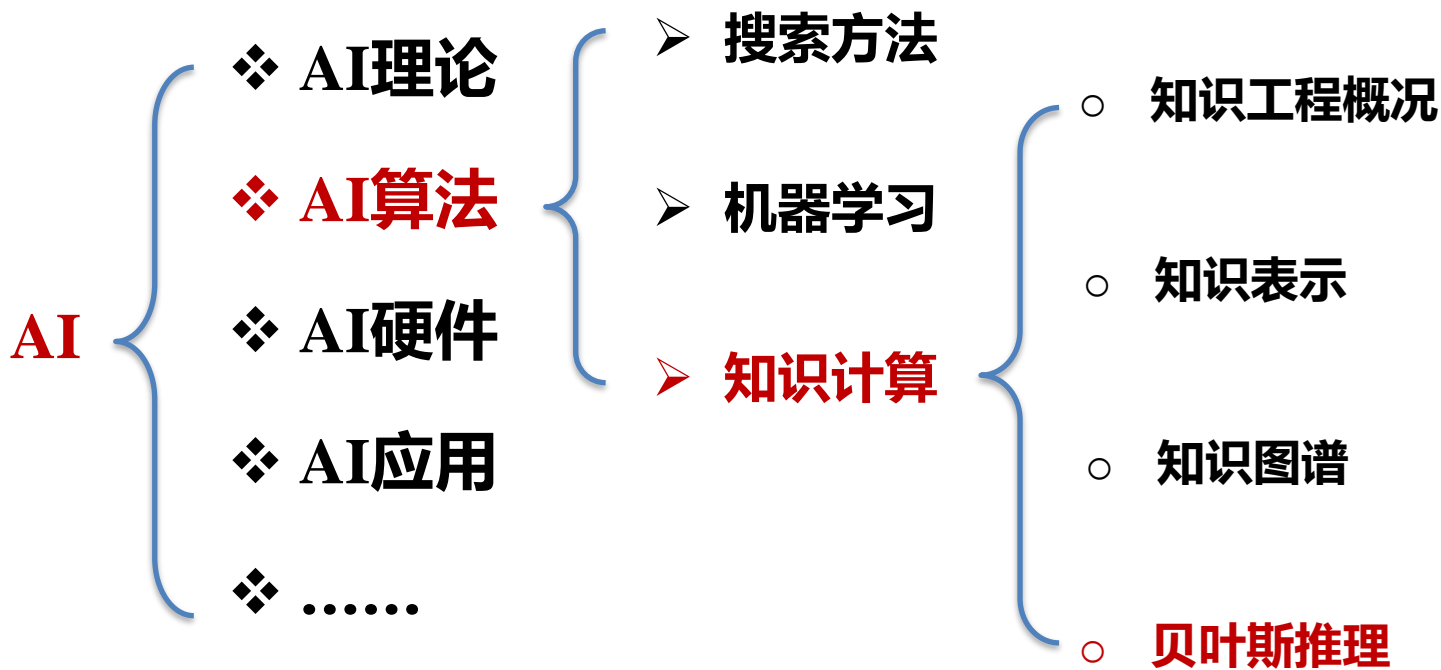




# 《人工智能》

## 第十五讲：贝叶斯网络







# 15.1 贝叶斯推理概述



- **贝叶斯网络是一种概率图模型**
  - 节点表示变量
  - 有向边表示变量间的依赖关系
- **表达能力强、适用面广**
  - 专家系统知识表示（80年代）
  - 贝叶斯学习（90年代）
- **特点：先验推导后验**



先看一个关于概率推理的例子。图15.1中有6个结点：参加晚会 (party, PT)、宿醉 (hangover, H0)、患脑瘤 (brain tumor, BT)、头疼 (headache, HA)、有酒精味 (smell alcohol, SA) 和X射线检查呈阳性 (posxray, PX)。可以把图15.1想象成为这样一个场景：一个中学生回家后，其父母猜测她参加了晚会，并且喝了酒；第二天这个学生感到头疼，她的父母带她到医院做头部的X光检查……

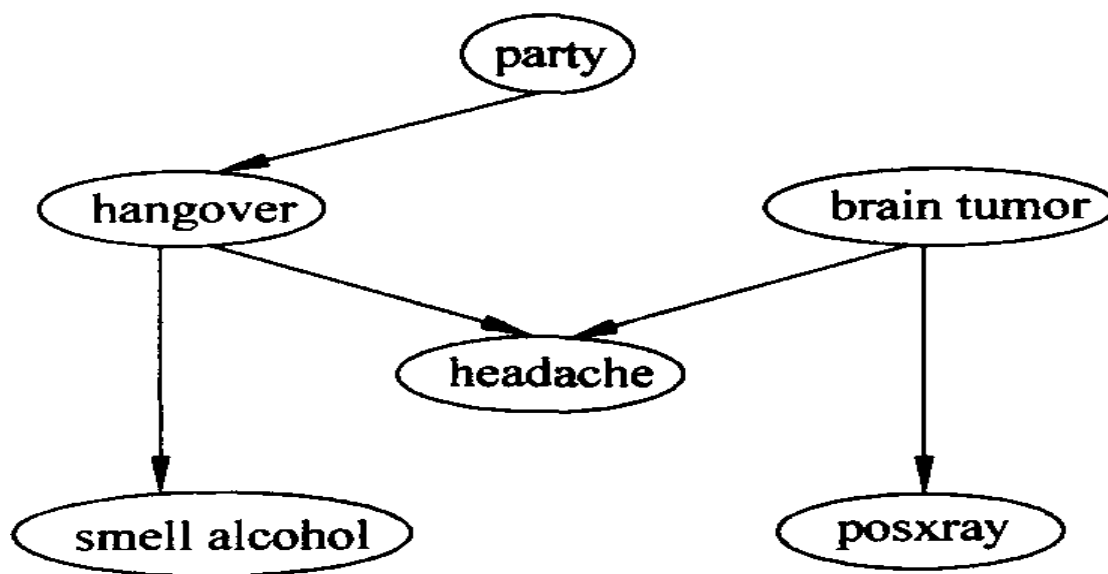


图15.1 贝叶斯推理例子



通过长期的观察，或者从别人那里了解，这个中学生的父母知道他们的女儿参加晚会的概率。通过长时间的数据积累，他们也知道他们的女儿参加晚会后宿醉的概率。因此，结点party和结点hangover之间有一条连线。同样，有明显的因果关系或相关关系的结点之间都有一条连线，并且连线从原因结点出发，指向结果结点。

