COMP9414: 人工智能第5a讲：不确定性

韦恩-沃布克

电邮：w.wobcke@unsw.edu.au

用不确定性进行推理

- 代理人并不总是能够确定所有命题的真实性，因此可能不仅有"平白无故"的信念（*P*或¬P）。
- 由于不可预测性或非确定性，一些环境本身对代理人产生了不确定性，所以命题对这些环境的建模是不充分的。

本讲座

- 不确定性
- 概率论
- 条件概率和贝叶斯法则
- 贝叶斯网络
  - 贝叶斯网络的语义学
  - 贝叶斯网络中的推理

逻辑方法的问题

- 代理人的理性决策需要在目标的重要性和实现目标的可能性，以及行动和不实现目标的成本之间进行权衡。

- 考虑尝试将医疗诊断系统正规化

$$p\forall(Symptom(p, AbdominalPain) \rightarrow Disease(p, Appendicitis))$$

- 这一规则是不正确的，因为腹痛的病人可能患有其他疾病。

$$p\forall(Symptom(p, AbdominalPain) \rightarrow$$

$$疾病(p, 阑尾炎)\vee 疾病(p, 溃疡)\vee 疾病(p, 印第安)\cdots)$$

- 因果规则如何？

$$p\forall(Disease(p, Ulcer) \rightarrow Symptom(p, AbdominalPain))$$

新南威尔士大学
©W.Wobcke等人，2019-2022年

不确定性的来源

- 由于不完全性，逻辑方法会出现困难：代理人可能没有完整的领域理论 无知：代理人可能没有足够的领域信息
  噪声：代理人拥有的信息可能不可靠
  非确定性：环境本身可能是随机的
  不可预测性：环境可能是固有的不可预测的
- **概率**给出了一种总结这种不确定性的方法
  - 例如，代理人认为，如果病人有腹痛，患阑尾炎的概率为0.75。

这些数字意味着什么？

统计学/常识分子观点

一组 "事件 "的长距离频率，例如，抛硬币时出现 "头 "的事件的概率=抛硬币时出现 "头 "的长距离频率

客观看法

概率是世界的真实方面--客观的

个人/主观/贝叶斯观点

基于代理人的知识对命题的信念的衡量，例如，人头的概率是基于对硬币的信念而相信硬币会落在人头上的程度，也可

样品空间和活动

- 抛掷三次硬币
- 可能的结果是

  タイトTTH THT THH
  HTT HTH HHT HHH
- 所有可能结果的集合

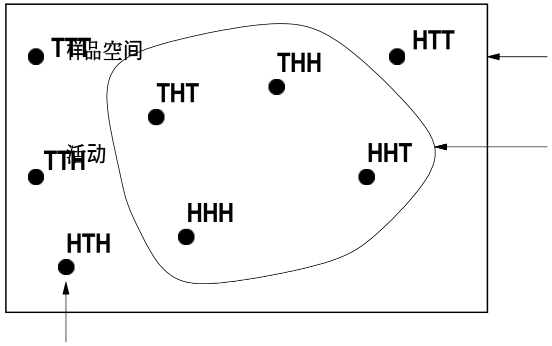
  s={ttt, tth, tht, thh, htt, hth, hht, hhh}。
- 样本空间的任何子集被称为事件
- 样本空间的任何单子集被称为简单事件

样品空间和活动

能只是一种猜测；不同的

代理人可能会分配不同的概率--主观的。

简单事件



之前的概率

- $P(A)$是一个事件A发生的先验或无条件的概率
- 例如，P（*阑尾炎*）=0.3
- **在没有任何其他信息的情况下，代理人**认为病人患有阑尾炎的概率为0.3（30%）。
- 为了说明新信息对概率的影响，代理人必须用条件概率进行推理以更新概率

随机变量

- 命题是随机变量，可以有几个值
  P（*天气=晴天*）=0.8 P（
  天气=雨天）=0.1 P（
  天气=阴天）=0.09 P（
  天气=雪天）=0.01
- 每个随机变量X*都*有一个可能的值域
  $(x_1, x_2, - - , x_n)$
- 所有可能值的概率 P(*Weather*) = (0.8, 0.1, 0.09, 0.01)
  是一个概率分布
- P(*Weather*,
  Appendicitis)是随机变量的组合，由交叉乘积表示(也可以

