课后练习1-3章

人工智能学院 181220076 周韧哲

1.1

- Agent: 用传感器来感受环境并用执行器来与环境交互作用的事物
- Agent函数：一种数学描述，将Agent的感知序列映射为动作
- Agent程序：Agent函数在代码上的具体实现
- 理性：每一步动作都选择让目标期望收益最大化的
- 自主：有学习能力，能够通过学习改正或完善自身知识
- 反射Agent：仅仅根据当前的感知来选择行动
- 基于模型的Agent：内部有模型会根据感知历史来维持内部状态，模型是关于世界如何运作的世界模型
- 基于目标的Agent：能用目标信息来描述想要达到的状态
- 基于效用的Agent：能使用更为细致通用的性能度量来对状态进行赋值，即效用函数
- 学习Agent：由学习元件、评判元件、性能元件和问题产生器组成的Agent

1.2

```
#-------基于目标的Agent--------#

def goal_based_agent():
    state=UPDATE_STATE(state,action,percept,model)
    actions=GET_AVALIABLE_ACTION(state)     #获得当前状态下所有合法的行动
    next_states=UPDATE_STATE_IN_MODEL(state,actions,action_history)   #在model中采取行动得到下一个状态列表
    action=CHOOSE_ACTION_BASED_GOAL(actions,next_states,goal) #根据目标选择最优动作
    return action

#-------基于效用的Agent--------#

def utility_based_agent():
    state=UPDATE_STATE(state,action,percept,model)
    actions=GET_AVALIABLE_ACTION(state)     #获得当前状态下所有合法的行动集合
    next_states=UPDATE_STATE_IN_MODEL(state,actions,action_history)   #在model中采取行动得到下一个状态集合
    action=CHOOSE_ACTIONS_BASED_GOAL(actions,next_states,goal) #根据目标选择最优动作集合
    best_action=CHOOSE_ACTION_BASED_UTILITY(action,state) #根据效用函数选择最优动作
    return best_action
```

2.1

令X,Y表示随机变量生命和水，由题意知：

X	Y	P(X,Y)
0	0	0.25
0	1	0
1	0	0.25
1	1	0.5

则有：$P(X=1|Y=1)=\frac{P(X=1,Y=1)}{P(Y=1)}=\frac{0.5}{0.5}=1$

故在给定有水的前提下，火星上有生命的概率为 1

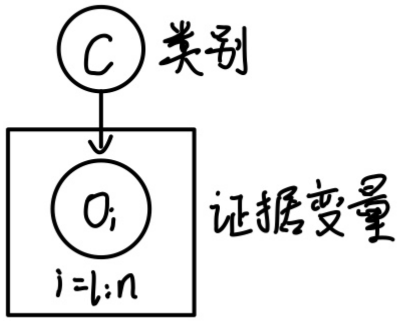
2.2

- 我们有：
$$\begin{aligned}P(S_t|O_{0:t}) &= P(S_t|O_{0:t-1},O_t)\\ &= \frac{P(O_t|S_t,O_{0:t-1})P(S_t|O_{0:t-1})}{P(O_t|O_{0:t})}\end{aligned}$$
而:
$$\begin{aligned}P(O_t|O_{0:t-1}) &= \frac{P(O_{0:t})}{P(O_{0:t-1})}\\ &= \frac{\sum_i^N \alpha_t(i)}{\sum_i^N \alpha_{t-1}(i)}\end{aligned}$$
由前向概率算法，此式可算出，从而得到
$$P(S_t|O_{0:t}) \propto P(O_t|S_t,O_{0:t-1})P(S_t|O_{0:t-1})$$

- 由观测独立性假设：$P(O_t|S_t,O_{0:t-1})=P(O_t|S_t)$,
再由全概率公式：
$$\begin{aligned}P(S_t|O_{0:t-1}) &= \sum_{s_{t-1}} P(S_t|S_{t-1},O_{0:t-1})P(S_{t-1}|O_{0:t-1})\\ &= \sum_{s_{t-1}} P(S_t|S_{t-1})P(S_{t-1}|O_{0:t-1})\end{aligned}$$
故$P(O_t|S_t,O_{0:t-1})P(S_t|O_{0:t-1})=P(O_t|S_t)\sum_{s_{t-1}}P(S_t|S_{t-1})P(S_{t-1}|O_{0:t-1})$.

