathrm{P5},\mathrm{P4})p(\mathrm{P6},\mathrm{P7}|\mathrm{P5},\mathrm{P4})$$

替换产生

$$p(\mathrm{Q}|\mathrm{P5},\mathrm{P4})=\sum_{\mathrm{P6},\mathrm{P7}}p(\mathrm{Q}|\mathrm{P6},\mathrm{P7},\mathrm{P5},\mathrm{P4})p(\mathrm{P6},\mathrm{P7}|\mathrm{P5},\mathrm{P4})$$

现在，因为一个节点条件独立于它的双亲给定的非后继节点，

$$p(\mathrm{Q}|\mathrm{P5},\mathrm{P4})=\sum_{\mathrm{P6},\mathrm{P7}}p(\mathrm{Q}|\mathrm{P6},\mathrm{P7})p(\mathrm{P6},\mathrm{P7}|\mathrm{P5},\mathrm{P4})$$

那么，d分离允许我们分割双亲：

$$p(\mathrm{Q}|\mathrm{P5},\mathrm{P4})=\sum_{\mathrm{P6},\mathrm{P7}}p(\mathrm{Q}|\mathrm{P6},\mathrm{P7})p(\mathrm{P6}|\mathrm{P5},\mathrm{P4})p(\mathrm{P7}|\mathrm{P5},\mathrm{P4})$$

最后，在计算其他的概率中，d分离允许我们忽略一个双亲上的证据：

$$p(\mathrm{Q}|\mathrm{P5},\mathrm{P4})=\sum_{\mathrm{P6},\mathrm{P7}}p(\mathrm{Q}|\mathrm{P6},\mathrm{P7})p(\mathrm{P6}|\mathrm{P5})p(\mathrm{P7}|\mathrm{P4})$$

它是非常重要的——注意到正被求和的项是:（a）给定的父节点的各个值，询问节点的概率（父节点的概率与贝叶斯网一同给出）；（b）只要给出那个父节点上的部分证据，它是每个父

节点的概率（递归调用我们正在执行的算法）。这些结果直接利用了我们用一个polytree产生的事实。

这个相同的过程被递归应用，直到最终到达一个有一证据节点做为父节点的节点，或到达没有父节点的节点（不是自身证据节点的节点）。在计算p（P7|P4）中，我们有这两种情况的第一种，证据节点P4是询问节点P7的一个父节点。此时，"包括双亲"的步骤比较简单，因为其中之一已经被包括了。因为p（P7，P3|P4）=p（P7 |P3，P4）p（P3|P4），我们能写出

$$p(\mathrm{P7}|\mathrm{P4}) = \sum_{\mathrm{P3}} p(\mathrm{P7}|\mathrm{P3},\mathrm{P4})p(\mathrm{P3}|\mathrm{P4}) = \sum_{\mathrm{P3}} p(\mathrm{P7}|\mathrm{P3},\mathrm{P4})p(\mathrm{P3})$$

（因为I（P3，P4），故有最后一步。）这个求和中的所有项由贝叶斯网给出，因此过程连同这个例子的分枝一同结束。

在计算p（P6|P5）时，得到

$$p(\mathrm{P6}|\mathrm{P5}) = \sum_{\mathrm{P1},\mathrm{P2}} p(\mathrm{P6}|\mathrm{P1},\mathrm{P2})p(\mathrm{P1}|\mathrm{P5})p(\mathrm{P2})$$

下一步必须计算p（P1|P5），注意到证据节点不在询问节点的"上方"，而是在"下面"，我们不能再用这个递归过程，而必须用"证据在下方"的过程——将要描述。在这个例子中，我们仅仅用贝叶斯法则获得$p(\mathrm{P1}|\mathrm{P5}) = \frac{p(\mathrm{P5}|\mathrm{P1})p(\mathrm{P1})}{p(\mathrm{P5})}$。现在计算$p$（P6|P5）需要的所有量都由贝叶斯网给出了。我们能集成所有的这些结果（执行所有的求和）得到p（Q|P5，P4）的最终结果。

