人工智能 第三次作业

第一题

（1）

要证：$(a \vee b) \wedge (\neg a \vee c) \wedge (\neg b \vee d) \wedge (\neg c \vee g) \wedge (\neg d \vee g) \Rightarrow g$

即证：$(a \vee b) \wedge (\neg a \vee c) \wedge (\neg b \vee d) \wedge (\neg c \vee g) \wedge (\neg d \vee g) \wedge \neg g$ 不可满足

归结 $(a \vee b), (\neg b \vee d)$：$(a \vee d) \wedge (\neg a \vee c) \wedge (\neg c \vee g) \wedge (\neg d \vee g) \wedge \neg g$3

归结 $(a \vee d), (\neg a \vee c)$：$(c \vee d) \wedge (\neg c \vee g) \wedge (\neg d \vee g) \wedge \neg g$

归结 $(g \vee d), (\neg d \vee g)$：$(g \vee d) \wedge (\neg d \vee g) \wedge \neg g$

归结 $(c \vee d), (\neg c \vee g)$：$g \wedge \neg g$

归结 $g, \neg g$：NIL

得证

（2）

语义不同：可能的语义数为 $4^{C_n^2}$

语义可相同：符号不同的子句个数为 $4^{C_n^2}$，符号相同的子句个数为$2^n + 1$，一共$4^{C_n^2} + 2^n + 1$个

第二题

（1）

(a) axiom 4 and $\{x_1/7\}$

(b) axiom 8 and $\{x2/(7+0), y2/(0+7), z2/(3+9)\}$

(c) axiom 3 and $\{x3/7，y3/0\}$

(d) axiom 7 and $\{x4/0, y4/3, w4/7, z4/9\}$

（2）

事实上，前向链接算法要求在每次迭代中便利所有推理规则，将所有可以推出的结论加入知识库内。本题受限于篇幅，仅在每次迭代中列出对最终结论有贡献的中间结论。

- Iteration 1

  Goal 1:$0 \leq 3$ is satisfied with axiom.1, and $0 \leq 3$ is added.
- Iteration 2

  Goal 2:$7 \leq 9$ is satisfied with axiom.2, and $7 \leq 9$ is added.
- Iteration 3

  Goal 3:$0 + 7 \leq 3 + 9$ is satisfied with axiom 3, and $0 + 7 \leq 3 + 9$ is added.
- Iteration 4

  Goal 4:$7 + 0 \leq 0 + 7$ is satisfied with axiom ,6 and $7 + 0 \leq 0 + 7$ is added.
- Iteration 5

  Goal 5:$7 + 0 \leq 3 + 9$ is satisfied with axiom 8, and $7 + 0 \leq 3 + 9$ is added.

- Iteration 6

  Goal 6:$7 \leq 7+0$ is satisfied with axiom 4, and $7 \leq 7+0$ is added.

- Iteration 7

  Goal 7:$7 \leq 3+9$ is satisfied with axiom 8 and the theorem is proved.

第三题

（1）

$P(B,A,E,J.M)=P(B)P(E)P(A|B,E)P(J|A)P(M|A)$

（2）

给定变量{A, E}，则B⊥J、B⊥M

（3）

给定变量A：B⊥J、B⊥M、E⊥J、E⊥M、J⊥M

不给定变量A：B⊥E

（4）

$$P(B=b_0,J=j_0,M=m_0)=\sum_e\sum_a P(b_0)P(e)P(a|b_0,e_0)P(j_0|a)P(m_0|a)$$

消元：

$$\begin{aligned}
& P(B)P(E)P(A|B,E)P(J|A)P(M|A) \\
= \quad & P(B)P(E)P(A|B,E)P(J|A)m_M(A) \quad [m_M(A)=P(M=m_0|A)] \\
= \quad & P(B)P(E)P(A|B,E)m_J(A) \quad [m_J(A)=P(J=j_0|A)m_M(A)] \\
= \quad & P(B)P(E)m_A(B,E) \quad [m_A(B,E)=\sum_A P(A|B,E)m_J(A)] \\
= \quad & P(B)m_E(B) \quad [m_E(B)=\sum_E P(E)m_A(B,E)] \\
= \quad & m_B() \quad [m_B()=P(B=b_0)m_E(b_0)]
\end{aligned}$$

$P(B=b_0,J=j_0,M=m_0)=m_B()$（m的计算方法如上式中括号中标出）

第四题

（1）Rejection Sampling

1. r=0.31, r<P(+a), 对于节点A取 +a
2. r=0.58, r<P(+b | +a), 对于节点B取 +b。拒绝
3. r=0.04, r<P(+a), 对于节点A取 +a
4. r=0.94, r≥P(+b | +a), 对于节点B取 -b
5. r=0.67, r<P(+c | +a), 对于节点C取 +c
6. r=0.49, r≥P(+d | -b,+c), 对于节点D取 -d。得到样本(+a, -b, +c, -d)
7. r=0.37, r<P(+a), 对于节点A取 +a
8. r=0.42, r<P(+b | +a), 对于节点B取 +b。拒绝

采用8个随机数，进行了1次采样，得到一个样本，为(+a, -b, +c, -d)

$$P(-d|-b)=\frac{1}{1}=1$$

（2）Likelihood Weighting Sampling

1. r=0.31, r<P(+a), 对于节点A取 +a, w=1.0
2. b 是 evidence，$w \leftarrow w \times P(-b|+a)$, w=0.2
3. r=0.58, r<P(+c|+a), 对于节点C取 +c
4. r=0.04, r<P(+d|-b,+c), 对于节点D取 +d。得到样本(+a, -b, +c,+d), w=0.2
5. r=0.94, r≥P(+a), 对于节点A取 -a
6. b 是 evidence，$w \leftarrow w \times P(-b|-a)$, w=0.5
7. r=0.67, r≥P(+c|-a), 对于节点C取 -c
8. r=0.49, r<P(+d|-b, -c), 对于节点D取 +d。得到样本(-a, -b, -c,+d), w=0.5

采用6个随机数，进行了2次采样，得到两个样本，为(+a, -b, +c,+d), w=0.2和(-a, -b, -c,+d), w=0.5

$$P(-d|-b) = 0$$

（3）Gibbs Sampling (+a, -b, +c, +d)

