第二部分：概率模型

章宗长

2021年3月17日

内容安排

- 不确定性的表示
- **概率推理**
- 参数学习
- 结构学习
- 应用案例：视频监控

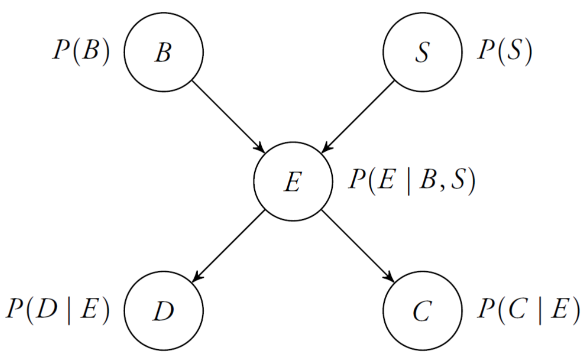
概率推理

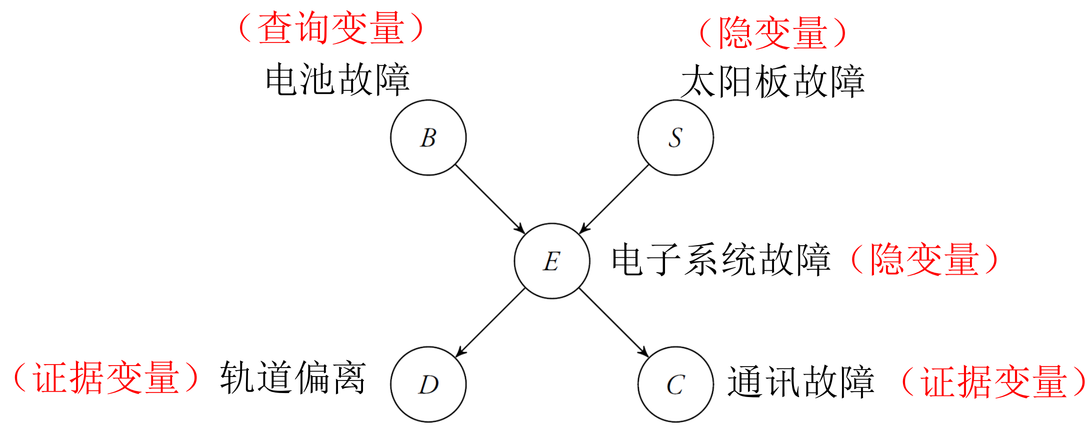
- **贝叶斯网络中的推理**
- 分类推理
- 时序模型中的推理
- 精确推理
- 精确推理的复杂度
- 近似推理

贝叶斯网络中的推理

- 推理：由一组证据变量的值确定一个或多个查询变量的分布

- 假设：想要推出分布$P(B \mid d^1, c^1)$

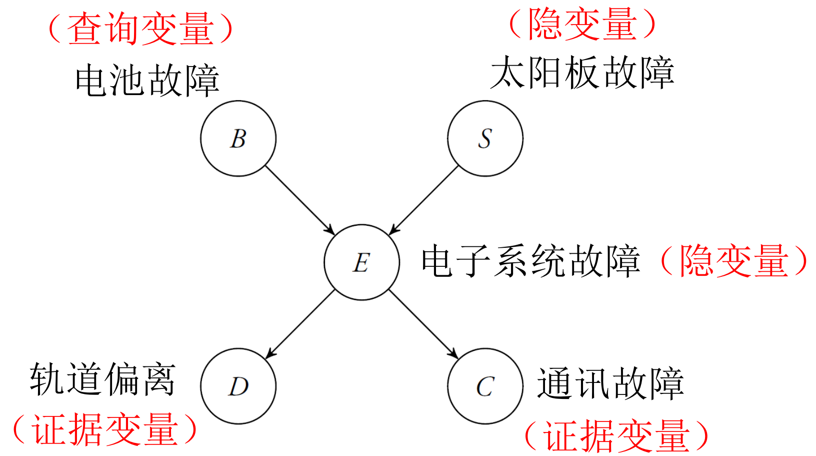




卫星监控问题的贝叶斯网络

- 通过枚举进行推理
- 利用贝叶斯网络的结构作高效的推理

枚举推理



- 例子：想要推出分布$P(b^1 \mid d^1, c^1)$
- 由条件概率的定义，有：

$$P(b^1 \mid d^1, c^1) = \frac{P(b^1, d^1, c^1)}{P(d^1, c^1)}$$

- 由全概率法则，有：

$$P(b^1, d^1, c^1) = \sum_s \sum_e P(b^1, s, e, d^1, c^1)$$

- 由链式规则，有：

$$P(b^1, d^1, c^1) = \sum_s \sum_e P(b^1)P(s)P(e \mid b^1, s)P(d^1 \mid e)P(c^1 \mid e)$$

枚举推理中的运算操作

$$P(b^1, d^1, c^1) = \sum_s \sum_e P(b^1)P(s)P(e \mid b^1, s)P(d^1 \mid e)P(c^1 \mid e)$$

$$P(\boxed{\mathcal{Y}}) = \sum_{\boxed{\mathcal{X}}} \prod_i \boxed{P(X_i \mid \mathrm{Pa}_{X_i})}$$