针对图15.1所示的网络，有许多问题需要解决。例如：

- 1) 如果父母已知他们的女儿参加了晚会，那么第二天一早，她呼出的气体中有酒精味的概率有多大？也就是说，当party发生时，smell alcohol发生的概率有多大？
- 2) 如果他们的女儿头疼，那么她患脑瘤的概率有多大？这时，如果他们又知道昨晚她参加了晚会，那么综合这些情况，她患脑瘤的可能性有多大？

第一个例子是从原因推理结果的。第二个例子从结果反推原因，还有许多从结果反推原因的例子。例如，如果父母早晨闻到他们的女儿呼出的气体中有酒精味，那么她昨晚参加晚会的概率有多大？等等。

为了系统地解决上面的各类问题，需要先掌握一定的概率基础知识。



# 15.2 贝叶斯推理基础知识



**定义1 统计概率** 若在大量重复试验中，事件 $A$ 发生的频率稳定地接近于一个固定的常数 $p$ ，它表明事件 $A$ 出现的可能性大小，则称此常数 $p$ 为事件 $A$ 发生的概率，记为 $P(A)$ ，即

$$p = P(A) \quad (2)$$

可见概率就是频率的稳定中心。任何事件 $A$ 的概率为不大于1的非负实数，即

$$0 < P(A) < 1$$

**定义2 古典概率** 我们设一种次试验有且仅有有限的 $N$ 个可能结果，即 $N$ 个基本事件，而 $A$ 事件包含着 $K$ 个可能结果，则称 $K/N$ 为事件 $A$ 的概率，记为 $P(A)$ 。即

$$P(A) = K/N$$



**定义3 条件概率** 我们把事件 $B$ 已经出现的条件下，事件 $A$ 发生的概率记做为 $P(A|B)$ 。并称为在 $B$ 出现的条件下 $A$ 出现的**条件概率**，而称 $P(A)$ 为**无条件概率**。

若事件 $A$ 与 $B$ 中的任一个出现，并不影响另一事件出现的概率，即当

$$P(A) = P(A \cdot B) \text{ 或 } P(B) = P(B \cdot A) \text{ 时,}$$

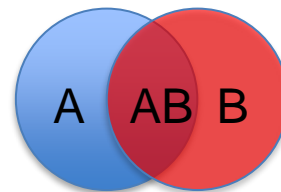
则称 $A$ 与 $B$ 是**相互独立的事件**。

**定理1 加法定理** 两个不相容(互斥)事件之和的概率, 等于两个事件概率之和, 即

$$P(A+B) = P(A) + P(B)$$

两个互逆事件 $A$ 和 $A^{-1}$ 的概率之和为1。即当 $A+A^{-1} = \Omega$ , 且 $A$ 与 $A^{-1}$ 互斥, 则 $P(A) + P(A^{-1}) = 1$ , 或常有 $P(A) = 1 - P(A^{-1})$ 。

若 $A$ 、 $B$ 为两任意事件, 则



$$P(A+B) = P(A) + P(B) - P(AB)$$



**定理2 乘法定理** 设 $A$ 、 $B$ 为两个不相容(互斥)非零事件，则其乘积的概率等于 $A$ 和 $B$ 概率的乘积，即

$$P(AB) = P(A)P(B) \quad \text{或} \quad P(AB) = P(B)P(A)$$

设 $A$ 、 $B$ 为两个任意的非零事件，则其乘积的概率等于 $A$ (或 $B$ )的概率与在 $A$ (或 $B$ )出现的条件下 $B$ (或 $A$ )出现的条件概率的乘积。

$$P(A \cdot B) = P(A) \cdot P(B|A) \quad \text{或} \quad P(A \cdot B) = P(B) \cdot P(A|B)$$



**(1) 先验概率。**先验概率是指根据历史的资料或主观判断所确定的各事件发生的概率，该类概率**没能经过实验证实**，属于检验前的概率，所以称之为**先验概率**。先验概率一般分为两类，一是**客观先验概率**，是指利用过去的历史资料计算得到的概率；二是**主观先验概率**，是指在无历史资料或历史资料不全的时候，只能凭借人们的主观经验来判断取得的概率。



**(2) 后验概率。** 后验概率一般是指**利用贝叶斯公式**，**结合调查等方式**获取了新的附加信息，对先验概率进行修正后得到的更符合实际的概率。

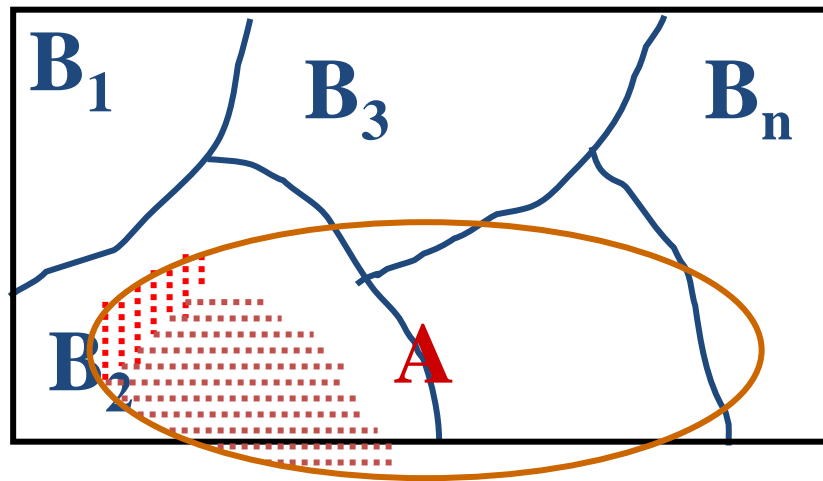
**(3) 联合概率。** 联合概率也叫**乘法公式**，是指**两个任意事件的乘积的概率**，或称之为**交事件的概率**。