1. r=0.31

$$\begin{aligned} P(+a|mb(A)) &= \frac{P(+a)P(-b|+a)P(+c|+a)}{P(+a)P(-b|+a)P(+c|+a)+P(-a)P(-b|-a)P(+c|-a)} \\ &= \frac{0.112}{0.112+0.1} = 0.528 \end{aligned}$$

   r<P(+a|mb(A)), 对于节点A取 +a

   当前状态 (+a, -b, +c, +d)

2. r=0.58

$$\begin{aligned} P(+c|mb(C)) &= \frac{P(+c|-a)P(+d|-b,+c)}{P(+c|-a)P(+d|-b,+c)+P(-c|-a)P(+d|-b,-c)} \\ &= \frac{0.02}{0.02+0.81} = 0.024 \end{aligned}$$

   r≥P(+c|mb(C)), 对于节点C取 -c

   当前状态 (+a, -b, -c, +d)

3. r=0.04

$$\begin{aligned} P(+d|mb(D)) &= \frac{P(+d|-b,-c)}{P(+d|-b,-c)+P(-d|-b,-c)} \\ &= \frac{0.9}{0.9+0.1} = 0.9 \end{aligned}$$

   r<P(+d|mb(D)), 对于节点D取 +d

   当前状态 (-a, -b, -c, +d)

4. r=0.94

$$\begin{aligned} P(+a|mb(A)) &= \frac{P(+a)P(-b|+a)P(-c|+a)}{P(+a)P(-b|+a)P(-c|+a)+P(-a)P(-b|-a)P(-c|-a)} \\ &= \frac{0.048}{0.048+0.09} = 0.348 \end{aligned}$$

   r≥P(+a|mb(A)), 对于节点A取 -a

   当前状态 (-a, -b, -c, +d)

5. r=0.67

$$P(+c|mb(C)) = \frac{P(+c|-a)P(+d|-b,+c)}{P(+c|-a)P(+d|-b,+c)+P(-c|-a)P(+d|-b,-c)}$$
$$= \frac{0.02}{0.02+0.81} = 0.024$$

r≥P(+c|mb(C)), 对于节点C取 -c

当前状态 (-a, -b, -c, +d)

6. r=0.49

$$P(+d|mb(D)) = \frac{P(+d|-b,-c)}{P(+d|-b,-c)+P(-d|-b,-c)}$$
$$= \frac{0.9}{0.9+0.1} = 0.9$$

r<P(+d|-b, -c), 对于节点D取 +d

当前状态 (-a, -b, -c, +d)

采用6个随机数，进行了6次采样，得到六个样本：1个 (+a, -b, +c, +d)， 1个(+a, -b, -c, +d)， 4个(-a, -b, -c, +d)

$$P(-d|-b) = \frac{0}{6} = 0$$

第五题

（1）

$$\begin{aligned} & P(x_1,\cdots,x_t,y_{t-1}=s_v,y_t=s_j) \\ = \quad & P(x_t|s_j)\times P(x_1,\cdots,x_{t-1},y_{t-1}=s_v)\times P(s_j|s_v) \\ = \quad & b_j(x_t)\alpha_{t-1}^i a_{v,j} \end{aligned}$$

（2）

(a)

$$p_{y=C}^3 = p(A|A)p(C|A) = A_1A_2$$

(b)

$$p_{y=A}^t = 1-b_0-c_0$$

$$p_{x=2}^t = \sum_y p_y^t p(x=2|y) = 0.3p_A^t = 0.3(1-b_0-c_0)$$

第三次作业

提交 DDL：2021 年 12 月 6 日 0 时

作业完成形式有三种：

（1）你可以手写自己的解答并拍照，再将照片整理成一份 word/pdf 文件并提交。

（2）你可以使用 word 文档进行编辑，最后提交 word/pdf 文件。

（3）你可以使用 latex 进行编辑，最后提交 pdf 文件。

如果你没有在 DDL 之前提交作业，请及时在微信群里联系助教进行补交。如果对作业有任何问题，你可以在从微信里询问助教谢瑜璋，或者发邮件到 constantjxyz@sjtu.edu.cn。

1 本次作业可能用到的知识点

本次作业可能会用到以下知识点：

（1）命题逻辑（propositional logic）、谓词逻辑（first-order logic）的定义、语义

（2）逻辑证明的假言推理规则（Modus Ponens）、归结证明（resolution）。

（3）逻辑推理的前向链接算法（forward chaining）、后向链接算法（backward chaining）。

（3）贝叶斯网络节点联合概率表示，马尔科夫毯（Markov Blanket）的定义，贝叶斯网络中的条件独立（conditional independence）关系

（4）贝叶斯网络中的精确推理、消元法

（5）贝叶斯网络中的近似推理采样方法

（6）HMM 模型的建模方式、概率推导

2 第一题

2-CNF 表达式类似于 3-CNF，其中每个子句 (clause) 里含有两个文字 (literals)，例如：

$$(a \vee b) \wedge (\neg a \vee c) \wedge (\neg b \vee d) \wedge (\neg c \vee g) \wedge (\neg d \vee g)$$

（1）运用反证法，证明上面的表达可以推导出 g (the above sentence entails g)，可以参考课件的归结证明 resolution 部分。

解答：要使用反证法证明上述式子蕴含 g，即证明 $(a \vee b) \wedge (\neg a \vee c) \wedge (\neg b \vee d) \wedge (\neg c \vee g) \wedge (\neg d \vee g) \wedge (\neg g)$ 不可满足。

运用归结证明的算法，我们此时有 $(a \vee b), (\neg a \vee c), (\neg b \vee d), (\neg c \vee g), (\neg d \vee g)$ 与 $(\neg g)$ 六个子句。

将 $(\neg g)$ 分别与 $(\neg d \vee g)$、$(\neg c \vee g)$ 合并，可以得到子句 $(\neg d)$ 与 $(\neg c)$。