查询变量和证据变量的集合　隐变量　表格

- 因子相乘
- 因子边际化
- 设置证据

因子

- 表格：表示离散型的联合概率分布和条件概率分布

X	Y	Z	$P(X,Y,Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

X	Y	Z	$P(Z \mid X,Y)$
0	0	0	0.205
0	0	1	0.795
0	1	0	0.196
0	1	1	0.804
1	0	0	0.167
1	0	1	0.833
1	1	0	0.222
1	1	1	0.778

因子相乘

- 结合两个因子，以产生一个更大的因子

$\phi = P(X,Y)$

X	Y	$P(X,Y)$
0	0	0.3
0	1	0.4
1	0	0.2
1	1	0.1

$\psi = P(Y,Z)$

Y	Z	$P(Y,Z)$
0	0	0.2
0	1	0.0
1	0	0.3
1	1	0.5

X	Y	Z	$P(X,Y)P(Y,Z)$
0	0	0	0.06
0	0	1	0.00
0	1	0	0.12
0	1	1	0.20
1	0	0	0.04
1	0	1	0.00
1	1	0	0.03
1	1	1	0.05

如果有$(X \perp Z \mid Y)$，则

$$\phi \cdot \psi = P(X,Y)P(Y,Z) \propto \frac{P(X,Y)P(Y,Z)}{P(Y)} = P(Y)P(X \mid Y)P(Z \mid Y)$$
$$= P(Y)P(X,Z \mid Y) = P(X,Y,Z)$$

因子边际化

- 对因子中某一变量的所有值求和，以产生一个新的因子

X	Y	Z	$P(X,Y,Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

X	Z	$P(X,Z)$
0	0	0.17
0	1	0.68
1	0	0.03
1	1	0.12

设置证据

- 对因子中的证据变量赋值，以产生一个新的因子

X	Y	Z	$P(X,Y,Z)$
0	0	0	0.08
0	0	1	0.31
0	1	0	0.09
0	1	1	0.37
1	0	0	0.01
1	0	1	0.05
1	1	0	0.02
1	1	1	0.07

$Y \leftarrow 1$

X	Z	$P(X,Z)$
0	0	0.09
0	1	0.37
1	0	0.02
1	1	0.07

概率推理

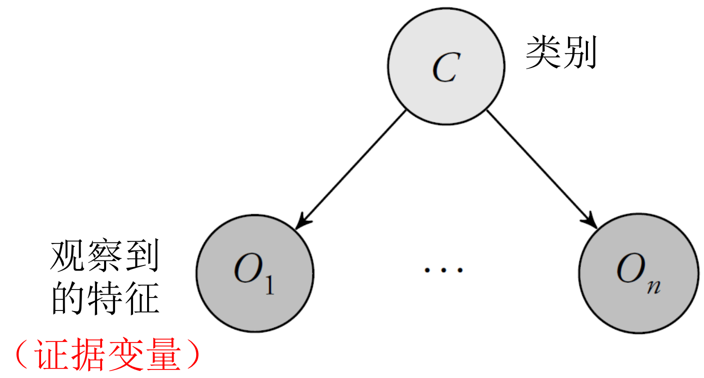
- 贝叶斯网络中的推理
- **分类推理**
- 时序模型中的推理
- 精确推理
- 精确推理的复杂度
- 近似推理

分类推理

- 用于分类任务，从给定的一组观察或特征中推理所属类别
- 例子：给定雷达目标的轨道数据（速度、头部的变化幅度等），确定雷达目标是鸟还是飞机
- 朴素贝叶斯模型：一种简单的概率模型，常用于分类任务
- （朴素）假设：给定所属类别，证据变量之间条件独立，即对于所有$i \neq j$，有

$$(o_i \perp o_j \mid C)$$

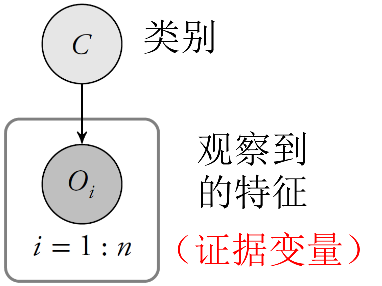
- 若假设不成立，可以在观察到的特征间添加必要的有向边



朴素贝叶斯模型

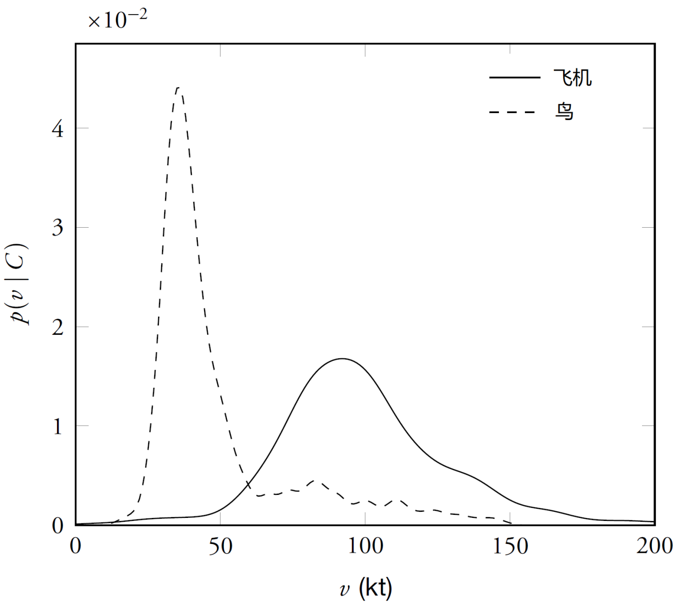
分类推理（续）

- 朴素贝叶斯模型的盘式（plate）记法



- 在朴素贝叶斯模型中，需要指出先验$P(C)$和条件于类别的分布$P(O_\mathrm{i} \mid \mathrm{C})$

- 雷达目标的分类问题
  - 先验：所跟踪目标的所有信息缺失时，目标是鸟还是飞机的概率
  - 条件于类别的分布：给定目标的类别，飞行速度的概率密度



分类推理（续）

- 需要求解条件概率$P(\mathrm{c} \mid o_1, \ldots, o_n)$，简写为$P(\mathrm{c} \mid o_{1:n})$
- 使用链式规则推出朴素贝叶斯模型的联合分布：