19.7.2 证据在下方

接着，我们计算p（Q|P12，P13，P14，P11），其中，所有的证据节点都在Q的下方。我们的计算再次沿着一个递归算法执行。它按如下进行：在顶级，我们用贝叶斯规则写出

$$p(\mathrm{Q}|\mathrm{P12},\mathrm{P13},\mathrm{P14},\mathrm{P11}) = \frac{p(\mathrm{P12},\mathrm{P13},\mathrm{P14},\mathrm{P11}|\mathrm{Q})p(\mathrm{Q})}{p(\mathrm{P12},\mathrm{P13},\mathrm{P14},\mathrm{P11})}$$
$$= kp(\mathrm{P12},\mathrm{P13},\mathrm{P14},\mathrm{P11}|\mathrm{Q})p(\mathrm{Q})$$

其中$k = \frac{1}{p(\mathrm{P12},\mathrm{P13},\mathrm{P14},\mathrm{P11})}$是一个在后面要按与前面的例子同样的方式计算的标准化因子。用d分离，I（{P12，P13}，{P14，P11}|Q），产生

$$p(\mathrm{Q}|\mathrm{P12},\mathrm{P13},\mathrm{P14},\mathrm{P11}) = kp(\mathrm{P12},\mathrm{P13}|\mathrm{Q})p(\mathrm{P14},\mathrm{P11}|\mathrm{Q})p(\mathrm{Q})$$

注意我们已把集合{P12，P13，P14，P11}分割成对应Q的两个孩子的子集。p（P12，P13|Q）和p（P14，P11|Q）的每一项包括给定其上的一个单一证据节点，计算一个询问节点集合的概率，因此，我们能使用类似于前面的算法。因为只有一个证据节点，故使用一个自顶而下的递归算法，而不用前面所用的自底而上算法。

通过首先计算p（P12， P13|Q），说明自顶而下方案是如何进行的。关键步骤是包括Q的单一孩子P9，它在询问节点集{P12，P13}的上方。注意，根据条件独立的定义，p（P12，P13，P9|Q）= p（P12，P13|P9，Q）p（P9|Q）。

那么，

$$p(\mathrm{P12},\mathrm{P13}|\mathrm{Q}) = \sum_{\mathrm{P9}} p(\mathrm{P12},\mathrm{P13}|\mathrm{P9},\mathrm{Q})p(\mathrm{P9}|\mathrm{Q})$$

现在，用d分离，I（{P12,P13},Q|P9)，故

$$p(\text{P12, P13}|\text{Q}) = \sum_{\text{P9}} p(\text{P12, P13}|\text{P9})p(\text{P9}|\text{Q})$$

这个求和中的项p（P9|Q）的计算涉及到P9的所有父节点：

$$p(\text{P9}|\text{Q}) = \sum_{\text{P8}} p(\text{P9}|\text{P8, Q})p(\text{P8})$$

p（P9|P8，Q）由网络给出。另一项p（P12，P13|P9）是对相同的自顶而下的过程的递归调用，该过程在给定询问节点集上的一个单一证据节点的情况下，计算它们的概率。在这种情况下，递归调用在一步后终止，因为P9的孩子是证据节点。由于P12和P13独立于给定的P9，因此有p（P12，P13|P9）= p（P12|P9）p（P13|P9）。这两个概率都由网络给出。

把自顶而下过程应用到p（P14，P11|Q），产生

$$p(\text{P14, P11}|\text{Q}) = \sum_{\text{P10}} p(\text{P14, P11}|\text{P10})p(\text{P10}|\text{Q})$$

那么由于I（P14，P11|P10），

$$p(\text{P14, P11}|\text{Q}) = \sum_{\text{P10}} p(\text{P14}|\text{P10}), p(\text{P11}|\text{P10})p(\text{P10}|\text{Q})$$

仅仅这个结果的中间项不是由网络直接给出。我们再次用自顶向下过程计算该项：

$$p(\text{P11}|\text{P10}) = \sum_{\text{P15}} p(\text{P11}|\text{P15, P10})p(\text{P15}|\text{P10})$$

其中

$$p(\text{P15}|\text{P10}) = \sum_{\text{P11}} p(\text{P15}|\text{P10, P11})p(\text{P11}) \text{（为什么）？}$$

但在p（P11|P15，P10）中，询问节点P11在证据节点的上方，因此我们必须再次应用这个过程的顶级（用贝叶斯规则）：

$$p(\text{P11}|\text{P15, P10}) = \frac{p(\text{P15, P10}|\text{P11})p(\text{P11})}{p(\text{P15, P10})} = k_1 p(\text{P15, P10}|\text{P11})p(\text{P11})$$