将子句 $(\neg d)$ 与 $(\neg b \vee d)$ 合并，可以得到子句 $(\neg b)$

将子句 $(\neg c)$ 与子句 $(\neg a \vee c)$ 合并，可以得到子句 $(\neg a)$

将子句 $(\neg b)$、$(\neg a)$ 与子句 $(a \vee b)$ 合并，可以得到空集，根据 resolution 算法，得到空集即证明原始的 $(a \vee b) \wedge (\neg a \vee c) \wedge (\neg b \vee d) \wedge (\neg c \vee g) \wedge (\neg d \vee g) \wedge (\neg g)$ 不可满足。因此表达式 $(a \vee b) \wedge (\neg a \vee c) \wedge (\neg b \vee d) \wedge (\neg c \vee g) \wedge (\neg d \vee g)$ 可以推导出 g。

归并过程中的顺序可以改变，例如也可以将第一个子句与第二个子句归并，然后再依次将归并结果与后面一个子句归并。只要最后能得到空集即可证明结果。

（2）对于 2-CNF 问题，假设现在我们有 n 个不同的符号。如果我们规定每个子句要求用**不同**的符号组成，那么我们可以用这 n 个符号组成多少种**语义不同（semantic distinct）**的子句

(clause)？如果我们规定每个子句可以用**相同**的符号组成，那么我们可以用这 n 个符号组成多少种**语义不同**的子句？

解答：如果我们用不同的符号组成语义不同的子句，子句内的顺序与子句的语义无关，例如 $(a\vee b)$ 与 $(b\vee a)$ 是语义相同的子句。那么在挑选符号的时候，一共有 $C_n^2 = n\cdot(n-1)\cdot\frac{1}{2}$ 种组合可能。而对于每种挑选时的组合可能，注意我们可以在各个位置加入 $\neg$ 符号，例如 $(a\vee b)$, $(\neg a\vee b)$, $(a\vee\neg b)$, $(\neg a\vee\neg b)$ 是四种不同的子句，因此我们最终可以得到 $C_n^2\cdot 4 = 2n^2-2n$ 种语义不同的子句。

如果我们使用相同的符号，我们有 n 种挑选符号的可能性。如果我们挑选出 a，我们可以组成 $(a\vee a)$, $(\neg a\vee a)$, $(a\vee\neg a)$, $(\neg a\vee\neg a)$ 四种子句。注意 $(\neg a\vee a)$ 与 $(a\vee\neg a)$ 语义相同，其语义都是 ***True***。因此当我们挑选出 a 的时候，我们可以组成 3 种语义不同的子句，其语义分别为 a, $\neg a$, ***True***。最终，将所有情况相加，如果我们规定每个子句可以用相同的符号组成，我们可以得到 $2n^2-2n+2n+1 = 2n^2+1$ 种语义不同的子句。

本题的关键是理解语义的概念，语义 (semantics) 决定了一个逻辑语句的含义，可以用于判断其什么时候为真、什么时候为假。因此 $(a\vee b)$ 与 $(b\vee a)$ 两种子句在任何情况下的真假情况都相同，可以视为语义相同的语句。与之类似的是，$(a\vee\neg a)$ 与 $(b\vee\neg b)$ 都恒等于 ***True***，可以视为语义相同的语句。

3 第二题

我们建立了一个新的数学空间，在这个空间里有以下公理：

- 1. $0\leq 3$
- 2. $7\leq 9$
- 3. $\forall x \quad x\leq x$
- 4. $\forall x \quad x\leq x+0$
- 5. $\forall x \quad x+0\leq x$
- 6. $\forall x,y \quad x+y\leq y+x$
- 7. $\forall w,x,y,z \quad w\leq y\wedge x\leq z \Rightarrow w+x\leq y+z$
- 8. $\forall x,y,z \quad x\leq y\wedge y\leq z \Rightarrow x\leq z$

我们希望用以上原子语句进行推理，得到 $7\leq 3+9$。注意在推理的过程中，我们只能使用以上 8 条公理，不能使用现实数学中的各种运算。

（1）假如我们使用反向链接算法（见课件上 backward chaining 部分）。我们可以得到以下推理过程，请完善推理过程。

Goal G0: $7\leq 3+9$, resolve with axiom 8 and $\{x0/7, z0/(3+9), y0/(7+0)\}$
*/*Use axiom 8, and substitute the (x, y, z) in axiom with (7, (7+0), (3+9)*/*
*/*To achieve Goal G0, we need to find a intermediate (7+0) and achieve Goal G1 and G2.*/*
Goal G1: $7\leq 7+0$, resolve with (a) ____________________. Goal G1 Succeeds.
Goal G2: $7+0\leq 3+9$, resolve with (b) ____________________.
*/*To achieve Goal G2, we need to find a intermediate and achieve Goal G3 and G4.*/*
Goal G3: $7+0\leq n$, resolve with (c) ____________________. Goal G3 Succeeds.
Goal G4: $n\leq 3+9$, resolve with (d) ____________________.
*/*To achieve Goal G4, we need to find a intermediate and achieve Goal G5 and G6.*/*

Goal G5: $0 \leq 3$, resolve with axiom 1. Goal G5 Succeeds.
Goal G6: $7 \leq 9$, resolve with axiom 2. Goal G6 Succeeds.
Goal G4 succeeds.
Goal G2 succeeds.
Goal G0 succeeds.

解答：(a) axiom 4 and $\{x1/7\}$
(b)axiom 8 and $\{x2/(7+0), z2/(3+9), y2/n\}$
(c)axiom 6 and $\{x3/7, y3/0, n/(0+7)\}$
(d)axiom 7 and $\{n/(0+7), w4/0, x4/7, y4/3, z4/9\}$

(2) 假如我们使用前向链接算法（见课件上 forward chaining 部分），我们可以怎样推理得到结论？请写出推理过程。

结合公理（1，2，7），并且带入 $\{w0/0, x0/7, y0/3, z0/9\}$，可以推理得到

- 9. $0+7 \leq 3+9$

使用公理 6，并且带入 $\{x1/7, y1/0\}$，可以推理得到

- 10. $7+0 \leq 0+7$

结合式子（9，10，8），并且带入 $\{x2/(7+0), y2/(0+7), z2/(3+9)\}$，可以推理得到

- 11. $7+0 \leq 3+9$

使用式子 4，并且带入 $\{x3/7\}$，可以推理得到

- 12. $7 \leq 7+0$

使用式子（12，11，8），并且带入 $\{x4/7, y4/(7+0), z2/(3+9)\}$，可以推理得到

- 13. $7 \leq 3+9$

4 第三题

给定下图1所示的贝叶斯网络。网络中有 (B, A, E, J, M) 五个变量。
（1）根据给定的贝叶斯网络对联合概率 $P(B, E, A, J, M)$ 进行因子分解。