$$P(\mathrm{c}, o_{1:n}) = \mathrm{P}(\mathrm{c}) \prod_{i=1}^{n} P(o_i \mid c)$$

- 由条件概率的定义有：

$$P(\mathrm{c} \mid o_{1:n}) = \frac{P(\mathrm{c}, o_{1:n})}{P(o_{1:n})}$$

其中，分母 $P(o_{1:n}) = \sum_c P(\mathrm{c}, o_{1:n})$是使$\sum_c P(\mathrm{c} \mid o_{1:n}) = 1$的常数，令其为$\frac{1}{\chi}$，则有：

$$P(\mathrm{c} \mid o_{1:n}) = \chi P(\mathrm{c}, o_{1:n})$$

分类推理（续）

- 通常省去χ，从而有：

$$P(\mathrm{c} \mid o_{1:n}) \propto P(\mathrm{c}, o_{1:n})$$

$\propto$ 表示左边正比于右边

- 例子
  - P(鸟，慢，头部鲜有波动) = 0.03
  - P(飞机，慢，头部鲜有波动) = 0.01

- 给定证据，确定目标是鸟的概率

$$P(\text{鸟} \mid \text{慢, 头部鲜有波动}) = \frac{0.03}{0.03+0.01} = \mathit{0.75}$$

概率推理

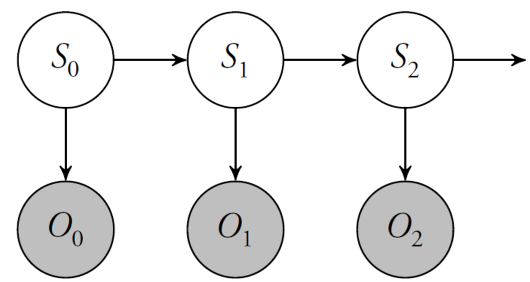
- 贝叶斯网络中的推理
- 分类推理
- **时序模型中的推理**
- 精确推理
- 精确推理的复杂度
- 近似推理

时序模型中的推理

- 很多重要的应用涉及时序模型中的推理
  - 语音识别、飞机跟踪、密码分析…

常见的推理任务：

- 滤波： $P(s_t \mid o_{0:t})$
- 预测： $P(s_{t'} \mid o_{0:t})$，其中$t' > t$
- 平滑： $P(s_{t'} \mid o_{0:t})$，其中$t' < t$
- 最可能序列： $\mathrm{argmax}_{s_{0:t}} P(s_{0:t} | o_{0:t})$



隐马尔科夫模型

滤波

- 假设离散状态和观察变量，求解$P(s_t \mid o_{0:t})$
- 由贝叶斯规则，有：

$$P(s_t \mid o_{0:t}) \propto P(o_t \mid s_t, o_{0:t-1})\, P(s_t \mid o_{0:t-1})$$

- 由条件独立性假设($o_t \perp o_{0:t-1} \mid s_t$)有：

$$P(o_t \mid s_t, o_{0:t-1}) = P(o_t \mid s_t)$$

- 再结合全概率法则，有：

$$P(s_t \mid o_{0:t}) \propto P(o_t \mid s_t) \sum_{s_{t-1}} P(s_t, s_{t-1} \mid o_{0:t-1})$$

滤波（续）

- 由条件概率的定义，有：

$$P(s_t \mid o_{0:t}) \propto P(o_t \mid s_t) \sum_{s_{t-1}} P(s_t \mid s_{t-1}, o_{0:t-1}) P(s_{t-1} \mid o_{0:t-1})$$

- 由于$(s_t \perp o_{0:t-1} \mid s_{t-1})$，有：

$$P(s_t \mid o_{0:t}) \propto P(o_t \mid s_t) \sum_{s_{t-1}} P(s_t \mid s_{t-1}) P(s_{t-1} \mid o_{0:t-1})$$

  其中$P(o_t \mid s_t)$和$P(s_t \mid s_{t-1})$可以直接由模型得到

- 假设：状态转移分布$P(s_t \mid s_{t-1})$和观察分布$P(o_t \mid s_t)$是稳态的，即不随时间变化

滤波（续）

- 递归贝叶斯估计（前向算法）：得到观察o_t后，由$P(s_{t-1} \mid o_{0:t-1})$计算$P(s_t \mid o_{0:t})$
- b_t：t时刻的后验状态估计

Algorithm 2.1 Recursive Bayesian estimation

1: **function** RecursiveBayesianEstimation
2: $\quad b_0(s) \leftarrow P(o_0 \mid s)P(s)$ for all s
3: $\quad$ Normalize b_0
4: $\quad$ **for** $t \leftarrow 1$ **to** ∞
5: $\quad\quad b_t(s) \leftarrow P(o_t \mid s)\sum_{s'} P(s \mid s')b_{t-1}(s')$ for all s
6: $\quad\quad$ Normalize b_t