**(4)全概率公式。** 设 $B_1, B_2, \dots, B_n$ 是两两互斥的事件, 且 $P(B_i) > 0, i = 1, 2, \dots, n, B_1 + B_2 + \dots + B_n = \Omega$ 。

另有一事件 $A = AB_1 + AB_2 + \dots + AB_n$

$$P(A) = \sum_{i=1}^n P(B_i)P(A | B_i)$$

称满足上述条件的  
 $B_1, B_2, \dots, B_n$ 为**完备事件组**。





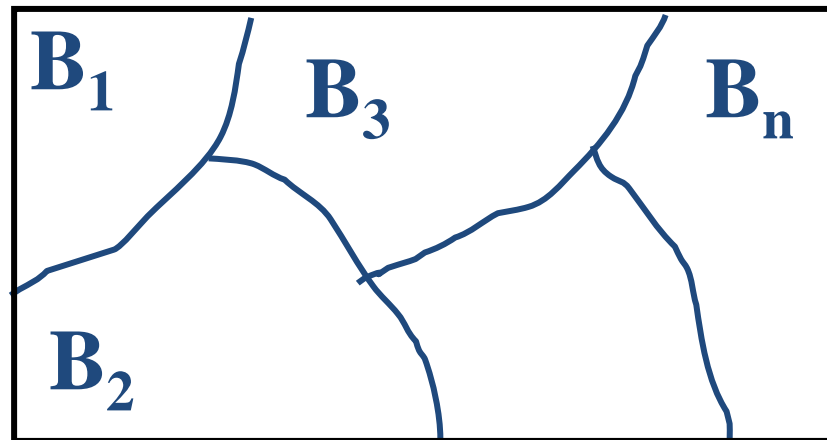
- 熵（信息熵）反应了系统的不确定性

$$H = - \sum_{i=1}^n P(B_i) \log_2 P(B_i)$$

(1)  $B_1=0.5, B_2=0.5, H=1$

(2)  $B_1=0.9, B_2=0.1, H=0.47$

(3)  $B_1=0.33, B_2=0.33, B_3=0.33, H=1.58$

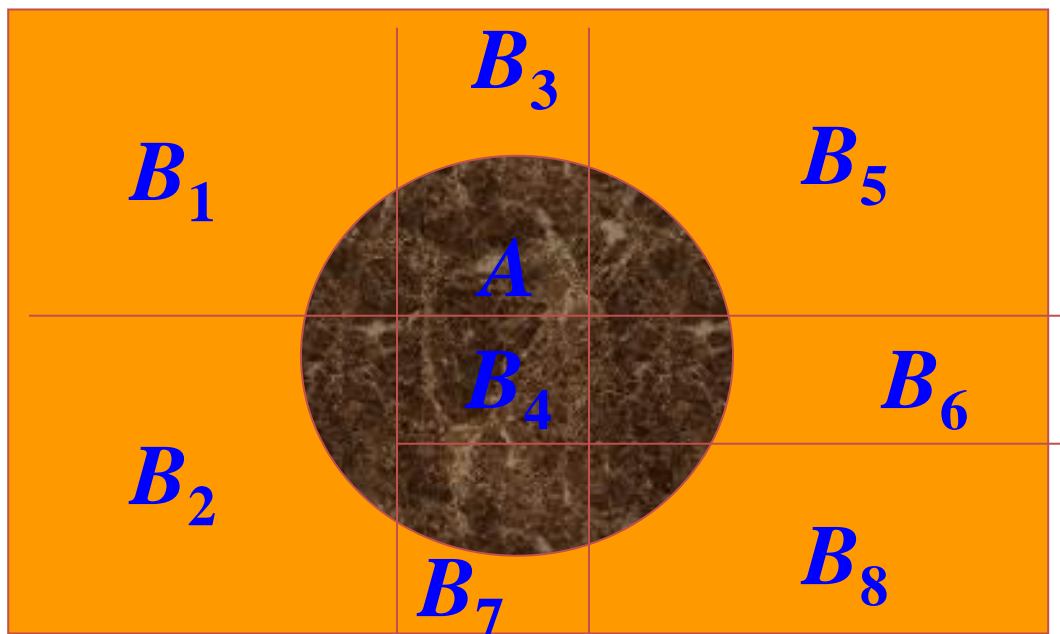




# 贝叶斯概率

诸 $B_i$ 是**原因**  
 $A$ 是**结果**

由此可以形象地把全概率公式看成为“**由原因推结果**”，每个原因对结果的发生有一定的“作用”，即结果发生的可能性与各种原因的“作用”大小有关。全概率公式表达了它们之间的关系。



**(5)贝叶斯公式。** 贝叶斯公式也叫**后验概率公式**，亦叫**逆概率公式**，其用途很广。设先验概率为 $P(B_i)$ ，调查所获的新附加信息为 $P(A_j|B_i)$  ( $i=1,2,\dots,n; j=1,2,\dots,m$ )，则贝叶斯公式计算的后验概率为