概率论的公理

1. $0 \leq P(A) \leq 1$
   - 所有的概率都在0和1之间
2. P(*真*)=1 $P(False) = 0$
   - 有效命题的概率是1
   - 不能满足的命题的概率为0
3. $p(a \vee b) = p(a) + p(b) - p(a \wedge b)$

- **可以确定所有其他命**题的概率
- 例如，$P(A \vee \neg A) = P(A) + P(\neg A) - P(A \wedge \neg A)$
  $P(True) = P(A) + P(\neg A) - P(False)$
  $1 = p(a) + p(\neg a) - 0$
  因此，P（¬A）=1−P（A）。

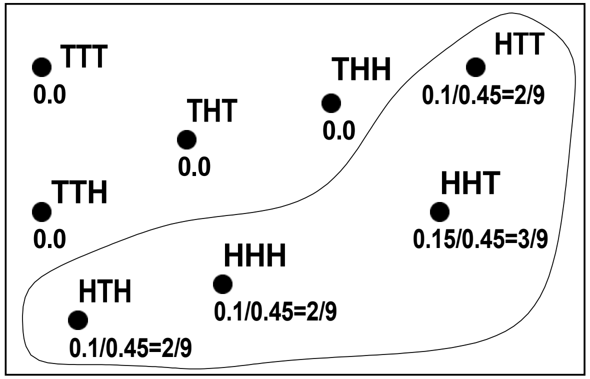
条件概率

- 需要根据新的信息来更新概率
- 使用条件或后验概率
  $P(A|B)$是指在我们只知道B*的情况*下，A的概率。
  例如，P（*阑尾炎|腹痛*）=0.75
- 定义。$P(A|B)=\frac{P(A \wedge B)}{P(B)}$前提是$P(B)>0$
- 产品规则。P（$A \wedge B$）=P（$A|B$）。$P(B)$
- $P(X|Y) = P(X = x_i | Y = y_j)$ 对于所有的i, j
  用逻辑连接词$P(A \wedge B)$来表示复合事件)

$\mathrm{P}(x, y) = \mathrm{P}(x \mid y)$。$\mathrm{P}(Y)$--一组方程

新南威尔士大学 ©W.Wobcke等人，2019-2022年

正常化



- 鉴于第一枚硬币是H的条件概率分布

联合概率分布

- 对领域内所有事件的概率的完整说明
- 假设随机变量X_1，X_2，- - ，X_n
- 一个原子（简单）事件是对所有变量的赋值
- 联合概率分布P(X_1，X_2，- - ，X_n)为所有可能的原子事件分配概率
- 具有两个布尔随机变量的简单医疗领域

	腹痛	¬腹痛
阑尾炎	0.04	0.06
¬阑尾炎	0.01	0.89

联合概率 分布

- 简单的事件是相互排斥和共同穷尽的
- 复杂事件的概率是兼容的简单事件的概率之和

$P(阑尾炎) = 0.04 + 0.06 = 0.10$

$P(Appendicitis \vee AbdominalPain) = 0.04 + 0.06 + 0.01 = 0.11$

$P(Appendicitis|AbdominalPain)$

$= \frac{P(Appendicitis \wedge AbdominalPain)}{P(腹痛)} = \frac{0.04}{0.04+0.01} = 0.8$

- 问题：在许多变量中，概率的数量是巨大的

联合概率分布

假设在三个随机变量*牙痛*、*蛀牙*和*接球*上有一些潜在的联合概率分布，我们可以用表格的形式写出来

	牙痛		*牙痛*	
	接住	*接住*	*接住*	*接住*
空腔	.108	.012	.072	.008
空腔	.016	.064	.144	.576

请注意，表中的条目之和为1.0。

对于任何命题，将其为真的原子事件相加

枚举推理的例子

	牙痛		牙痛	
	接住	接住	接住	接住
空腔	.108	.012	.072	.008
空腔	.016	.064	.144	.576

对于任何一个命题，将其为真的原子事件相加。

P（*牙痛*）= 0.108 + 0.012 + 0.016 + 0.064 = 0.2

枚举推理的例子

	牙痛		牙痛	
	接住	接住	接住	接住
空腔	.108	.012	.072	.008
空腔	.016	.064	.144	.576

对于任何一个命题，将其为真的原子事件相加。

P(*蛀牙* ∨ *牙痛*)