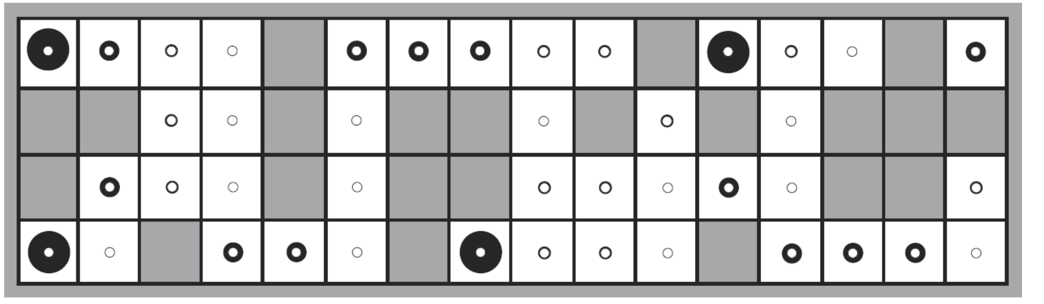
- 当观察连续时，$P(o \mid s)$是概率密度函数
- 当状态连续时，第5行的求和变成积分，b变成密度函数

实例：定位

- 得到一个观察结果NSW后机器人位置的后验分布

假设机器人的初始位置均匀分布在各方块

东南西北4个方向各有一个传感器，NSW表示感知到北边、南边和西边有障碍物



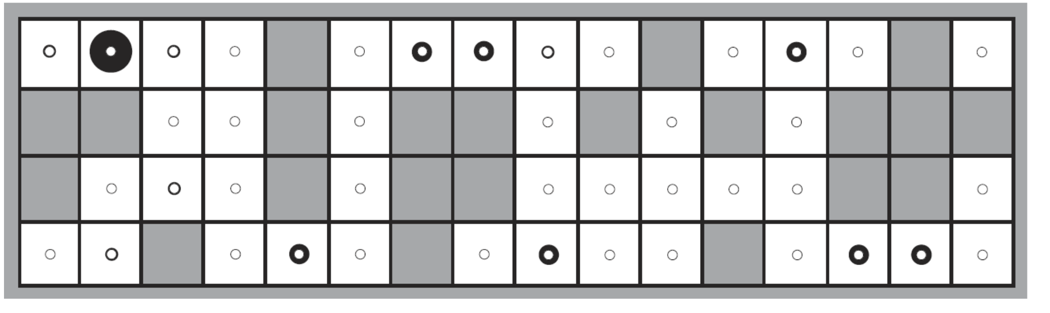
机器人只有一个非确定性行动Move，等可能地移动到各相邻方块

- 得到第二个观察结果NS后机器人位置的后验分布

NS表示感知到北边、南边有障碍物

每个圆圈的大小表示机器人在那个位置的概率

传感器错误率是0.2



预测

- 求解$P(s_{t+k+1} | o_{0:t})$，其中$k \geq 0$

- 单步预测（滤波）

$$P(s_t | o_{0:t}) \propto P(o_t | s_t) \sum_{s_{t-1}} P(s_t | s_{t-1}, o_{0:t-1}) P(s_{t-1} | o_{0:t-1})$$

- 没有增加新观察的条件下的滤波

$$P(s_{t+k+1} | o_{0:t}) = \sum_{s_{t+k}} P(s_{t+k+1} | s_{t+k}) P(s_{t+k} | o_{0:t})$$

只涉及转移分布，不涉及观察分布

平滑

- 求解$P(s_k | o_{0:t})$，其中$0 \le k < t$

$$P(s_k \mid o_{0:t}) = P(s_k \mid o_{0:k}, o_{k+1:t})$$

$$\propto P(s_k \mid o_{0:k}) P(o_{k+1:t} \mid s_k, o_{0:k})$$

$$= \underbrace{P(s_k \mid o_{0:k})}_{\text{前向算法}} \underbrace{P(o_{k+1:t} \mid s_k)}_{\text{后向算法}}$$

- 前向-后向算法

平滑（续）

- 后向算法

$$P(o_{k+1:t} \mid s_k) = \sum_{s_{k+1}} P(o_{k+1:t} \mid s_k, s_{k+1}) P(s_{k+1} \mid s_k)$$

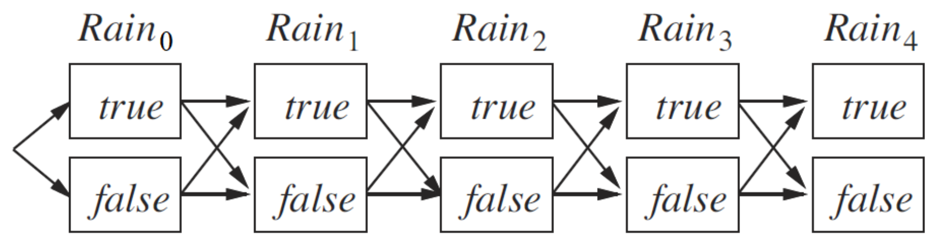
$$= \sum_{s_{k+1}} P(o_{k+1:t} \mid s_{k+1}) P(s_{k+1} \mid s_k)$$