解答:用链式法则来完成联合概率 $P(B, E, A, J, M)$ 的因子分解,分解结果为 $\prod_{i=1}^{n} P(X_i|parents(X_i))$。
回到本题，本题中有五个变量 (B, E, A, J, M)。我们一个一个来看。
对于变量 B，它没有父节点（parents），故我们可以得到式子 $P(B)$。
对于变量 E，它也没有父节点（parents），故我们可以得到式子 $P(E)$。
对于变量 A，它的父节点为 B 与 E, 故我们可以得到式子 $P(A|B, E)$。
对于变量 J，它的父节点为 A, 故我们可以得到式子 $P(J|A)$。
对于变量 M，它的父节点为 A, 故我们可以得到式子 $P(M|A)$。
最后将上面所有的式子连续相乘即可得到所有变量的联合概率的表示，即为

$$P(B, E, A, J, M) = P(B) \cdot P(E) \cdot P(A|B, E) \cdot P(J|A) \cdot P(M|A)$$

（2）对于变量 B 而言，我们给定哪一个或者两个变量的值，能够使得该变量条件独立于贝叶斯网络中的其他变量?

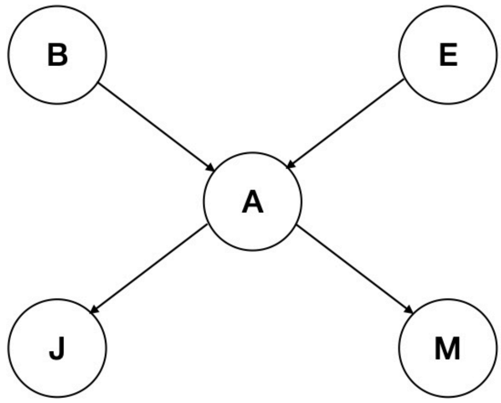


图 1: 第三题的贝叶斯网络

解答：对于一个贝叶斯网络中的节点而言，如果其马尔科夫毯中的节点的值被给定了，那么它就与网络中的其他节点条件独立了。这个马尔科夫毯包含了一个节点的子节点、父节点、子节点的其他父节点。

对于本题而言，对于变量 B 而言，属于其马尔可夫毯的变量有变量 A 与变量 E。如果给定了变量 A 与变量 E 的值，那么变量 B 就与网络中的变量 J，变量 M 条件独立。如果未给定变量 A 与变量 E 的值，那么变量 B 就与网络中的变量 J，变量 M 无条件独立的关系。

(3) 如果我们给定变量 A 的值，其他的变量 (B, E, J, M) 之间有哪些是条件独立的？如果我们不给定变量 A 的值呢？注意，贝叶斯网络间两个节点条件独立可以写为 $X_i \perp X_j$。

解答：寻找条件独立的过程，我们一般观察变量间的结构是否符合"common parent"，"cascade"，"V-structure"这三种结构。

观察发现，(B, A, M)，(B, A, J)，(E, A, J)，(E, A, M) 之间符合"cascade"结构。(B, A, E) 之间符合"V-structure"结构。(J, A, M) 符合"common parent"结构。

如果固定 A，那么根据三种结构的结论，可以发现 $B \perp M, B \perp J, E \perp J, E \perp M, J \perp M$，而在给定了 A 的值的情况下，B 与 E 耦合，两者间没有条件独立的关系。

如果不给定 A 的值，那么与上述的结果反过来，$B \perp E$，而 $(B, M), (B, J), (E, J), (E, M), (J, M)$ 间没有条件独立的关系。

(4) 我们希望求解贝叶斯网络中 $J = j, M = m, B = b$ 的概率，即 $P(B = b_0, J = j_0, M = m_0)$。请根据联合概率的表达式列出计算 $P(B = b_0, J = j_0, M = m_0)$ 的表达式。除此以外，请列出使用消元法求解 $P(B = b_0, J = j_0, M = m_0)$ 的过程。为了方便，我们要求消元的顺序是 $M \to J \to A \to E \to B$。

解答：我们要求解 $P(B = b, J = j, M = m)$，那么根据贝叶斯网络的链式法则，我们可以列出下面的式子 $P(B = b_0, J = j_0, M = m_0)$

$$= \sum_e \sum_a P(B = b_0)P(E = e)P(A = a|B = b_0, E = e)P(J = j_0|A = a)P(M = m_0|A = a)$$

在上面的式子中，变量 (B, J, M) 是给定的，变量 (A, E) 未给定、需要将其所有取值情况下算出的概率进行求和，上面的式子是比较严谨的表达方式。为了简便，有时我们也可以写成 $P(b)P(e)P(a|e, b)P(j|a)P(m|a)$，课件上面就是这么表述的。要注意这个式子里的 (b, e, a, j, m) 表

示的是变量 (B,E,A,J,M) 的所有合理取值。例如对于变量 J, 就只有 j_0 这么一种合理的取值。

如果要应用消元法进行求解，我们需要进行以下步骤：

根据规定的消元顺序，此时应该消去 M，那么就在 $\sum_e\sum_a P(B=b_0)P(E=e)P(A=a|B=b_0,E=e)P(J=j_0|A=a)P(M=m_0|A=a)$ 中找到所有与变量 M 有关的项，只找到 $P(M=m_0|A=a)$ 这一项，我们将该项列出，然后用一个函数 f 对其进行替代。可以发现，该函数 f 除了与变量 M 有关以外，还与变量 A 有关，但是基于变量 M 在此步被消去，因此剩下的求解式子中不应当出现 M、我们将该函数记为 $f(A)$，因此我们可以得到

$$P(B=b_0,J=j_0,M=m_0)=\sum_e\sum_a P(B=b_0)P(E=e)P(A=a|B=b_0,E=e)P(J=j_0|A=a)f(A)$$