= 0.108 + 0.012 + 0.072 + 0.008 + 0.016 + 0.064 = 0.28

枚举的条件概率

	牙痛		牙痛	
	接住	接住	接住	接住
空腔	.108	.012	.072	.008
空腔	.016	.064	.144	.576

$$P(\neg\text{蛀牙}|\text{牙痛}) = \frac{P(\neg\text{蛀牙}\wedge\text{牙痛})}{P(\text{牙痛})}$$

$$\frac{0.016 + 0.064}{0.108 + 0.012 + 0.016 + 0.064} = 0.4$$

贝叶斯法则

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

- **人工智能系**统放弃了联合概率，直接使用贝叶斯法则的条件概率工作
- 推导出贝叶斯规则。

  $p(a\wedge b)=p(a|b)p(b)$ (定义)

  $p(b\wedge a)=p(b|a)p(a)$ (定义)

  所以$P(A|B)P(B)=P(B|A)P(A)$ 因为$P(A\wedge B)=P(B\wedge A)$

  因此，如果$P(A) /= 0$，则$P(B|A)=\frac{P(A|B)P(B)}{P(A)}$

■ 注：如果$P(A)=0$，则$P(B|A)$是未定义的。

新南威尔士大学

©W.Wobcke等人，2019-2022年

应用贝叶斯规则

- 例子(Russell & Norvig, 1995)
- **医生知道**
  - 脑膜炎有50%的时间会导致脖子僵硬
  - 病人患脑膜炎的机会是$\frac{1}{50000}$
  - 病人出现颈部僵硬的机会是$\frac{1}{20}$
- $P(StiffNeck|Meningitis) = 0.5$
  $P(\text{脑膜炎}) = \frac{1}{50000}$
  $P(StiffNeck) = \frac{1}{20}$
  $P(Meningitis|StiffNeck) = \frac{P(StiffNeck|Meningitis)\text{。}P(\text{脑膜炎})}{P(StiffNeck)}$
  $= 0.5\frac{1}{50000}\frac{1}{\frac{1}{20}} = 0.0002$

正常化

- 避免对症状进行评估
  $P(Meningitis|StiffNeck) = \frac{P(StiffNeck|Meningitis)\text{。}P(\text{脑膜炎})}{P(StiffNeck)}$
  $P(\neg\text{脑膜炎}|\text{僵硬颈椎}) = \frac{P(\text{僵硬颈椎}|\neg\text{脑膜炎})\text{。}P(\neg\text{脑膜炎})}{P(StiffNeck)}$
- $P(StiffNeck) = P(StiffNeck|Meningitis)\text{。}P(\text{脑膜炎})$
  $+ P(StiffNeck|\neg Meningitis)\text{。}P(\neg\text{脑膜炎})$
- 所以P（$\text{脑膜炎}|\text{硬颈}$）$=$
  $\frac{P(StiffNeck|Meningitis)\text{。}P(\text{脑膜炎})}{P(StiffNeck|Meningitis)\text{。}P(Meningitis)+P(StiffNeck|\neg Meningitis)\text{。}P(\neg\text{脑膜炎})}$
  同样，对于$P(\neg\text{脑膜炎}|\text{颈椎病})$也是如此。
- 因此，从这两个$P(StiffNeck|---)$可以推导出$P(StiffNeck)$

使用贝叶斯法则

- **假**设阑尾炎有两个条件概率
  $P(Appendicitis|AbdominalPain) = 0.8$
  $P(Appendicitis|Nausea) = 0.1$
- $P(Appendicitis|AbdominalPain \wedge Nausea)$
  $\frac{P(AbdominalPain \wedge Nausea|Appendicitis)\text{。}P(\text{阑尾炎})}{P(\text{腹痛} \wedge \text{恶心})}$
- 需要知道$P(AbdominalPain \wedge Nausea|Appendicitis)$
- 对于许多症状来说，这是一项艰巨的任务……。