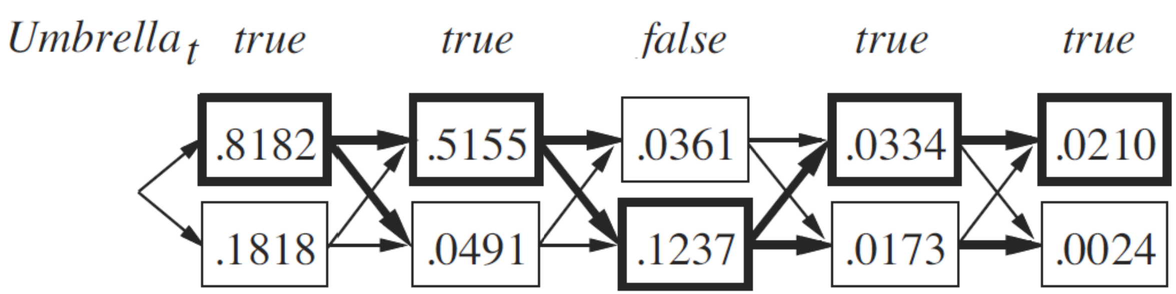
$$= \sum_{s_{k+1}} P(o_{k+1}, o_{k+2:t} \mid s_{k+1}) P(s_{k+1} \mid s_k)$$

$$= \sum_{s_{k+1}} P(o_{k+1} \mid s_{k+1}) P(o_{k+2:t} \mid s_{k+1}) P(s_{k+1} \mid s_k)$$

寻找最可能序列

- 求解 $\mathrm{argmax}_{s_{0:t}} P(s_{0:t} \mid o_{0:t})$
- 维特比（Viterbi）算法
- 雨伞世界
  - t时刻的状态：$Rain_t$
  - t时刻的观察：$Umbrella_t$





$P(S_0 \mid o_0)$ $P(S_1 \mid o_{0:1}, s_0^1)$ $P(S_2 \mid o_{0:2}, s_0^1, s_1^1)$ $P(S_3 \mid o_{0:3}, s_0^1, s_1^1, s_2^0)$ $P(S_4 \mid o_{0:4}, s_0^1, s_1^1, s_2^0, s_3^1)$

最可能序列 s_0^1 s_1^1 s_2^0 s_3^1 s_4^1

概率推理

- 贝叶斯网络中的推理
- 分类推理
- 时序模型中的推理
- **精确推理**
- 精确推理的复杂度
- 近似推理

精确推理

- 计算查询变量的边际分布或条件分布的精确值
- 枚举推理

$$P(b^1, d^1, c^1) = \sum_s \sum_e P(b^1)P(s)P(e \mid b^1, s)P(d^1 \mid e)P(c^1 \mid e)$$

表格

利用贝叶斯网络的结构

$$P(\boxed{\mathcal{Y}}) = \sum_{\boxed{\mathcal{X}}} \prod_i \boxed{P(X_i \mid \mathrm{Pa}_{X_i})}$$

隐变量

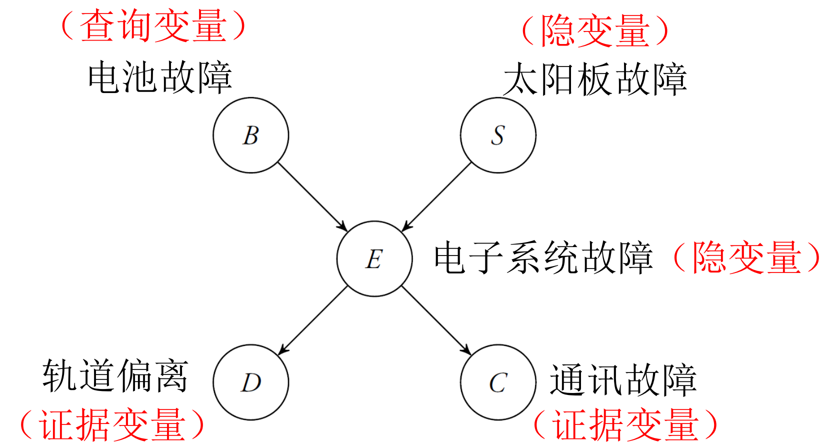
查询变量和证据变量的集合

变量消去法

信念传播法

- 对隐变量求和
  - 随着隐变量数目的增加，求和项的数目将呈指数级增长
- 有n个二值变量的网络，枚举推理算法的复杂度为$O(n \cdot 2^n)$

变量消去法



- 例：计算分布$P(B \mid d^1, c^1)$
- 在网络中与结点B、D、C关联的条件概率分布可用如下表格来表示：

$$T_1(B), T_2(S), T_3(E,B,S), T_4(D,E), T_5(C,E)$$

- 因为D和C是观察到的变量，仅需保存$T_4(D,E)$和$T_5(C,E)$中$D=1$和$C=1$的对应行
- 使用设置证据的操作对表格进行简化，简化后的表格记为$T_6(E)$和$T_7(E)$
- 然后，按一定的顺序来消除隐变量