$$f(A)=P(M=m_0|A=a)$$

如果题目中给定了概率表，那么我们就可以通过计算得到 $f(A)$。计算得到的结果是一个常数，原始求解的式子就被化简了。

接下来消去 J，与上面的过程类似，先找到上一步化简后的表达式中所有与 J 有关的式子，发现只有 $P(J=j_0|A=a)$ 这一项，那么我们可以得到

$$P(B=b_0,J=j_0,M=m_0)=\sum_e\sum_a P(B=b_0)P(E=e)P(A=a|B=b_0,E=e)f_2(A)f(A)$$

$$f_2(A)=P(J=j_0|A=a)$$

接下来消去 A，我们可以在上一步化简后的表达式中找到与 A 有关的式子，发现有 $P(A=a|B=b_0,E=e),f_2(A),f(A)$ 这三项。那么我们可以得到

$$P(B=b_0,J=j_0,M=m_0)=\sum_e P(B=b_0)P(E=e)f_3(e)$$

$$f_3(B,E)=\sum_a P(A=a|B=b_0,E=e)f_2(A)f(A)$$

接下来消去 E，我们可以在上一步化简后的表达式中找到与 E 有关的项 $P(E=e),f_3(B,E)$，由此可以得到

$$P(B=b_0,J=j_0,M=m_0)=P(B=b_0)f_4(B)$$

$$f_4(B)=\sum_e P(E=e)f_3(B,E)$$

最后再消去 B，可以得到

$$P(B=b_0,J=j_0,M=m_0)=f_5$$

$$f_5=P(B=b_0)f_4(B)$$

通过以上的过程我们就成功用消元法化简了 $P(B=b_0,J=j_0,M=m_0)$，在求解过程中只需要求解出 $f(A),f_2(A),f_3(B,E),f_4(B),f_5$ 这几个中间量就可以求解得到最终的结果。这个消元的过程与课件上消元求解过程在表达方法上有些差异，但是思想是一致的。你需要把握几个关键点：

- 寻找与待消去变量相关的项，即待消去元素作为条件概率的条件（分母）或条件概率的变量（分子）或函数的变量
- 用一个函数 $f(v)$ 表示所有相关项的求解积，除了待消去元素以外，这些相关项里的所有其他变量都需要写入函数的变量 v 中

- 将函数 $f(v)$ 放入原来的式子以代替之前提取出的所有待消去变量相关项，这样就从原来的式子中消去了一个变量。

本题中各种式子都比较复杂，大家可以注意一下，一般而言，我们用大写字母比如 A, B 来表示变量，它可以有多种可能的取值。我们一般用小写字母比如 a, b 来笼统表示变量的取值，一般用 a_0, b_0 等方式来表示某一个变量具体的取值。

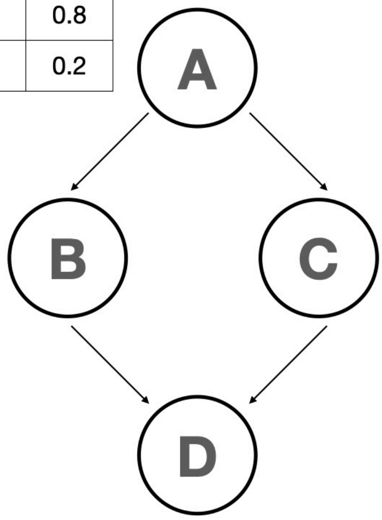
5 第四题

给定如下图2所示的贝叶斯网络模型。我们希望从给定的模型中通过取样的方式进行一些概率的估算，且我们规定采样时该模型各节点的拓扑顺序为 $(A \to B \to C \to D)$

A	P(A)
+a	0.8
-a	0.2

A	B	P(B|A)
+a	+b	0.8
+a	-b	0.2
-a	+b	0.5
-a	-b	0.5

B	C	D	P(D|B,C)
+b	+c	+d	0.3
+b	+c	-d	0.7
+b	-c	+d	0.1
+b	-c	-d	0.9
-b	+c	+d	0.2
-b	+c	-d	0.8
-b	-c	+d	0.9
-b	-c	-d	0.1



A	C	P(C|A)
+a	+c	0.7
+a	-c	0.3
-a	+c	0.1
-a	-c	0.9

图 2: 第四题的贝叶斯网络

示例：采用先验采样（prior sample）的方法生成样本。给定采样过程中的随机数为 $(0.31 \to 0.58 \to 0.04 \to 0.94 \to 0.67 \to 0.49 \to 0.37 \to 0.42)$。

- 首先我们取随机数 $r = 0.31$，而 $r = 0.31 < P(+a) = 0.8$，因此对于节点 A 我们取 $+a$
- 接着随机数 $r = 0.58 < P(+b| + a) = 0.8$，因此对于节点 B 我们取 $+b$
- 随机数 $r = 0.04 < P(+c| + a) = 0.7$，因此对于节点 B 我们取 $+c$
- 随机数 $r = 0.94 > P(+d| + b, +c) = 0.3$，因此对于节点 D 我们取 $-d$。这样我们通过一次采样得到了一个样本 $(+a, +b, +c, -d)$
- 随机数 $r = 0.67 < P(+a) = 0.8$，因此对于节点 A 我们取 $+a$
- 随机数 $r = 0.49 < P(+b| + a) = 0.8$，因此对于节点 B 我们取 $+b$

- 随机数 $r = 0.37 < P(+c|+a) = 0.7$，因此对于节点 C 我们取 $+c$
- 随机数 $r = 0.42 > P(+d|+b, +c) = 0.3$，因此对于节点 D 我们取 $-d$。这样我们通过一次取样获得了一个新的样本 $(+a, +b, +c, -d)$
- 采用了八个随机数，进行了两次采样，得到了两个样本，均为 $(+a, +b, +c, -d)$

（1）采用拒绝采样的方法（rejection sampling）计算 $P(-d|-b)$，请参照模型上方的示例写出采样过程（包括最后进行了几次采样，得到了怎样的样本）。采样时的随机数为 $(0.31 \to 0.58 \to 0.04 \to 0.94 \to 0.67 \to 0.49 \to 0.37 \to 0.42)$，且规定当随机数 $r < P(+a)$ 时采样 $+a$，当 $r \geq P(+a)$ 时采样 $-a$，随机数用完以后即废弃、当所有随机数用完整个算法停止。