$$P(B_i|A_j) = P(B_i)P(A_j | B_i) / \sum_{k=1}^n P(B_k)P(A_j | B_k) \quad (5)$$

该公式于1763年由贝叶斯(Bayes)给出。它是在观察到事件A已发生的条件下，寻找导致A发生的每个原因的概率。

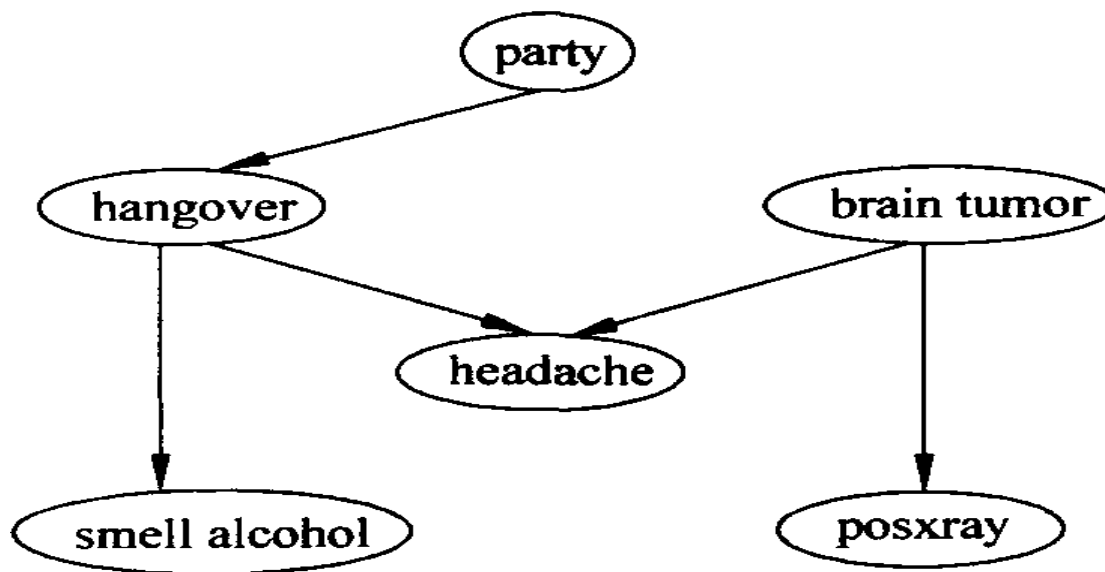


# 15.3 贝叶斯网络



- 贝叶斯网络由**网络结构**和**条件概率表**组成

- 网络结构是一个有向无环图
- 每个结点代表一个事件或者随机变量
- 结点的取值是完备互斥的





- **对已有信息要求低, 可以在信息不完全情况下推理**
  - 采用条件独立性假设
- **具有良好的逻辑性和可理解性**
  - 贝叶斯网络是白盒, 神经元模型是黑盒
- **可以与专家知识深度融合, 避免过学习**
  - 突出主要矛盾
- **推理可解释性性强**



- **贝叶斯网络预测**

- 贝叶斯网络的预测是指从起因推测一个结果的推理，也称为由顶向下的推理
- 原因推导出结果，求出由原因导致的结果发生的概率

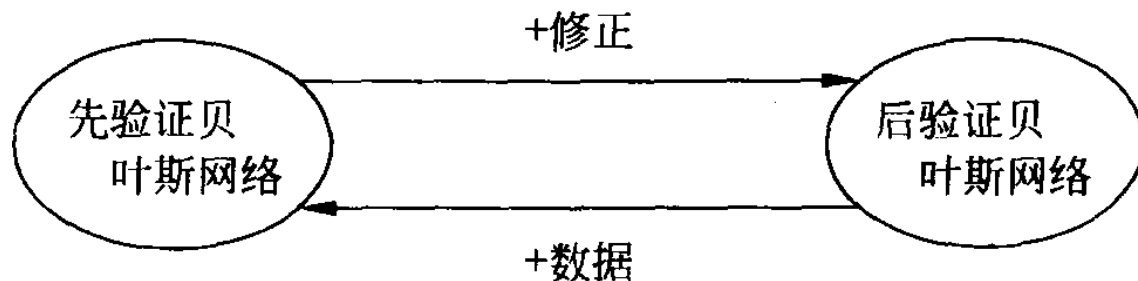
- **贝叶斯网络诊断**

- 贝叶斯网络的诊断是指从结果推测一个起因的推理，也称为由底至上的推理。
- 已知发生了某些结果，根据贝叶斯网络推理计算造成该结果发生的原因和发生的概率



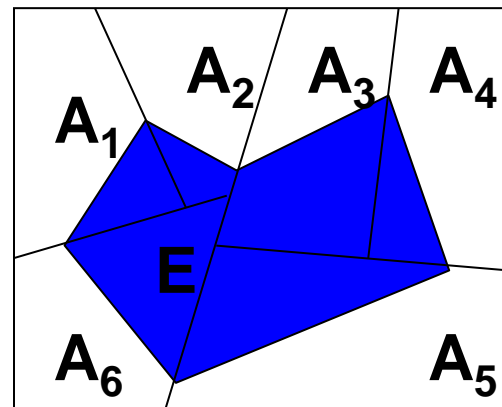
## • 贝叶斯网络学习

- 贝叶斯网络学习是指由先验的贝叶斯网络得到后验的贝叶斯网络的过程
- 贝叶斯网络学习的实质是用现有数据对先验知识的修正
- 贝叶斯网络可以持续迭代修正
- 上次学习得到的后验贝叶斯网络变成下一次学习的先验贝叶斯网络
- 结构学习：确定最合适的贝叶斯网络模型结构
- 参数学习：在给定结构下，确定贝叶斯网络模型的参数，即每个结点上的条件概率分布表（CPT表）。



$$p(A | B) = \frac{p(A, B)}{p(B)} = \frac{p(B | A)p(A)}{p(B)}$$

$$p(A_i | E) = \frac{p(E | A_i)p(A_i)}{p(E)} = \frac{p(E | A_i)p(A_i)}{\sum_i p(E | A_i)p(A_i)}$$



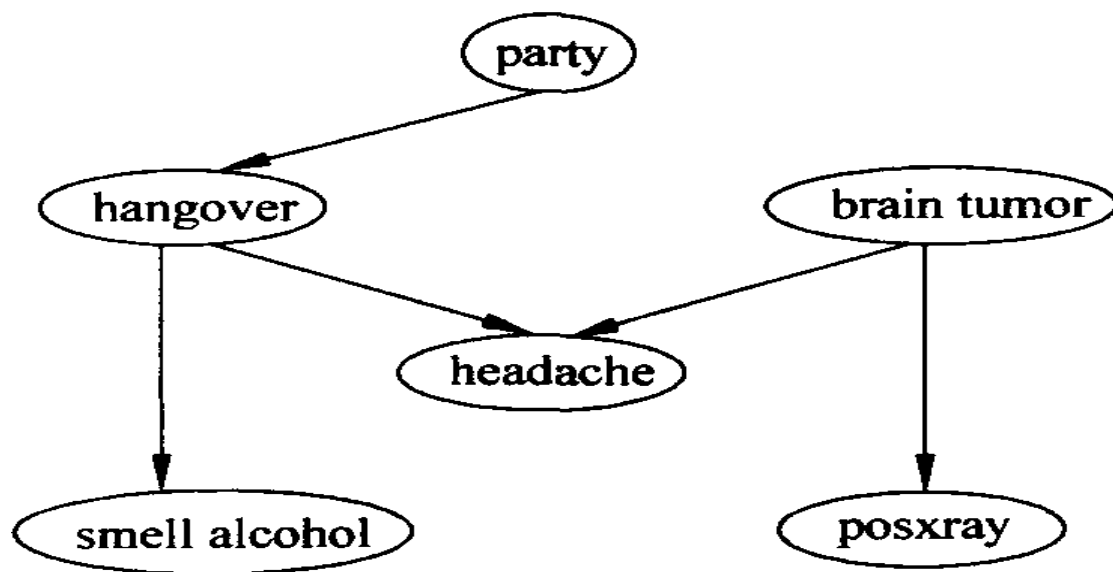
- 基于条件概率的定义
- $p(A_i|E)$  是在给定证据下的后验概率
- $p(A_i)$  是先验概率
- $P(E|A_i)$  是在给定 $A_i$ 下的证据似然
- $p(E)$  是证据的预定义后验概率