其中 $k_1 = \frac{1}{p(\text{P15,P10})}$ ，p（P11）由网络直接给出；因为P10和P11是独立的，算法终止于

$$p(\text{P15, P10}|\text{P11}) = p(\text{P15}|\text{P10, P11})p(\text{P10}|\text{P11}) = p(\text{P15}|\text{P10, P11})p(\text{P10})$$

现在，所有的结果能被收集，可以计算出总和以及k和k_1，以得到p（Q|P12，P13，P14，P11）的最终结果。

证据在上和证据在下算法的复杂度都和网络中节点数（对polytree而言）成线性关系。

19.7.3 证据在上下两方

如果在Q的上方和下方均有证据，如在p(Q|{P5,P4},{P12,P13,P14,P11})中，我们把证据分成上部ε^+和下部ε^-两部分，用一种贝叶斯规则

$$p(\text{Q}|\mathcal{E}^+, \mathcal{E}^-) = \frac{p(\mathcal{E}^-|\text{Q}, \mathcal{E}^+)p(\text{Q}|\mathcal{E}^+)}{p(\mathcal{E}^-|\mathcal{E}^+)}$$

像往常一样，我们把 $\frac{1}{p(\mathcal{E}^-|\mathcal{E}^+)} = k_2$ 作为一个标准化因子，并表达为：

$$p(Q|\mathcal{E}^+, \mathcal{E}^-) = k_2 p(\mathcal{E}^-|Q, \mathcal{E}^+) p(Q|\mathcal{E}^+)$$

注意，Q从ε^+中d分离ε^-，故

$$p(Q|\mathcal{E}^+, \mathcal{E}^-) = k_2 p(\mathcal{E}^-|Q) p(Q|\mathcal{E}^+)$$