2.3

- 分类任务是从给定的一组观察或特征中推理所属类别。
- 朴素贝叶斯模型的假设是给定所属类别，证据变量之间条件独立，即对于所有$i\neq j$,有 $(o_i\perp o_j|C)$.

- 

2.4

- 令随机变量S,F,E分别表示睡眠充足，上课睡觉和红眼，其值域都为{0,1}，从而可以得出四个表：

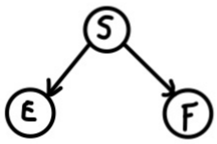
S	P
1	0.7
0	0.3

S_{t-1}	S_t	$P(S_t\|S_{t-1})$
1	1	0.8
1	0	0.2
0	1	0.3
0	0	0.7

S_t	E_t	$P(E_t\|S_t)$
1	1	0.2
1	0	0.8
0	1	0.7
0	0	0.3

S_t	F_t	$P(F_t\|S_t)$
1	1	0.1
1	0	0.9
0	1	0.3
0	0	0.7

  对应的贝叶斯网络结构为：

  表已给出了先验概率分布$P(S)$，状态转移分布$P(S_t|S_{t-1})$,观察分布$P(E_t|S_t), P(F_t|S_t)$,从而可以进行滤波和预测。

- 令O为观察变量，取值为O_1, O_2, O_3, O_4，分别表示*上课睡觉且有红眼*、*上课睡觉且无红眼*、*上课不睡觉且有红眼*、*上课不睡觉且无红眼*。由条件独立性可由E,F得出观察分布，从而可得出以下几个表：

S	P
1	0.7
0	0.3

S_{t-1}	S_t	$P(S_t\|S_{t-1})$
1	1	0.8
1	0	0.2
0	1	0.3
0	0	0.7

S_t	O_t	$P(O_t\|S_t)$
1	O_1	0.02
1	O_2	0.08
1	O_3	0.18
1	O_4	0.72
0	O_1	0.21
0	O_2	0.09
0	O_3	0.49
0	O_4	0.21

此为隐含状态S、观察状态O的隐马尔可夫模型。

2.5

- $o_0 = O_4, o_1 = O_3, o_2 = O_1$,令$S$表示$EnoughSleep$,由前向算法，可得（向量的第一项表示s=1，第二项表示s=0,以下运算均保留两位小数）：

$$
\begin{aligned}
&P(s_0) = (0.7, 0.3) \\
&P(s) = \sum_{s_0} P(s|s_0)P(s_0) = (0.65, 0.35) \\
&b_0(s) = P(o_0|s)P(s) = (0.8643, 0.1357) \to (0.86, 0.14) \\
&b_1(s) = (0.51, 0.49) \\
&b_2(s) = (0.10, 0.90)
\end{aligned}
$$

故

$$
\begin{aligned}
&P(s_0|o_0) = (0.86, 0.14) \\
&P(s_1|o_{0:1}) = (0.51, 0.49) \\
&P(s_2|o_{0:2}) = (0.10, 0.90)
\end{aligned}
$$

- 由前向-后向算法：

$$
\begin{aligned}
&P(s_o|o_{0:2}) = P(s_0|o_0)P(o_{1:2}|s_0) = P(s_0|o_0)\sum_{s_1} P(o_1|s_1)P(s_1|s_0)\sum_{s_2} P(o_2|s_2)P(s_2|s_1) = (0.73, 0.27) \\
&P(s_1|o_{0:2}) = P(s_1|o_{0:1})P(o_2|s_1) = (0.28, 0.72) \\
&P(s_2|o_{0:2}) = (0.10, 0.90)
\end{aligned}
$$