变量消去法（续）

- 如先消除E，找到所有含有E的表格：

$$T_3(E,B,S), T_6(E),\ T_7(E)$$

- 对这些表执行因子相乘，得到一张新的表格：

$$T_8(E,B,S) = T_3(E,B,S)T_6(E)T_7(E)$$

- 对表格T_8执行因子边际化，得到一张新的表格：

$$T_9(B,S) = \sum_{\mathrm{e}} T_8(e,B,S)$$

- 丢掉T_3, T_6, T_7，因为需要从它们获得的信息都包含在T_9中

变量消去法（续）

- 再消除S，收集余下的含有S的表格：$T_2(S)$，$T_9(B,S)$

- 对这两张表执行因子相乘，得到：

$$T_{10}(B) = T_2(S)T_9(B,S)$$

- 丢掉T_2, T_9，剩余$T_1(B)$和$T_{10}(B)$，执行因子相乘，得到：

$$T_{11}(B) = T_1(B)T_{10}(B)$$

- 归一化表T_{11}，即可得到$P(B \mid d^1, c^1)$

变量消去法（续）

- 枚举法

$$P(B,d^1,c^1)=\sum_s\sum_e T_1(B)\,T_2(s)T_3(e,B,s)T_4(D=1,e)T_5(C=1,e)$$

$$=\sum_s\sum_e T_1(B)\,T_2(s)T_3(e,B,s)T_6(e)T_7(e)$$

- 变量消去法

先消去E，
再消去S

$$P(B,d^1,c^1)=T_1(B)\sum_s\left(T_2(s)\sum_e\big(T_3(e,B,s)T_6(e)T_7(e)\big)\right)$$

先消去S，
再消去E

$$P(B,d^1,c^1)=T_1(B)\sum_e\left(T_6(e)T_7(e)\sum_s\big(T_2(s)T_3(e,B,s)\big)\right)$$

变量消去法（续）

- 贝叶斯网络中的变量消去
  - B：贝叶斯网络
  - $\mathcal{Q}$：查询变量集合
  - $\boldsymbol{o}$：观察到的值

Algorithm 2.2 Variable elimination in Bayesian networks

function VariableElimination($B, \mathcal{Q}, \mathbf{o}$)
  $\mathcal{T} \leftarrow$ set of conditional probability tables associated with nodes in B
  Remove rows that are inconsistent with $\mathbf{o}$ from all the tables in $\mathcal{T}$
  for $i \leftarrow 1$ **to** n
    $\mathcal{T}' \leftarrow$ all the tables in $\mathcal{T}$ that involve X_i
    $T \leftarrow$ the product of the tables in $\mathcal{T}'$ with X_i summed out
    Remove $\mathcal{T}'$ from $\mathcal{T}$ and add T
  $T \leftarrow$ product of the tables remaining in $\mathcal{T}$
  $P(\mathcal{Q} \mid \mathbf{o}) \leftarrow$ normalize T
  return $P(\mathcal{Q} \mid \mathbf{o})$

变量消去法（续）

- 变量消去的顺序会影响变量消去法的计算时间
- 选择最优的消元顺序是**NP-难**问题，即最坏情况下不能在多项式时间内求解
- 在最坏情况下，即使有最优的消元顺序，变量消去法所需的计算时间仍然可以是网络大小的指数级复杂度

变量消去法（续）

- 缺点：若需计算多个边际分布，重复使用变量消去法将会造成大量的冗余计算
- 假定在计算了$P(B,d^1,c^1)$之外，还希望计算$P(S,d^1,c^1)$

先消去E，再消去S

$$P(B,d^1,c^1) = T_1(B)\sum_s\left(T_2(s)\boxed{\sum_e\left(T_3(e,B,s)T_6(e)T_7(e)\right)}\right)$$

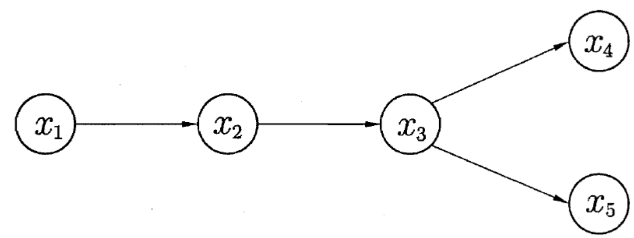
这部分的信息是重复计算的

先消去E，再消去B

$$P(S,d^1,c^1) = T_2(S)\sum_b\left(T_1(b)\boxed{\sum_e\left(T_3(e,b,S)T_6(e)T_7(e)\right)}\right)$$

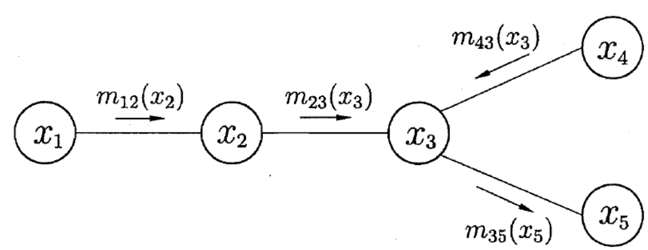
信念传播法

- 将变量消去法中的求和操作看作一个消息传递过程
- 例子



(a) 贝叶斯网络结构

- 若采用$\{x_1, x_2, x_4, x_3\}$的顺序消去变量，计算$P(x_5)$



(b) 消息传递过程

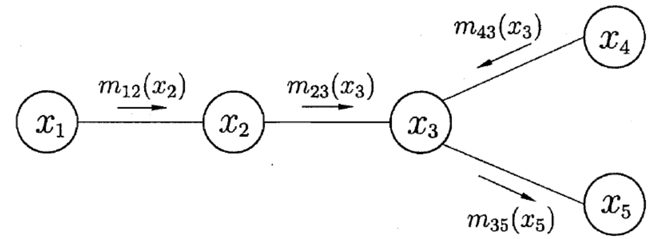
ψ：因子
$n(i)$：结点x_i的邻接结点

- 从x_i向x_j传递了一个消息

$$m_{ij}(x_j) = \sum_{x_i} \psi(x_i, x_j) \prod_{k \in n(i) \setminus j} m_{ki}(x_i)$$