# 贝叶斯网络的预测、诊断和训练算法

将从图15.1所示的简单贝叶斯网络的例子入手，分别介绍贝叶斯网络的预测、诊断和训练算法。

假定网络中的概率和条件概率都已经知道，也就是说网络已经训练完毕。这些数据给出如下。



参加晚会 (party, PT)、宿醉 (hangover, H0)、患脑瘤 (brain tumor, BT)、头疼 (headache, HA)、有酒精味 (smell alcohol, SA) 和X射线检查呈阳性 (posxray, PX)

# 概率和条件概率数据

图15. 1中的Party和Brain Tumor两个结点是原因结点，没有连线以它们作为终点。首先给出这两个结点的无条件概率，如表15. 2所示。

表15. 2结点PT、BT的无条件概率分布

	$P(PT)$	$P(BT)$
True	0.200	0.001
False	0.800	0.999

表15. 2中的第二列是关于Party(参加晚会)的概率：参加晚会的概率是0.2，不参加晚会的概率是0.8。第三列是关于患脑瘤的概率：患脑瘤的概率是0.001，不患脑瘤的概率是0.999。

下面还将给出几组条件概率，分别是PT已知的情况下HO的条件概率，如表15. 3所示；HO已知的情况下SA的条件概率，如表15. 4所示；BT已知的情况下PX的概率，如表15.5所示。

表15.3 已知结点PT时HO的条件概率

$P(HO PT)$	$PT = \text{True}$	$PT = \text{False}$
True	0.700	0
False	0.300	1.000

表15.4 已知结点HO时SA的条件概率

$P(SA HO)$	$HO = \text{True}$	$HO = \text{False}$
True	0.800	0.100
False	0.200	0.900

表15.5 已知结点BT时PX的条件概率

$P(PX BT)$	$BT = \text{True}$	$BT = \text{False}$
True	0.980	0.010
False	0.020	0.990

上面三个表的结构相似，给出的都是条件概率。表15.3中第2列的意思是：当参加晚会后，宿醉的概率是0.7；不宿醉的概率是0.3。第3列的意思是：当不参加晚会后，不会发生宿醉的情况。

对表15.4和表15.5的解释类似。

最后给出的是一个联合条件概率：已知HO和BT时HA的概率，如表15.6所示。

表15.6 已知HO和BT时HA的概率

$P(HA   HO, BT)$	HO = True		HO = False	
	BT = True	BT = False	BT = True	BT = False
True	0.990	0.700	0.900	0.020
False	0.010	0.300	0.100	0.980

表15.6中数据的意义是：

当宿醉发生和有  
脑瘤的情况下，  
头疼的概率是  
0.99，不头疼  
的概率是0.01

当宿醉发生但  
没有脑瘤的情  
况下，头疼的  
概率是0.7，  
不头疼的概率  
是0.3。

当没有宿醉但患  
有脑瘤的情况下  
，头疼的概率是  
0.9，不头疼的  
概率是0.01。

。



# 贝叶斯网络预测算法

为了方便，约定：对于一个结点Point， $P(+Point)$ 表示Point发生的概率， $P(-Point)$ 表示不发生的概率。

**【例15.1】**下面计算结点HA的概率。

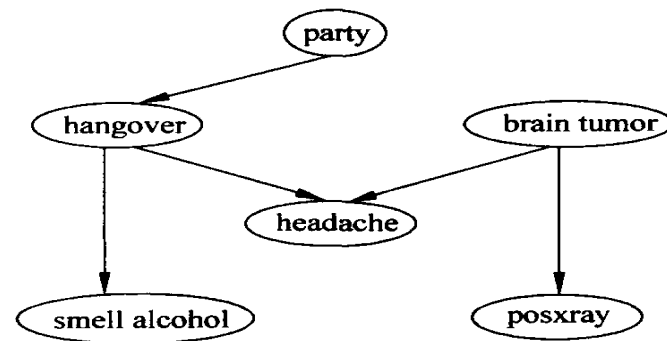
$$\begin{aligned} P(+H0) &= P(+H0|PT) * P(PT) + P(+H0|-PT) * P(-PT) \\ &= 0.2 * 0.7 + 0 = 0.14 \end{aligned}$$

$$P(-H0) = 0.86$$

根据全概率公式，有

$$\begin{aligned} P(+HA) &= P(+BT)P(+HO)*0.99 + P(+BT)P(-HO)*0.9 + P(-BT)P(+HO) \\ &\quad *0.7 + P(-BT)P(-HO)*0.02 \end{aligned}$$

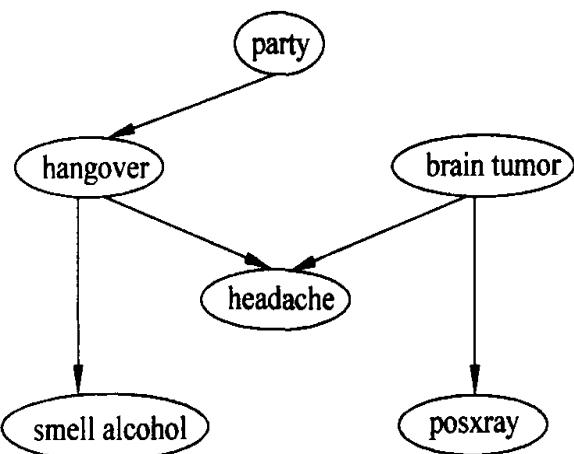
$$\begin{aligned} &= 0.001*0.14*0.99 + 0.001*0.86*0.9 + 0.999*0.14*0.7 + 0.999*0.86*0.02 \\ &= 0.116 \end{aligned}$$