- $t = 0$时的滤波概率为$(0.86, 0.14)$，平滑概率为$(0.73, 0.27)$，滤波下s=1的概率比平滑下的大，即o_1, o_2信息的加入降低了s=1的概率，而由于其都表现了有红眼故这是容易理解的：未来的观察对当前状态有影响。
  $t = 1$时的滤波概率为$(0.51, 0.49)$，平滑概率为$(0.28, 0.72)$,滤波下s=0的概率比平滑下的小，即o_2信息的加入增加了s=0的概率，而由于其有红眼故这也是容易理解的。
  即平滑整合了未来的信息，比滤波更具有预测性。

2.6

- 图的拓扑排序是一个结点的有序列表，使得如果图中有边$A \to B$，则A出现在B之前。

- 拓扑排序使得贝叶斯网络中的采样可以从条件概率分布中采样，因为由于链式法则，在采样一个随机变量之前需要先采样其父节点。
- 拓扑排序总是存在，但不唯一。
- 拓扑排序：A,B,D,C,F,E。

2.7

- 吉布斯采样法的缺点是，样本之间存在相关性，不容易进入到稳态分布。为了减少样本之间的相关性，可以每隔固定次数采样一次。为了采样进入到一个稳态分布的样本序列，舍弃从初始样本出发不久采样到的样本。
- 极大似然估计的一个重大缺陷是当数据集足够小时，使得某些事件不能被观测到，则会认为其发生的概率为0。
- 贝叶斯学习的优点有：用所有假说做预测，而不是使用单个“最好”的假说，可归约为概率推理，缺点是需要大规模求和或积分。

2.8

对数似然为：

$$\begin{aligned} l(m,b,\sigma^2) &= ln\mathcal{N}(x_{1:n}|m,b,\sigma^2) \\ &= \sum_{j=1}^{n} ln(\frac{1}{\sqrt{2\pi}\sigma}exp(-\frac{(x_j-my_j-b)^2}{2\sigma^2})) \\ &= n(-ln\sqrt{2\pi}-ln\sigma)-\sum_{j=1}^{n}\frac{(x_j-my_j-b)^2}{2\sigma^2} \end{aligned}$$

对m,b,σ分别求偏导令其为0：

$$\begin{aligned} \frac{\partial l}{\partial m} &= -\sum_{j=1}^{n}\frac{1}{\sigma^2}(my_j+b-x_j)y_j = 0 \\ \frac{\partial l}{\partial b} &= -\sum_{j=1}^{n}\frac{1}{\sigma^2}(my_j+b-x_j) = 0 \\ \frac{\partial l}{\partial \sigma} &= -\frac{n}{\sigma}+\sum_{j=1}^{n}\frac{1}{\sigma^3}(my_j+b-x_j)^2 = 0 \end{aligned}$$

解得：

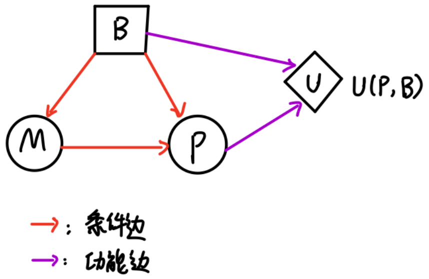
$$\begin{aligned} m^* &= \frac{n\sum_{j=1}^{n}x_jy_j-\sum_{j=1}^{n}x_j\sum_{j=1}^{n}y_j}{n\sum_{j=1}^{n}y_j^2-(\sum_{j=1}^{n}y_j)^2} \\ b^* &= \frac{\sum_{j=1}^{n}x_j}{n}-\frac{\sum_{j=1}^{n}y_j(n\sum_{j=1}^{n}x_jy_j-\sum_{j=1}^{n}x_j\sum_{j=1}^{n}y_j)}{n(n\sum_{j=1}^{n}y_j^2-(\sum_{j=1}^{n}y_j)^2)} \\ \sigma^{*2} &= \frac{1}{n}\sum_{j=1}^{n}(x_j-m^*y_j-b^*)^2 \end{aligned}$$