注意到在这个结果中，我们计算的第一个概率已作为计算 $p(Q|\varepsilon^-)$的自顶而下的一部分。第二个概率直接用自底而上过程计算。

19.7.4 一个数值例子

在图19-5中，用一个更小的抽象 polytree数值化地说明这些方法的使用。我们的目的是计算p（Q | U）。

像平常的诊断推理一样，首先应用贝叶斯法则得到

$$p(Q|U) = kp(U|Q)p(Q), \quad \text{其中} \quad k = \frac{1}{p(U)}。$$

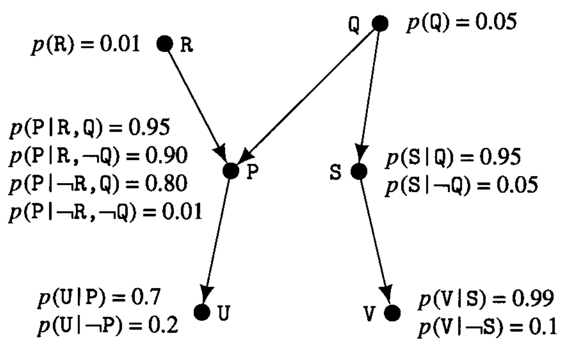


图19-5 一个小polytree

自顶而下算法连续计算

$$p(U|Q) = \sum_P p(U|P)p(P|Q)$$

$$p(P|Q) = \sum_R p(P|R, Q)p(R)$$

$$= p(P|R, Q)p(R) + p(P|\neg R, Q)p(\neg R)$$

$$= 0.95 \times 0.01 + 0.8 \times 0.99 = 0.80, \text{ 因此}$$

$$p(\neg P|Q) = 0.20$$

$$p(U|Q) = p(U|P) \times 0.8 + p(U|\neg P) \times 0.2$$

$$= 0.7 \times 0.8 + 0.2 \times 0.2 = 0.60, \text{ 这样}$$

$$p(Q|U) = k \times 0.6 \times 0.05 = k \times 0.03$$

$$p(\neg Q|U) = kp(U|\neg Q)p(\neg Q)$$

$$p(U|\neg Q) = \sum_P p(U|P)p(P|\neg Q)$$

$$p(P|\neg Q) = \sum_R p(P|R, \neg Q)p(R)$$

$$= p(P|R, \neg Q)p(R) + p(P|\neg R, \neg Q)p(\neg R)$$

$$= 0.90 \times 0.01 + 0.01 \times 0.99 = 0.019, \text{这样}$$

$$p(\neg P|\neg Q) = 0.98$$

$$p(U|\neg Q) = p(U|P) \times 0.019 + p(U|\neg P) \times 0.98$$

$$= 0.7 \times 0.019 + 0.2 \times 0.98 = 0.21, \text{这样}$$

$$p(\neg Q|U) = k \times 0.21 \times 0.95 = k \times 0.20$$

因此，k=4.35，最终

p（Q|U）= 4.35 ×0.03 = 0.13

像这个例子的这些计算能被组织以避免重复的子计算，这样做的一个方法是所谓的桶排除法（*bucket elimination*）[Dechter 1996]。

当网络不是一个polytree时，由于在节点之间有多条路径，上述的递归过程将不会终止。已有一些其他的技术被提出来解决这些更复杂的网络。其中之一是Monte Carlo方法(也叫逻辑采样[Henrion 1988])。在这个技术中，用无双亲节点的边缘概率来给那些节点分配随机值（如True或False）。用那些值，它们的后继的CPT被用来分配随机值给这些后继，等等，顺着网络向下。最终，网络中的每个节点都有一个值。这个过程被重复很多次，我们跟踪被分配给节点的所有值。在无限次试验的限制中，节点值将和被网络和它的 CPT规定的节点的联合概率相一致。经过大量的试验后，我们能用P和E被分配真值的次数除以E被分配真值的次数评估$p(Q|E)$的值。显然，给定一个证据节点集，能用同样的方法来计算一个询问节点集的联合概率。

另一个方法叫群集（*clustering*）[Lauritzen & Spiegelhalter 1988]，它把网络中的节点用一种方式分组成“超节点”以便超节点图是一个 polytree。超节点的可能值是它们的构成节点值的所有组合。然后可以使用 polytree算法，但现在对每个超节点有很多 CPT——给定所有超节点值的条件概率，以所有父节点值为条件（父节点自己可以是超节点）。

19.8 补充读物和讨论

有几本关于概率的课本可以用来补充本章内容；[Feller 1968]是其中的一本。

一些研究者认为非单调推理能用概率方法最好地处理。例如，参见 [Goldszmidt,Morris & Pearl 1990]。

在AI中，关于使用贝叶斯网的概率推理工作开始于[Pearl 1982a, Kim & Pearl 1983]，他们为树和polytree网分别开发了“消息传递”算法。在本章描述的 polytree方法是基于[Russell & Norvig 1995,pp.447以后]的。对贝叶斯网的处理只限于离散变量，对连续随机变量也已做了一些工作，参见[Shachter & Kenley 1989]。[Wellman 1990]研究了“定性”网络。

引用了由[Pearl 1984]写的关于概率推理的书。[Neapolitan 1990]是一本关于概率方法在专家系统中应用的课本。[Henrion 1990]是一篇有关贝叶斯网中概率推理的文章。[Jensen 1996]是一本关于贝叶斯网的课本，它以HUGIN系统为特征。[Neal 1991]调查了贝叶斯网和神经网络之间的联系。《Communicaions of the ACM》中关于“ AI的不确定性”是由 David Heckerman、Michael Wellman和Abe Mamdani编辑的（1995 年3月，第38卷，第3期）。

贝叶斯网已用在很多专家系统中。一个典型的例子是 PATHFINDER，它帮助病理学者诊断淋巴节疾病 [Heckerman 1991, Heckerman & Nathwani 1992]。另一个是用于内科医学的CPCSBN[Pradhan,et al. 1994]，它有448个节点和908个弧，可以和世界上内科医学的一流的诊

断专家相媲美。

除了贝叶斯网之外，还有几个可选的方法可对不确定信息进行推理。用于医疗诊断和治疗建议的MYCIN专家系统使用确定性因素[Shortliffe 1976,Buchanan & Shortliffe 1984]。[Duda,Hart,& Nilsson 1976]在他们的PROSPECTOR专家系统中使用充分性和必要性索引帮助矿物勘探。

其他的方法基于模糊逻辑和"概率理论：[Zadeh 1975,Zadeh 1978,Elkan 1993]以及组合Dempster-Shafer规则[Dempster 1968, Shafer 1979]。[Nilsson 1986]开发了一个"概率逻辑"，引用了概率论和多值逻辑中的相关工作。我认为贝叶斯网的概率推理主宰着用于大多数专家系统应用的其他方法，但是这个说法是有争议的。