解答：注意拒绝采样的过程中，如果 evidence varibale 的值与规定不符合 (本题中规定了 $-b$)，则本次采样立即终止。

- 首先我们取随机数 $r = 0.31$，而 $r = 0.31 < P(+a) = 0.8$，因此对于节点 A 我们取 $+a$
- 接着随机数 $r = 0.58 < P(+b|+a) = 0.8$，因此对于节点 B 我们取 $+b$，由于 $+b$ 不符合 $P(-d|-b)$ 条件部分对 B 节点的取值，因此我们拒绝该采样，我们终止此次采样
- 随机数 $r = 0.04 < P(+a) = 0.8$，因此对于节点 A 我们取 $+a$
- 随机数 $r = 0.94 > P(+b|+a) = 0.8$，因此对于节点 B 我们取 $-b$
- 随机数 $r = 0.67 < P(+c|+a) = 0.7$，因此对于节点 C 我们取 $+c$
- 随机数 $r = 0.49 > P(+b|-b, +c) = 0.2$，因此对于节点 D 我们取 $-d$，这样我们就采样得到了一个样本 $(+a, -b, +c, -d)$
- 随机数 $r = 0.37 < P(+a) = 0.37$，因此对于节点 A 我们取 $+a$
- 随机数 $r = 0.42 < P(+b|+a) = 0.8$，因此对于节点 B 我们取 $+b$，由于 $+b$ 不符合 $P(-d|-b)$ 条件部分对 B 节点的取值，因此我们拒绝该采样，我们终止此次采样。
- 采用了八个随机数，进行了三次采样，得到了一个样本，为 $(+a, -b, +c, -d)$

（2）采用似然采样的方法（likelihood weighting sampling）计算 $P(-d|-b)$，请参照模型上方的示例写出采样过程（包括最后进行了几次采样，得到了怎样的样本）。采样时的随机数为 $(0.31 \to 0.58 \to 0.04 \to 0.94 \to 0.67 \to 0.49)$，且规定当随机数 $r < P(+a)$ 时采样 $+a$，当 $r \geq P(+a)$ 时采样 $-a$，随机数用完以后即废弃、当所有随机数用完整个算法停止。

解答：注意在似然采样的过程中 evidence variable 的取值固定，并且每个样本要计算参数 w。

- 首先我们判断节点 A 不属于 evidence variable，那么我们取随机数 $r = 0.31$，而 $r = 0.31 < P(+a) = 0.8$，因此对于节点 A 我们取 $+a$
- 接下来，节点 B 在 $P(-d|-b)$ 中属于 evidence variable，我们取 $-b$，并且参数 $w = 1 \cdot P(-b|+a) = 0.2$
- 随机数 $r = 0.58 < P(+c|+a) = 0.7$，因此对于节点 C 我们取 $+c$
- 随机数 $r = 0.04 < P(+d|+c, -b) = 0.2$，因此对于节点 D 我们取 $+d$。这样我们就生成了一个样本 $(+a, -b, +c, +d, w = 0.2)$
- 随机数 $r = 0.94 > P(+a) = 0.8$，因此对于节点 A 我们取 $-a$
- 节点 B 在 $P(-d|-b)$ 中属于 evidence variable，我们取 $-b$，并且参数 $w = 1 \cdot P(-b|-a) = 0.5$

- 随机数 $r = 0.67 > P(+c|-a) = 0.1$，因此对于节点 C 我们取 $-c$
- 随机数 $r = 0.49 < P(+d|-b,-c) = 0.9$，因此对于节点 D 我们取 $+d$，这样我们就采样得到了一个样本 $(-a,-b,-c,+d,w=0.5)$
- 采用六个随机数，进行两次采样，得到两个样本，为 $(+a,-b,+c,+d,w=0.2)$, $(-a,-b,-c,+d,w=0.5)$

(3) 采用吉布斯采样的方法（Gibbs sampling）计算 $P(-d|-b)$，请参照模型上方的示例写出采样过程（包括最后进行了几次采样，得到了怎样的样本）。采样时的随机数为 $(0.31 \to 0.58 \to 0.04 \to 0.94 \to 0.67 \to 0.49)$，采样初时刻的节点取值初始化为 $(+a,-b,+c,+d)$，且规定当随机数 $r < P(+a)$ 时采样 $+a$，当 $r \geq P(+a)$ 时采样 $-a$，随机数用完以后即废弃、当所有随机数用完整个算法停止。

解答：

- 首先我们判断节点 B 属于 evidence variable，因此我们将其初始化为 $-b$，其他节点初始化为 $(+a,+c,+d)$。
- 取随机数 $r = 0.31$，我们仅考虑在 A 节点的马尔可夫毯之内的节点 B, C 两个节点的当前值，而 $r = 0.31 < P(+a|-b,+c) = 0.918$，因此对于节点 A 我们保持其取值 $+a$。这样我们就生成了一个样本 $(+a,-b,+c,+d)$。计算 $P(+a|-b,+c)$ 的过程如下所示，如果有不清楚的地方，可以阅读课件 lecture10 的第 6 页与第 10 页，里面都有对应的计算公式。

$$
\begin{aligned}
P(+a|-b,+c) &= \frac{P(+a,-b,+c)}{P(-b,+c)} \qquad \text{(条件概率计算)} \\
&= \frac{P(+a,-b,+c)}{P(-b,+c,+a)+P(-b,+c,-a)} \qquad \text{(节点 B 与 C 的取值都与 A 有关)} \\
&= \frac{P(+a)\cdot P(-b|+a)\cdot P(+c|+a)}{[P(+a)\cdot P(-b|+a)\cdot P(+c|+a)]+[P(-a)\cdot P(-b|-a)\cdot P(+c|-a)]} \\
&\quad \text{(贝叶斯网络的联合概率因子分解)} \\
&= \frac{0.8\times 0.2\times 0.7}{0.8\times 0.2\times 0.7+0.2\times 0.5\times 0.1} \\
&\approx 0.918
\end{aligned} \tag{1}
$$

- 节点 B 固定为 $-b$。A B D 节点都在 C 的马尔科夫毯之内，随机数 $r = 0.58 > P(+c|+a,-b,+d) = 0.341$，因此对于节点 C 我们取值 $-c$, 生成一个样本 $(+a,-b,-c,+d)$。

$$
\begin{aligned}
P(+c|+a,-b,+d) &= \frac{P(+a,-b,+c,+d)}{P(+a,-b,+d)} \qquad \text{(条件概率计算)} \\
&= \frac{P(+a,-b,+c,+d)}{P(+a,-b,+c,+d)+P(+a,-b,-c,+d)} \qquad \text{(节点 D 的取值与 C 有关)}
\end{aligned} \tag{2}
$$