有条件的独立

而分母是一个归一化系数

- 阑尾炎是导致腹痛和恶心的直接原因
- **如果我**们知道一个病人患有阑尾炎，恶心的概率不应该取决于腹痛的存在；同样，腹痛的概率也不应该取决于恶心。
- 鉴于阑尾炎，恶心和腹痛是有条件地独立存在的
- 如果知道Y不影响X的条件概率，那么事件X是独立于事件Y的，以背景知识K为条件。

$$p(x\,|\mathrm{k}) = p(x\,|\mathrm{y}, k)$$

贝叶斯网络

- 贝叶斯网络（也是贝叶斯信念网络、概率网络、因果网络、知识地图）是一个有向无环图（DAG），其中
  - 每个节点对应于一个随机变量

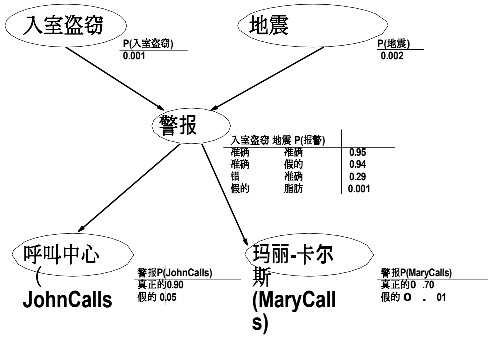
  有向链接连接成对的节点--一个来自节点的有向链接 *X*到节点*Y*意味着*X*对*Y*有直接影响
  - 每个节点都有一个条件概率表，量化了父母对节点的影响
- 贝叶斯网络的独立假设
  - 每个随机变量都是（有条件地）独立于其非后代的，因为它的父母都是独立的。

贝叶斯网络

- 例子（珍珠，1988）。
- **你家里有一个新的防盗**报警器，在探测小偷方面相当可靠，但有时也可能对地震作出反应。你还有两个邻居，约翰和玛丽，他们承诺在听到警报时给你打电话。约翰总是在他听到了警报，但有时会把电话铃声和警报混淆起来，然后再打电话，另外，玛丽喜欢大声的音乐，有时会错过警报。考虑到谁打过或没打过电话的证据，我们想估计一下发生入室盗窃的概率。

贝叶斯网络

- 例子（珍珠，1988）。



- 概率概括了潜在的无限可能的情况。

条件概率表

- 行中包含每个节点值在条件情况下的条件概率（父节点值的可能组合）。

入室盗窃	*地震*	P(*报警*|*入室盗窃*∧*地震*) 准确	错
真	真	0.950	0.050
真	假的	0.940	0.060
假的	真	0.290	0.710
假的	假的	0.001	0.999

贝叶斯网络的语义学

- 贝叶斯网络提供了对领域的完整描述
- 联合概率分布可以从网络中确定
  - $P(X_1, X_2, ---, X_n) = \prod_{i=1}^{n} P(X_i | \text{Parents}(X_i))$
- 例如，$P(J \wedge M \wedge A \wedge \neg B \wedge \neg E) =$
  $P(J|A).P(M|A).P(A|\neg B \wedge \neg E).P(\neg B).P(\neg E) =$
  $0.90 \times 0.70 \times 0.001 \times 0.999 \times 0.998 =$
  0.000628
- 贝叶斯网络是一个完整的、非冗余的领域表示（可以比联合概率分布紧凑得多）。

贝叶斯网络的语义学

- 联合概率分布的因子化
- 连锁规则。使用条件概率来分解连接词
  $P(X_1 \wedge X_2 \wedge \text{——} \wedge X_n) = P(X_1)$。$P(X_2 | X_1)$。$P(x_3 | x_1 \wedge x_2)$.- - - .$P(X_n$
  $|X \wedge\text{------})_1$
  $x_2 \wedge - - \wedge x_{n-1})$
- 现在，在贝叶斯网络中对变量$X_1, X_2, --, X_n$
  进行排序，使一个变量排在其父母之后--让π_{X_i}是变量X_i
  的父母的元组（这是一个复杂随机变量）。
  使用连锁规则，$P(X_1 \wedge X_2 \wedge --\wedge X_n) = P(X_1)$。$P(X_2 | X_1)$。$P(X_3$
  $|X_1 \wedge X_2)$。- - - .$P(X | X_{n1} \wedge X_2 \wedge - - \wedge X)_{n-1}$

贝叶斯网络的语义学

- 每个$P(X_i | X_1 \wedge X_2 \wedge - - \wedge X_{i-1})$的特性是它不是
  以X的一个后代为条件$_i$（考虑到贝叶斯网络中变量的排序）
- 因此，根据条件独立性
  - $P(X_i | X_1 \wedge X_2 \wedge - - \wedge X_{i-1}) = P(X_i | \pi_{X_i})$
- 重写给出了连锁规则
  - $p(x_1, x_2, - -, x_n) = \prod_{i=1}^{n} P(X_i | \pi_{X_i})$