当面对不确定性时，人类的行为可能是相当的不一致[Tversky & Kahneman 1982]。因此它不会对工程学提供有用的模型。

[Shafer & Pearl 1990]是关于不确定推理的一本论文集。在关于AI不确定性（UAI）的年会论文集中包含当前的研究进展。《International Journal of Approximate Reasoning》以及其他的AI期刊和AI会议学报也发表重要的论文。

习题

19.1 假定有颜色的小球分别在三个不能区分的盒子*B*1、*B*2和*B*3中，如下所示：

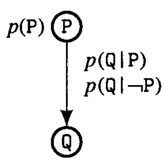
随机选择一个盒子，再从该盒子随机地选一个球，球是红色。被选择的盒子是*B*1、*B*2和*B*3的概率各是什么？解释你的推理。

	B1	B2	B3
红色	2	4	3
白色	3	2	4
蓝色	6	3	3

19.2 考虑下面显示的信念网络。

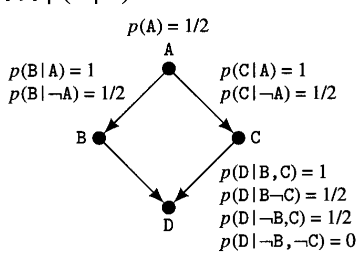
1) 推导出P Q的一个概率表达式。

2) p(P Q)和p(Q | P)什么时候相等？



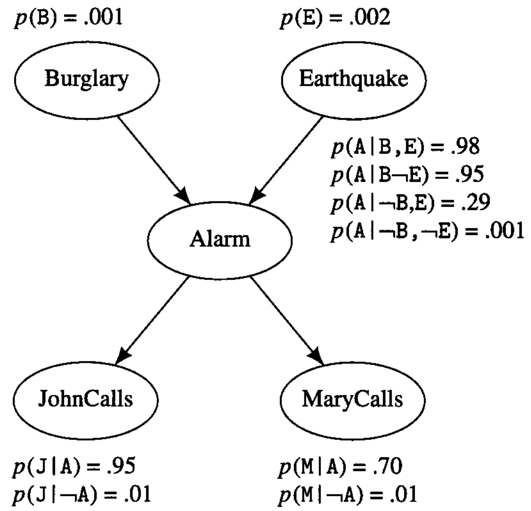
3) 假定不知道该网络的条件概率表，而是知道p(P)和p(P Q)的值。关于p(Q)的值能说些什么？

19.3 一个大学的招生委员会试图决定一个接受的申请人真正合格的概率。相关的概率给在如下的贝叶斯网中。计算p(A | D)



A=申请人合格
B=申请人有高平均分
C=申请人有优秀的推荐
D=申请通过

19.4 下面显示的信念网络形式化如下的情形：你有一个新的夜盗警铃按装在家里。它预防盗贼相当可靠，且偶尔会对小地震做出反应。你有两个邻居：John和Mary，他们都答应当听到警铃时会叫你。当John听到警铃时，他相当相信它，但有时却会混淆电话铃和警铃而叫你。另一方面，Mary喜欢大声的音乐，结果有时错过了警铃。



练习一下使用信念网络定义的联合概率工作的能力，计算在同时有地震和盗贼的情况下，John和Mary都没有呼叫的概率。即：计算p（¬J，¬M，B，E）。

19.5 在一个遥远的星球中。90%的出租车是绿色的，10%的是蓝色的。一个与一辆出租车有关的车祸发生了；我们假定绿车和蓝车的事故率相等。一个法庭处理这个事故，现场的一个记者说："出事的出租车是蓝色的"。记者常常是可信的；实际上，他的陈述在80%的时间内是正确的。即，如果出事的出租车确实是蓝色（或绿色的），我们的目击者说：蓝色（或绿色）的概率是 0.8。给定记者的陈述，出事出租车是蓝色的概率是多少？

19.6 变戏法机器人Orville，当它的电池电压低时，掉球是很常见的。在以前的测试中，已经决定了当电池低压时，它掉球的概率是 0.9。然而当电池电压不低时，掉球的概率仅仅是0.01。电池不久前刚被充电，我们最好的估计（在看 Orville的最新戏法记录之前）是电池低压时的几率是 10比1。一个有不太可靠视觉系统的机器人观察者，记录orviue的掉球。观察者的可信度由下面的概率给出：

p（观察者说Orville掉球了|Orville掉球了）= 0.9

p（观察者说Orville掉球了|Orville没有掉球）= 0.2

画出贝叶斯网，给定观察者的记录以计算电池是低压的概率。