根据贝叶斯网络的联合概率因子分解方法，我们有：

$$P(+a,-b,+c,+d) = P(+a)\cdot P(-b|+a)\cdot P(+c|+a)\cdot P(+d|-b,+c)$$

$$P(+a,-b,-c,+d) = P(+a)\cdot P(-b|+a)\cdot P(-c|+a)\cdot P(+d|-b,-c)$$

$$P(+c|+a,-b,+d) = \frac{0.8\times 0.2\times 0.7\times 0.2}{0.8\times 0.2\times 0.7\times 0.2+0.8\times 0.2\times 0.3\times 0.9} \approx 0.341$$

- 取节点 D 的马尔可夫毯之内的节点 B, C。随机数 $r = 0.04 < P(+d|-b,-c,) = 0.9$，$P(+d|-b,-c,)$ 是题目的已知条件，因此对于节点 D 我们保持取值 $+d$。这样我们就生成了一个样本 $(+a,-b,-c,+d)$。

- 接下来的计算过程与上面类似。随机数 $r = 0.94 > P(+a|-b,-c) = 0.348$，因此对于节点 A 的取值修改为 $-a$，因此取样 $(-a,-b,-c,+d)$

- 节点 B 在 $P(-d|-b)$ 中属于 evidence variable，我们取 $-b$。随机数 $r = 0.67 > P(+c|-a,-b,+d) = 0.024$，因此对于节点 C 的取值我们保持为 $-c$，因此我们取样 $(-a,-b,-c,+d)$

- 随机数 $r = 0.49 < P(+d|-b,-c) = 0.9$，因此对于节点 D 我们保持取值 $+d$，这样我们就采样得到了一个样本 $(-a,-b,-c,+d)$

- 采用了六个随机数,进行了六次采样,得到了六个样本,为 $(+a,-b,+c,+d)$, $(+a,-b,-c,+d)$, $(+a,-b,-c,+d)$, $(-a,-b,-c,+d)$, $(-a,-b,-c,+d)$, $(-a,-b,-c,+d)$。

- 本题的几个关键点：（1）在改变一个节点的状态时，仅考虑其马尔可夫毯以内的变量，并将其写入条件概率公式中的条件部分（2）计算贝叶斯网络中的联合分布概率，善于运用因子分解的方法将联合概率分布拆分成多个简单条件概率相乘，（3）对于形如 $P(+a,-b,+d)$ 的联合概率，因为该概率拆成多个简单条件概率的时候，发现节点 D 的条件之一为 C，因此该联合概率要拆分成 $P(+a,-b,+c,+d) + P(+a,-b,-c,+d)$（4）每尝试改变一个状态就取一个样，因此我们最后采到了六个样本，这点与前面介绍的贝叶斯采样方法不同。

6 第五题

给定一个隐马尔可夫模型（HMM）。

（1）运用 HMM 的建模方式，用条件概率的计算方式推导并化简 $P(x_1,...,x_t,y_{t-1}=s_v,y_t=s_j)$。

提示：一般来说，我们会对 HMM 模型有以下的建模方式（与课件上一致)——x_t 表示 t 时刻的观察状态，y_t 表示 t 时刻的隐藏状态, 隐藏状态可能取值为 $\{1,2,3,\cdots,M\}$。初始状态的概率（start probabilities）表示为 $\{\pi_1,\cdots,\pi_M\}$，隐藏状态之间从状态 i 转化为状态 j 的转移概率（transition probabilities）为 $a_{i,j}$，隐藏状态 y_j 与观察状态 x_i 之间的发射概率（emission probabilities）为 $b_j(x_i)$。在使用前向算法推导序列概率的时候, 设定的前向因子 $\alpha_t^i = P(x_1,...,x_t,y_t=s_i)$。

解答：本题的关键是理解条件概率的计算公式以及 HMM 模型中各种条件独立关系。

根据条件概率定义，我们有

$$P(x_1,...,x_t,y_{t-1}=s_v,y_t=s_j) = [\sum_{x_t} P(x_t,y_t=s_j|x_1,...,x_{t-1},y_{t-1}=s_v)] \cdot P(x_1,...,x_{t-1},y_{t-1}=s_v)$$

根据我们对前向因子 α_t^i 的定义，我们可知

$$P(x_1,...,x_{t-1},y_{t-1}=s_v) = \alpha_{t-1}^v$$

接下来化简 $\sum_{x_t} P(x_t,y_t=s_j|x_1,...,x_{t-1},y_{t-1}=s_v)$，根据 HMM 结构图，结合贝叶斯网络间节点的条件独立关系（HMM 可以视为一种特殊的贝叶斯网络），我们可知节点 X_t 与 Y_t 的马尔科夫毯合并后为 $\{X_t,Y_t,Y_{t-1}\}$。分子部分的 x_t,y_t 均已经固定，而条件部分中的变量 Y_{t-1} 已经固定为 s_v，因此节点 X_t、Y_t 两个节点都与 $(X_1,...,X_{t-1})$ 都条件独立，可以得到

$$\sum_{x_t} P(x_t,y_t=s_j|x_1,...,x_{t-1},y_{t-1}=s_v) = \sum_{x_t} P(x_t,y_t=s_j|y_{t-1}=s_v)$$

进一步化简 $\sum_{x_t} P(x_t, y_t = s_j | y_{t-1} = s_v)$，我们可以在用一次条件概率的计算公式，得到

$$\sum_{x_t} P(x_t, y_t = s_j | y_{t-1} = s_v) = \sum_{x_t} P(x_t | y_t = s_j, y_{t-1} = s_v) P(y_t = s_j | y_{t-1} = s_v)$$

再一次使用 HMM 节点间的条件独立关系，可知节点 Y_{t-1} 与节点 X_t 之间是 cascade 结构，在给定了中间节点 Y_t 的值为 s_j 的情况下，节点 Y_{t-1} 与节点 X_t 之间条件独立。因此可以得到

$$\sum_{x_t} P(x_t | y_t = s_j, y_{t-1} = s_v) = \sum_{x_t} P(x_t | y_t = s_j) = \sum_{x_t} b_j(x_t)$$