信念传播法（续）

- 若采用$\{x_1,x_2,x_4,x_3\}$的顺序消去变量，计算$P(x_5)$



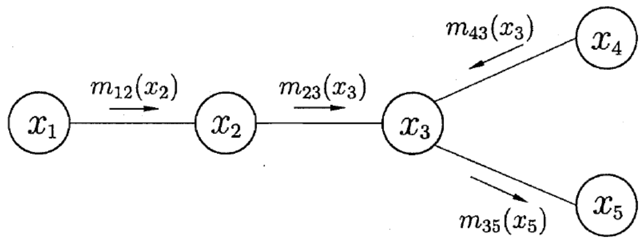
(b) 消息传递过程

$$
\begin{aligned}
P(x_5) &= \sum_{x_3} P(x_5 \mid x_3) \sum_{x_4} P(x_4 \mid x_3) \sum_{x_2} P(x_3 \mid x_2) \boxed{\sum_{x_1} P(x_1) P(x_2 \mid x_1)} \\
&= \sum_{x_3} P(x_5 \mid x_3) \sum_{x_4} P(x_4 \mid x_3) \boxed{\sum_{x_2} P(x_3 \mid x_2) \boxed{m_{12}(x_2)}} \\
&= \sum_{x_3} P(x_5 \mid x_3) \sum_{x_4} P(x_4 \mid x_3) \boxed{m_{23}(x_3)} \\
&= \sum_{x_3} P(x_5 \mid x_3) m_{23}(x_3) \boxed{\sum_{x_4} P(x_4 \mid x_3)} \\
&= \boxed{\sum_{x_3} P(x_5 \mid x_3) m_{23}(x_3) \boxed{m_{43}(x_3)}} \\
&= \boxed{m_{35}(x_5)}
\end{aligned}
$$

$$\boxed{m_{ij}(x_j) = \sum_{x_i} \psi(x_i, x_j) \prod_{k \in n(i) \setminus j} m_{ki}(x_i)}$$

信念传播法（续）

- 一个结点仅在接收到来自其他所有结点的消息后才能向另一个结点发送消息



(b) 消息传递过程

结点x_3向x_5发送消息，必须事先收到来自结点x_2和x_4的消息

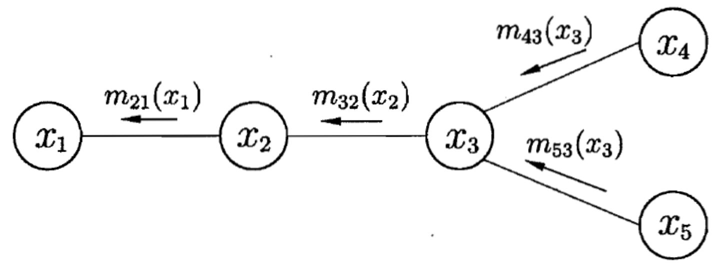
- 结点的边际分布正比于它所接收的消息的乘积：

$$P(x_i) \propto \prod_{k \in n(i)} m_{ki}(x_i)$$

- 如果图的每条边上都有方向不同的两条消息，则使用上式即可获得所有变量的边际概率

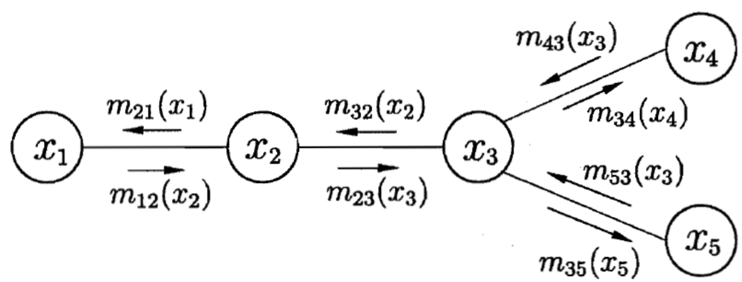
信念传播法（续）

- 指定一个根结点，从所有叶子结点开始向根结点传递消息，直到根结点收到所有邻接结点的消息



(a) 消息传向根结点

- 从根结点开始向叶子结点传递消息，直到所有叶子结点均收到消息



(b) 消息从根结点传出

概率推理

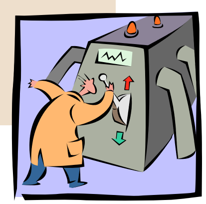
- 贝叶斯网络中的推理
- 分类推理
- 时序模型中的推理
- 精确推理
- **精确推理的复杂度**
- 近似推理