使用贝叶斯网络进行计算

- 事实1：考虑随机变量X的父代$Y_1, Y_2, - -, Y_n$ $P(X | Y_1 \wedge - - \wedge$

  $$Y_n \wedge Z) = P(X | Y_1 \wedge - - \wedge Y_n)$$

  如果Z不涉及X的后代(包括X本身)
- 事实2：如果$Y_1, - - -, Y_n$ 是成对不相交的，并穷尽所有的可能性

  $$P(x) = \sigma p(x \wedge y_i) = \sigma p(x | y_i)。P(Y_i)$$

  $$p(x | z) = \sigma p(x \wedge y_i | z)$$

  - 例如，$P(J|B) = \underline{P(J \wedge B)} = \underline{\Sigma P(J \wedge B \wedge e \wedge a \wedge m)}$，其中j的范围为J、¬J。

$P(B)$ $\Sigma P(j \wedge B \wedge e \wedge a \wedge m)$

e代表E，$\neg$E，a代表A，$\neg$A和m代表M，$\neg$M

新南威尔士大学
©W.Wobcke等人，2019-2022年

使用贝叶斯网络进行计算

- $P(J \wedge B \wedge E \wedge A \wedge M) = P(J|A)$。$P(B)$。$P(E)$。$P(A|B \wedge E)$。$P(M|A) =$
  $0.90 \times 0.001 \times 0.002 \times 0.95 \times 0.70 =$
  0.000001197
- $p(j \wedge b \wedge \neg e \wedge a \wedge m) = 0.0005910156$
- $p(j \wedge b \wedge e \wedge \neg a \wedge m) = 5 \times 10^{-11}$
- $p(j \wedge b \wedge e \neg a \wedge m) = 2.99 \times 10^{-8}$
- $p(j \wedge b \wedge e \wedge a \wedge \neg m) = 0.000000513$
- $p(j \wedge b \wedge e \wedge a \wedge m) = 0.000253292$
- $p(j \wedge b \wedge e \wedge \neg a \wedge \neg m) = 4.95 \times 10^{-9}$
- $p(j \wedge b \wedge \neg e \wedge \neg a \wedge \neg m) = 2.96406 \times 10^{-6}$

使用贝叶斯网络进行计算

- $p(\neg j \wedge b \wedge e \wedge a \wedge m) = 0.000000133$
- $p(\neg j \wedge b \wedge \neg e \wedge a \wedge m) = 6.56684 \times 10^{-5}$
- $p(\neg j \wedge b \wedge e \wedge \neg a \wedge m) = 9.5 \times 10^{-10}$
- $p(\neg j \wedge b \wedge \neg e \wedge \neg a \wedge m) = 5.6886 \times 10^{-7}$
- $p(\neg j \wedge b \wedge e \wedge a \wedge \neg m) = 0.000000057$
- $p(\neg j \wedge b \wedge \neg e \wedge a \wedge \neg m) = 2.81436 \times 10^{-5}$
- $p(\neg j \wedge b \wedge e \wedge \neg a \wedge \neg m) = 9.405 \times 10^{-8}$
- $p(\neg j \wedge b \wedge \neg e \wedge \neg a \wedge \neg m) = 5.63171 \times 10^{-5}$

使用贝叶斯网络进行计算

- 因此，$P(J|B) = \frac{P(J \wedge B)}{P(B)} = \frac{\Sigma P(J \wedge B \wedge e \wedge a \wedge m)}{\Sigma P(j \wedge B \wedge e \wedge a \wedge m)} = \frac{0.00849017}{0.001}$
- $p(j|b) = 0.849017$
- 通常可以简化计算，而不使用完整的联合概率--但并非总是如此