同时我们有

$$P(y_t = s_j | y_{t-1} = s_v) = a_{v,j}$$

至此化简完成，整体的计算过程如下所示：

$$\begin{aligned}
& P(x_1, ..., x_t, y_{t-1} = s_v, y_t = s_j) && (3) \\
= & \sum_{x_t} P(x_t, y_t = s_j | x_1, ..., x_{t-1}, y_{t-1} = s_v) P(x_1, ..., x_{t-1}, y_{t-1} = s_v) && (4) \\
= & \sum_{x_t} P(x_t, y_t = s_j | y_{t-1} = s_v) \alpha_{t-1}^v && (5) \\
= & \sum_{x_t} P(x_t | y_t = s_j, y_{t-1} = s_v) P(y_t = s_j | y_{t-1} = s_v) \alpha_{t-1}^v && (6) \\
= & \sum_{x_t} b_j(x_t) a_{v,j} \alpha_{t-1}^v && (7)
\end{aligned}$$

考虑到实际情况下，x_t 可能是给定的固定值，因此最后的结果写答案的简便形式 $P(x_1, ..., x_t, y_{t-1} = s_v, y_t = s_j) = b_j(x_t) a_{v,j} \alpha_{t-1}^i$ 也是正确的，但是要注意理解，式子里的 x_t 可能为给定的值、也可能有多种取值情况。

(2) 我们给定的 HMM 模型如下图3所示。假定图中的圆圈表示 A, B, C 三种可能的隐藏状态，而圆圈中的数字表示特定隐藏状态下的观察状态以及其发射概率（emission probabilities）。圆圈之间的箭头表示隐藏状态的转移以及相应的转移概率（transition probabilities）。我们用 $p_{y=A}^t$ 表示在 t 时刻，隐状态为 A 的概率，用 $p_{x=1}^t$ 表示在 t 时刻，观察状态为 1 的概率。为了简便，我们规定初始的隐状态为 A，即 $p_A^1 = 1$。在每个时刻，隐状态会通过转移概率确定下个时刻的状态。

(a) 请用图中给定的数值与符号计算 $p_{y=C}^3$。

(b) 假设在 t 时刻，$(p_{y=B}^t, p_{y=C}^t)$ 分别为 (b_0, c_0)，请用图中给定的数值与符号计算 $p_{x=2}^t$。

解答：本题的关键是看懂 HMM 模型的这张图，要注意只有存在箭头的部分之间可以实现状态的转移，且每个时刻会通过转移概率确定下个时刻的隐藏状态。

想要求 $p_{y=C}^3$，那么需要规划从 $t = 1$ 到 $t = 3$ 时刻的隐状态变化序列，且起始状态为 A，终止状态为 C。根据图中所示，只可能有 (A, A, C) 这一种存在的可能性，因此我们可以得到（a 表示状态转移概率）

$$p_{y=C}^3 = a_{AA} \cdot a_{AC} = A1 \cdot A2$$

想要求 $p_{x=2}^t$，那么我们需要寻找哪些隐藏状态下我们可能有观察状态 2，最终发现只有当隐藏状态 A 下，我们有观察状态 2。因此我们可以得到

$$p_{x=2}^t = p_{y=A}^t \cdot b_A(2) = (1 - p_{y=B}^t - p_{y=C}^t) \cdot b_A(2) = (1 - c_0 - b_0) \cdot 0.3$$

关键是理解在某一时刻，隐状态必定为 A, B, C 其一，因此 $p_{y=A}^t + p_{y=B}^t + p_{y=C}^t = 1$。

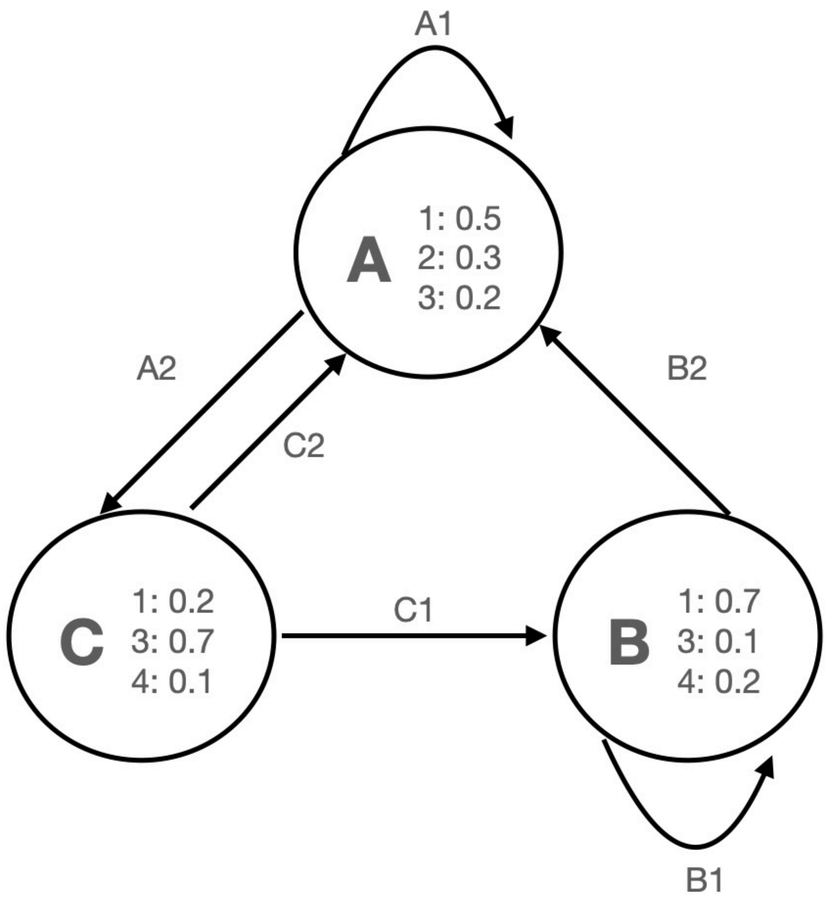


图 3: 第五题第二小问的 HMM 模型