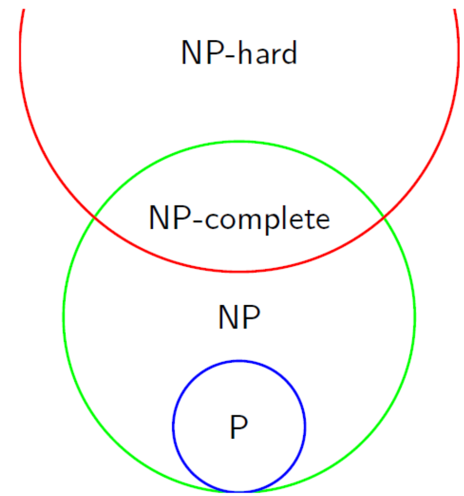
精确推理的复杂度

复杂度类

- P：可以在多项式时间内求解的问题
- NP：解可以在多项式时间内得到验证的问题
- NP-难：难度不小于最难的NP问题的问题
- NP-完全：同属于NP和NP-难的问题

$$P \neq NP?$$





- 右图是在这个假设的前提下，各复杂度类间的关系

精确推理的复杂度（续）

- 证明一个问题Q是**NP-难**的：
  - （通常的做法）把一个已知的NP-完全问题转化为Q的一个实例
- 可满足性（Satisfiability，SAT）问题
  - 确定一个布尔公式是否可满足
  - 第一个被证明为NP-完全的问题
- 通过3SAT问题来证明贝叶斯网络中的精确推理是NP-难问题
- 布尔公式由与（$\wedge$）、或（$\vee$）、非（$\neg$）及n个布尔变量 $x_1, \ldots, x_n$ 构成
- 文字：变量x_i或它的否定式$\neg x_\mathrm{i}$

精确推理的复杂度（续）

- 3SAT子句：3个文字构成的析取式，如：$x_3 \vee \neg x_5 \vee x_6$
- 3SAT公式： 3SAT子句的合取式

例如： $F(x_1, x_2, x_3, x_4) = (x_1 \vee x_2 \vee x_3) \wedge$
$(\neg x_1 \vee \neg x_2 \vee x_3) \wedge$
$(x_2 \vee \neg x_3 \vee x_4)$

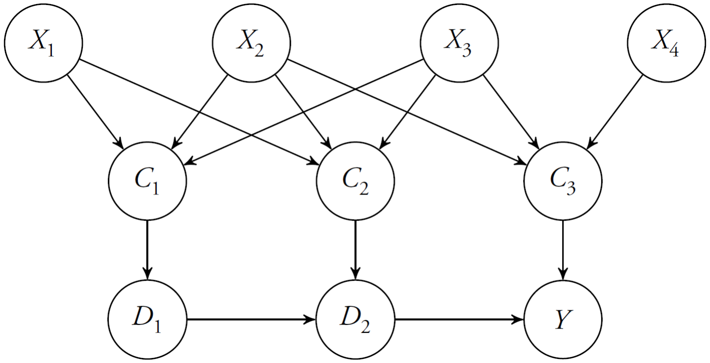
- 3SAT问题：确定是否存在一种对式中变量的可能赋值，使得3SAT公式为真
- 例如： $F(\text{真}, \text{假}, \text{假}, \text{真}) = \text{真}$，因此该式是可满足的

精确推理的复杂度（续）

- 验证一种变量赋值是否可满足只需要线性时间
- 在有些3SAT问题中，找到一组可满足的变量赋值很困难
- 最坏情况：测试2^n种可能的变量赋值
- 容易将任意3SAT问题表示成贝叶斯网络

精确推理的复杂度（续）

$$F(x_1,x_2,x_3,x_4) = (x_1 \vee x_2 \vee x_3) \wedge (\neg x_1 \vee \neg x_2 \vee x_3) \wedge (x_2 \vee \neg x_3 \vee x_4)$$



- 变量表示为$X_{1:4}$，子句表示为$C_{1:3}$
- 令
  - 变量服从均匀分布
  - 每个子句结点为1的概率为父结点构成的子句为真的概率
  - 剩余结点为1的概率为所有父结点均为1的概率
- 原始3SAT问题是可满足的 $\Leftrightarrow P(y^1) > 0$
- 因此贝叶斯网络中精确推理的难度不小于3SAT

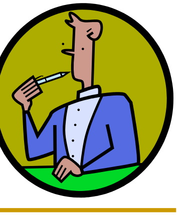
课后练习2.3

- 什么是分类任务？朴素贝叶斯模型使用了什么假设？使用盘式记法来画一个朴素的贝叶斯模型。

课后练习2.4

- 有一位教授想知道学生是否睡眠充足。每天，教授观察学生在课堂上是否睡觉，并观察他们是否红眼。教授有如下的领域理论：
  - 没有观察数据时，学生睡眠充足的先验概率为0.7。
  - 给定学生前一天睡眠充足为条件，学生在晚上睡眠充足的概率是0.8；如果前一天睡眠不充足，则是0.3。
  - 如果学生睡眠充足，则红眼的概率是0.2，否则是0.7。
  - 如果学生睡眠充足，则在课堂上睡觉的概率是0.1，否则是0.3。

  将这些信息形式化为一个动态贝叶斯网络，使教授可以使用这个网络从观察序列中进行滤波和预测。然后再将其形式化为一个只有一个观察变量的隐马尔科夫模型。给出这个模型的完整的概率表。



课后练习2.5

- 对于上一道练习描述的动态贝叶斯网络以及证据变量值
  - o_0=没有红眼，没有在课堂上睡觉
  - o_1=有红眼，没有在课堂上睡觉
  - o_2=有红眼，在课堂上睡觉

  执行下面的计算：

  a. 状态估计：针对每个$t = 0, 1, 2$，计算$P(EnoughSleep_t \mid o_{0:t})$。

  b. 平滑：针对每个$t = 0, 1, 2$，计算$P(EnoughSleep_t \mid o_{0:2})$。

  c. 针对$t = 0$和$t = 1$，比较滤波概率和平滑概率。