**【例15.2】** 计算已知参加晚会的情况下，第二天早晨呼吸有酒精味的概率。

首先，由表7.3可以看出，当PT发生时，H0发生的概率是0.7。也就是说，当参加晚会后，宿醉发生的概率是0.7，不发生的概率是0.3。由**全概率公式**

$$\begin{aligned}
 P(+SA) &= P(+H0) P(+SA | +H0) + P(-H0) P(+SA | -H0) \\
 &= 0.7 \times 0.8 + 0.3 \times 0.1 \\
 &= 0.59
 \end{aligned}$$



$P(H0 PT)$	$PT= True$	$PT= False$
True	0.700	0
False	0.300	1.000

表15.4 已知结点  $HO$  时  $SA$  的条件概率

$P(SA HO)$	$HO= True$	$HO= False$
True	0.800	0.100
False	0.200	0.900

表15.5 已知结点  $BT$  时  $PX$  的条件概率

$P(PX BT)$	$BT= True$	$BT= False$
True	0.980	0.010
False	0.020	0.990

【例15.3】计算已知参加晚会的情况下，头疼发生的概率。

由表7.3可知，当PT发生时，H0发生的概率是0.7，不发生的概率是0.3；由表7.2可以看出，BT发生的概率是0.001，不发生的概率是0.999。已知H0和BT后，根据全概率公式，得到

$$\begin{aligned} P(+HA) &= P(+H0)P(+BT)P(+HA \mid +H0+BT) + P(+H0)P(-BT) \\ &\quad \times 0.7 + P(-H0)P(+BT) \times 0.9 + P(-H0)P(-BT) \times 0.02 \\ &= 0.7 \times 0.001 \times 0.99 + 0.7 \times 0.999 \times 0.7 + 0.3 \times 0.001 \times 0.9 + 0.3 \times 0.999 \times 0.02 \\ &= 0.496467 \end{aligned}$$

$$P(-HA) = 1 - P(+HA) = 0.503533$$

也就是说，如果知道已经参加了晚会，而没有其他方面的任何证据，则这个人头疼的概率是0.496，不头疼的概率是0.504。

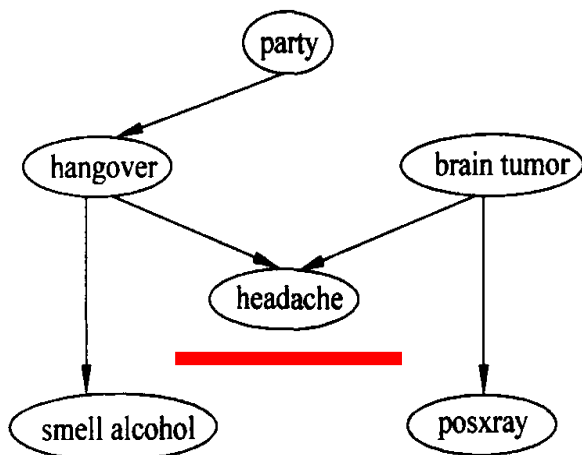


表15.3

$P(H0 PT)$	$PT = \text{True}$	$PT = \text{False}$
True	0.700	0
False	0.300	1.000

表15.6

$P(HA HO,BT)$	$HO = \text{True}$		$HO = \text{False}$	
	$BT = \text{True}$	$BT = \text{False}$	$BT = \text{True}$	$BT = \text{False}$
True	0.990	0.700	0.900	0.020
False	0.010	0.300	0.100	0.980

表15.5 已知结点BT时PX的条件概率

$P(PX BT)$	$BT = \text{True}$	$BT = \text{False}$
True	0.980	0.010
False	0.020	0.990

比较分析例15. 1和例15. 3的结果：由于参加晚会，头疼发生的概率大大增加了。

结合上面给出的三个例子，下面给出**贝叶斯网络预测算法的步骤**描述。如下所示。

**输入：**给定贝叶斯网络 $B$  (包括网络结构 $m$ 个结点以及某些结点间的连线、原因结点到中间结点的条件概率或联合条件概率)，给定若干个原因结点发生与否的事实向量 $F$  (或者称为证据向量)，给定待预测的某个结点 $t$ 。

**输出：**结点 $t$ 发生的概率。

(1) 把证据向量输入到贝叶斯网络 $B$ 中。

(2) 对于 $B$ 中的每一个没处理过的结点 $n$ ，如果它具有发生的事实(证据)，则标记它为已经处理过；否则继续下面的步骤。

(3) 如果它的所有父结点中有一个没有处理过，则不处理这个结点；否则，继续下面的步骤。

(4) 根据结点 $n$ 的所有父结点的概率以及条件概率或联合条件概率计算结点 $n$ 的概率分布，并把结点 $n$ 标记为已处理。

(5) 重复步骤(2) ~ (4)，共 $m$ 次。此时，结点 $f$ 的概率分布就是它的发生 / 不发生的概率。算法结束。

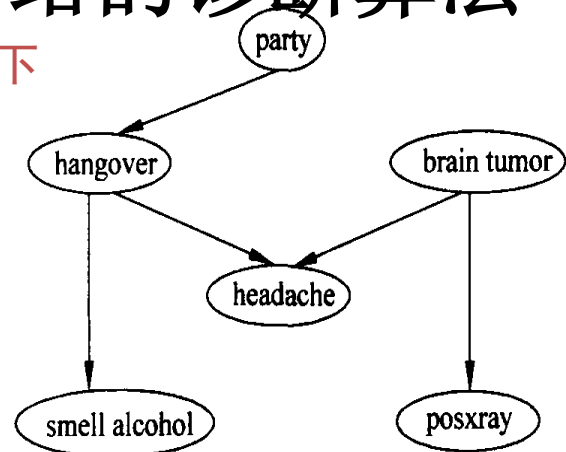
需要注意的是，第(5)步的作用是使得每个结点都有被计算概率分布的机会。



# 贝叶斯网络的诊断算法

本部分将做相反方向的工作：在已知结果结点发生与否的情况下推断条件结点发生的概率。

【例15.4】计算已知X光检查呈阳性的情况下，患脑瘤的概率。



根据条件概率公式

$$P(+BT \mid +PX) = P(+PX \mid +BT) \times P(+BT) / P(+PX)$$

$$= 0.98 \times 0.001 / 0.011 = 0.08909$$

$$P(-BT \mid +PX) = 1 - P(+BT \mid +PX) = 0.911$$

也就是说，当X光检查呈阳性的情况下，患脑瘤的概率是0.089，不患脑瘤的概率是0.911。

先验概率  
(0.011)