贝叶斯网络的推理

诊断性推理。从效果到原因

$P(Burglary|JohnCalls) = 0.016$

因果推理。从原因到结果

P（$JohnCalls|Burglary$）$=0.85$；P（$MaryCalls|Burglary$）$=0.67$

因果推理。解释了

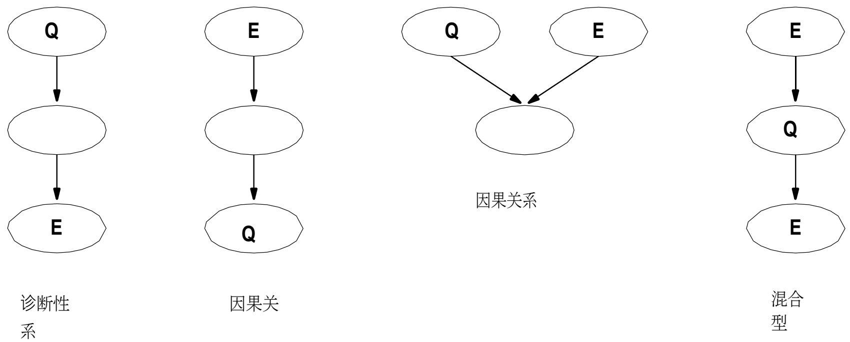
$P(Burglary|Alarm)$=0.3736，但加上证据，$P(Burglary|Alarm \wedge Earthquake)$=0.003；尽管入室盗窃和地震是独立的，但一个的存在使另一个的可能性大大降低

混合推理。上述模式的组合 诊断+因果。$P(Alarm|JohnCalls$

$\wedge\neg$Earthquake)

因果关系+诊断性。$P(Burglary|JohnCalls \wedge\neg$Earthquake)

贝叶斯网络的推理



- Q=询问；E=证据

例子 - 因果推理

- $P(JohnCalls|Burglary)$
- $p(j|b) = p(j|a \wedge b)$。$P(a|b) + P(j|\neg a \wedge b)$。$P(\neg A|B)$
  $= P(J|A).P(A|B) + P(J|\neg A)$。$P(\neg A|B)$
  $= P(J|A).P（A|B）+P（J|\neg A）$。$(1-P(A|B))$
- 现在$P(A|B)=P(A|B\wedge E)$。$P(E|B) + P(A|B\wedge\neg E)$。$P(\neg E|B)$
  $= P(A|B \wedge E)$。$P(E) + P(A|B\wedge\neg E)$。$P(\neg E)$
  $= 0.95 \times 0.002 + 0.94 \times 0.998 = 0.94002$
- 因此，$P（J|B）=0.90 \times 0.94002 + 0.05 \times 0.05998 = 0.849017$
- 事实3：$P(X|Z) = P(X|Y \wedge Z)$。$P(Y|Z) + P(X|\neg Y \wedge Z)$。$P(\neg Y$

例子 - 诊断性推断

- $P(Earthquake|Alarm)$
- $P（E|A）=\frac{P（A|E）。P(E)}{P(A)}$
  $\frac{P(A|B\wedge E)。P(B).P(E)+P(A|\neg B\wedge E)。P(\neg B).P(E)}{P(A)}$
  $= \frac{0.95\times0.001\times0.002+0.29\times0.999\times0.002}{P(A)} = \frac{5.8132\times10\text{-}4}{P(A)}$
- 现在$P(A)=P(A|B\wedge E)$。$P(B).P(E) + P(A|\neg B \wedge E)$。$P(\neg B).P(E)+$
  $P(A|B\wedge\neg E)$。$P(B).P(\neg E) + P(A|\neg B \wedge\neg E)$。$P(\neg B).P(\neg E)$
  而$P(A|B\wedge\neg E)$。$P(B).P(\neg E)+P(A|\neg B\wedge\neg E)$。$P(\neg B).P(\neg E)$
  $= 0.94 \times 0.001 \times 0.998 + 0.001 \times 0.999 \times 0.998 = 0.001935122$
  所以$P(A)=5.8132 \times 10^{-4} + 0.001935122 = 0.002516442$
- 因此，$P(E|A) = \frac{5.8132\times10^{-4}}{0.002516442} = 0.2310087$
- 事实4：$P(X \wedge Y) = P(X)$。如果X, Y是独立的，则$P(Y)$。

总结

|Z)，因为
$X \wedge Z \equiv (X \wedge Y \wedge Z) \vee (X \wedge \neg Y \wedge Z)$ (事实2的条件版本)

- 由于噪音或不确定性，用概率来推理是很有用的
- 由于有大量的数值，用联合概率分布进行计算很困难

- 使用贝叶斯法则和独立假设可以简化推理过程
- 贝叶斯网络允许紧凑地表示概率和有效地进行概率推理
- 可以给出优雅的递归算法，使贝叶斯网络的推理过程自动化