上面的例子比较简单，可以直接用条件概率公式计算获得。下面再看一个比较复杂的例子。

$P(HO \mid PT)$	$PT = \text{True}$	$PT = \text{False}$
True	0.700	0
False	0.300	1

$P(PX \mid BT)$	$BT = \text{True}$	$BT = \text{False}$
True	0.980	0.010
False	0.020	0.990

表15.5 已知结点 BT 时 PX 的条件概率

由：  $P(AB) = P(A \mid B) \times P(B)$  得到：

$$P(A \mid B) = P(AB) / P(B)$$

而：  $P(AB) = P(B \mid A) \times P(A)$

所以：  $P(A \mid B) = P(AB) / P(B)$

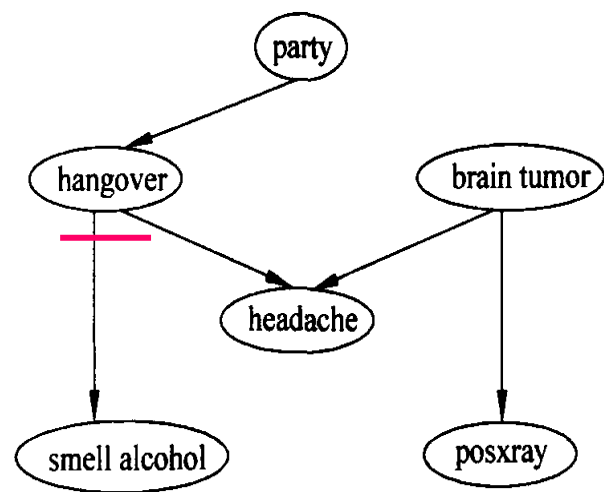
$$= P(B \mid A) \times P(A) / P(B)$$

【例15.5】计算已知头疼的情况下，患脑瘤的概率。

首先，根据表15.6给出的联合条件分布计算已知BT情况下HA的边缘条件概率。为此，要首先计算结点H0的概率分布。根据表15.3和全概率公式

$$\begin{aligned} P(+H0) &= P(+H0 \mid +PT) \times P(+PT) + P(+H0 \mid -PT) \times P(-PT) \\ &= 0.7 \times 0.2 + 0 \\ &= 0.14 \end{aligned}$$

$P(H0 PT)$	$PT = \text{True}$	$PT = \text{False}$
True	0.700	0
False	0.300	1.000



上面的计算表明，没有任何证据的情况下，宿醉发生的概率是0.14，不发生的概率是0.86。通过宿醉的发生概率，可以计算已知BT情况下HA的边缘条件概率。

$$\begin{aligned} P(+HA \mid +BT) &= P(+H0) \times P(+HA \mid +BT, +H0) \\ &\quad + P(-H0) \times P(+HA \mid +BT, -H0) \\ &= 0.14 \times 0.99 + 0.86 \times 0.9 \\ &= 0.9126 \end{aligned}$$

全概率公式

$$P(-HA \mid +BT) = 1 - P(+HA \mid +BT) = 0.087$$

上面的计算得到了已知患脑瘤的情况下头疼的概率是0.913，不头疼的概率是0.087。这个条件概率是一个边缘分布，它是从联合条件概率分布( $H0, BT \rightarrow HA$ )去掉一个条件 $H0$ 得到的。我们把这个边缘分布的内容整理在表7.7中。

表15.7 已知BT情况下HA的(边缘)条件概率

$P(HA \mid BT)$	$BT = \text{True}$	$BT = \text{False}$
True	0.913	0.115
False	0.087	0.885

最后，根据表15.7提供的条件概率，利用条件概率公式，可得

$$\begin{aligned} P(+BT \mid +HA) &= P(+HA \mid +BT) \times P(+BT) / P(+HA) \\ &= 0.9126 \times 0.001 / 0.116 \\ &= 0.007867 \end{aligned}$$

表15.6

$P(HA \mid HO, BT)$	$HO = \text{True}$		$HO = \text{False}$	
	$BT = \text{True}$	$BT = \text{False}$	$BT = \text{True}$	$BT = \text{False}$
True	0.990	0.700	0.900	0.020
False	0.010	0.300	0.100	0.980

例15.4和例15.5分别从简单和复杂两种情况进行了贝叶斯网络的诊断示例。下面的部分将介绍同时具有预测功能和诊断功能的算法。

根据上面的两个例子，可以总结出**贝叶斯网络诊断算法的一般步骤**，如下所示。

**输入：**给定贝叶斯网络 $B$ ，给定若干个结果结点发生与否的事实向量 $F$ （或者称为证据向量），给定待诊断的某个结点 $t$ 。

**输出：**结点 $t$ 发生的概率。

(1) 把证据向量输入到贝叶斯网络 $B$ 中。

(2) 对于 $B$ 中的每一个没处理过的结点 $n$ ，如果它具有发生的事实（证据），则标记它为已经处理过；否则继续下面的步骤。

(3) 如果它的所有子结点中有一个没有处理过，则不处理这个结点；否则，继续下面的步骤。

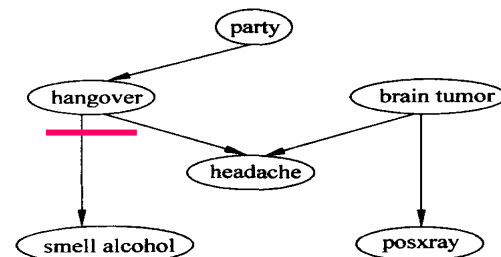
(4) 根据结点 $n$ 的所有子结点的概率以及条件概率或联合条件概率计算结点 $n$ 的概率分布，并把结点 $n$ 标记为已处理。

(5) 重复步骤(2)～(4)共 $m$ 次。此时，原因结点 $t$ 的概率分布就是它的发生 / 不发生的概率。算法结束。

需要注意的是，第(5)步的作用是使得每个结点都有被计算概率分布的机会。

# 贝叶斯网络预测和诊断的综合算法

利用贝叶斯网络进行单纯的预测或进行单纯的诊断的情况是比较少的，一般情况下，需要综合使用预测和诊断的功能。



【例15.6】计算已知参加晚会并且第二天早上呼吸有酒精味的情况下，宿醉的发生概率。

由于已知参加了晚会(+PT)，那么根据表7.3，宿醉发生的概率是0.7，不发生的概率是0.3。根据全概率公式

$$\begin{aligned}
 P(+SA) &= P(+SA \mid +H0) \times P(+H0) + P(+SA \mid -H0) \times P(-H0) \\
 &= 0.8 \times 0.7 + 0.1 \times 0.3 = 0.59
 \end{aligned}$$

这个结果就是已知参加晚会的情况下，有酒精味的发生概率。再利用条件概率公式，可得

$$\begin{aligned}
 P(+H0 \mid +SA) &= P(+SA \mid +H0) \times P(+H0) / P(+SA) \\
 &= 0.8 \times 0.7 / 0.59 = 0.94915
 \end{aligned}$$

这是最终的结果，也就是说，当参加晚会并且第二天早晨有酒精味的情况下，宿醉发生的概率是0.949。

从上面的计算过程可以总结出解决这类综合问题的一般思路。首先，要把原因结点的证据(此例中是+PT)进行扩散，得到中间结点(H0)或结果结点(SA)的概率分布。最后根据条件概率公式计算中间结点的概率分布。

这是解决预测和诊断综合问题的一般思路，下面将给出一个更复杂的综合问题的例子。

【例15.7】计算在已知有酒精味、头疼的情况下，患脑瘤的概率。

首先，由条件概率公式可以计算在有酒精味的情况下宿醉的发生概率

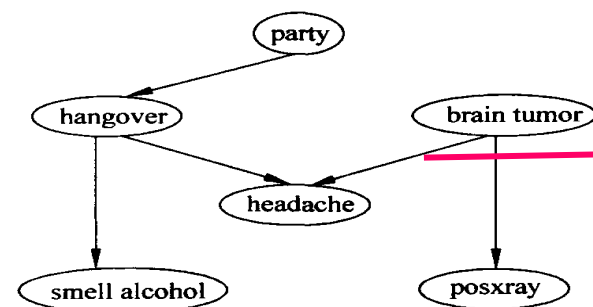
$$P(+H0 \mid +SA) = P(+SA \mid +H0) \times P(+H0) / P(+SA) = 0.5656$$

然后，由全概率公式可以计算患脑瘤的情况下头疼的发生概率（当然，这时宿醉的概率已经是0.5656，它参与了下面的运算）

$$\begin{aligned} P(+HA \mid +BT) &= P(+HA \mid +BT, +H0) \times P(+H0) \\ &+ P(+HA \mid +BT, -H0) \times P(-H0) \\ &= 0.99 \times P(+H0) + 0.9 \times P(-H0) = 0.9509 \end{aligned}$$

最后，再由条件概率公式可以计算患脑瘤的概率

$$\begin{aligned} P(+BT \mid +HA) &= P(+HA \mid +BT) \times P(+BT) / P(+HA) \\ &= 0.9509 \times 0.001 / 0.4052 \\ &= 0.002347 \end{aligned}$$



可以比较例15.7和例15.5的计算结果，例15.7中计算得到的患脑瘤的概率要相对小一些。同样患有头疼，两个例子中患脑瘤的概率是不一样的。这是因为，例15.7中的结果结点“有酒精味”发生，这意味着头疼的原因有更大的可能是因为宿醉，而不是患脑瘤。

除了上面的7个例子外，读者可以试着解决图15.1所示贝叶斯网络中更复杂的例子，或者解决本章后面的习题。

# 贝叶斯网络的建立和训练算法

如何构建贝叶斯网络？

1. 构建节点

2. 构建边

3. 构建CPT表

结点的概率分布和结点间的条件概率分布可以通过专家经验填入，但使用更多的方法是通过历史数据训练得到。贝叶斯网络的训练方式比较简单。作为示例，下面给出图15. 1所示的6个结点发生的历史数据，如表15. 8所示。

表15. 8贝叶斯网络中的历史数据

序号	<i>PT</i>	<i>HO</i>	<i>BT</i>	<i>HA</i>	<i>SA</i>	<i>PX</i>
1	1	1	0	1	1	0
2	0	0	1	1	0	1
3	1	0	0	0	0	0
4	1	1	1	0	1	1
5	0	0	0	0	0	1
6	1	1	0	1	1	0
7	1	0	1	0	1	0
8	0	0	1	0	0	0
9	1	0	0	0	1	0
10	1	1	1	1	1	1

根据表15. 8给出的数据，可以用统计的方式得到任意结点的概率分布。假设结点P有m个状态P1, P2, ..., Pm, 则有

$$P(P_m) = \frac{P_m \text{ 出现的数据条数}}{\text{总的数据条数}} \quad (7-8)$$



例如，对于结点PT，有 $P(+PT)=7 / 10=0.7$ ； $P(-PT)=3 / 10=0.3$ 。如果PS表示结点P的一个状态，QS表示结点Q的一个状态，则PS发生时QS发生的概率为

$$P(QS | PS) = \frac{PS \text{ 和 } QS \text{ 共同发生的次数}}{PS \text{ 发生的次数}}$$

例如，+PT共发生了7次，+PT和+H0共同发生了4次，因此有 $P(+H0 | +PT)=4 / 7=0.571$ 。同样的方式也可以计算 $P(QS | \sim PS)$ 。

同理，可以计算多个结点间的联合条件分布。假设PS表示结点P的一个状态，QS表示结点Q的一个状态，RS表示结点R的一个状态。那么PS和QS发生时RS的概率为

$$P(RS | PS, QS) = \frac{PS, QS, RS \text{ 共同发生的次数}}{PS \text{ 和 } QS \text{ 共同发生的次数}}$$

如果结点P、Q、R各有两个状态，那么类似以上形式的公式共有8个，共同构成了结点P、Q到结点R的联合条件概率分布。

例如，+H0和+BT共发生了2次，而+H0、+BT和+HA共发生了1次，因此 $P(+HA | +H0, +BT)=1 / 2=0.5$

如果某个结点是结果结点或中间结点，那么得到这个结点的概率分布的方式有如下两种。

(1) 直接从表15.8所示的数据中通过统计获得。

(2) 先从表格数据中通过统计获得原因结果的概率分布，再从表格数据中通过统计获得条件概率分布或联合条件概率分布，最后用全概率公式计算中间结点或结果结点的概率分布。

可以验证，这两种方式获得的概率分布是一致的。