3.1

Pat更可能买到更好的车，因为在其他条件相同的情况下，Pat获得的信息更多。如果以面值作为车的效用，则Pat更可能感到失望。

3.2

- 决策网络为：



- 首先用全概率公式计算$P(p|B)=(P(p|b),P(p|\neg b))$：

$$
\begin{aligned}
P(p|B) &= \sum_m P(p|B,m)P(m|B) \\
&= (0.9\times 0.9+0.5\times 0.1, 0.8\times 0.7+0.3\times 0.3) \\
&= (0.86, 0.65)
\end{aligned}
$$

  从而计算B的期望效用$EU(B)=(EU(b),EU(\neg b))$：

$$
\begin{aligned}
EU(B) &= \sum_p P(p|B)U(p,B) \\
&= (0.86\times(2000-100)+0.14\times(0-100), 0.65\times 2000+0.35\times 0) \\
&= (1620, 1300)
\end{aligned}
$$

  所以购买教材的期望效用为1620，不够买教材的期望效用为1300。

- 由**最大化期望效用原则**，Sam应该买教材。

3.3

- 信息价值是获取信息之后和之前的最优行动的期望价值之间的差。当一个观测不改变最优行动时，它的信息价值是0
- 假设Agent在观察**o**下的最优动作为a，则在变量O'下，由行动的期望效用定义，有：

$$
EU^*(a^{o'}|\mathbf{o},o')\geq EU(a|\mathbf{o},o')\text{，}a^{o'}\text{表示在观察}(\mathbf{o},o')\text{下的最优动作}
$$

  不等式左右两边都对o'做概率累加，得到：

$$
\sum_{o'} P(o'|\mathbf{o})EU^*(a^{o'}|\mathbf{o},o')\geq \sum_{o'} P(o'|\mathbf{o})EU^*(a|\mathbf{o},o')
$$

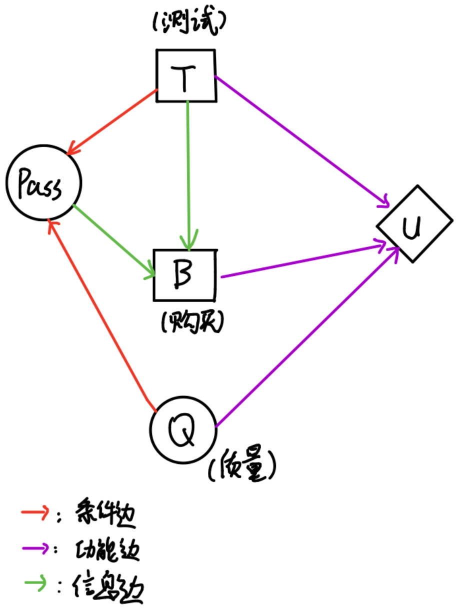
  右边即为$EU^*(\mathbf{o})$。左边即为$\sum_{o'} P(o'|\mathbf{o})EU^*(\mathbf{o},o')$。

  所以有：

$$
VOI(O'|\mathbf{o})=(\sum_{o'} P(o'|\mathbf{o})EU^*(\mathbf{o},o'))-EU^*(\mathbf{o})\geq 0
$$

3.4

- 决策网络为：



- 不测试购买的期望净获利为

$$
\begin{aligned}
P(q^+(c_1))U(q^+, b, \neg t) + P(q^-(c_1))U(q^-, b, \neg t) &= 0.7 \times (2000 - 1500) + 0.3 \times (2000 - 1500 - 700) \\
&= 290
\end{aligned}
$$

- 易得车通过测试的概率为

$$P(Pass) = P(Pass|q^+)P(q^+) + P(Pass|q^-)P(q^-) = 0.8 \times 0.7 + 0.35 \times 0.3 = 0.665$$

所以车不通过测试的概率为0.335。由贝叶斯定理，得到

$$
\begin{aligned}
P(q^+|Pass) &= \frac{P(Pass|q^+)P(q^+)}{P(Pass)} = 0.8421 \\
P(q^+|\neg Pass) &= \frac{P(\neg Pass|q^+)P(q^+)}{P(\neg Pass)} = 0.4179 \\
P(q^-|Pass) &= 1 - 0.8421 = 0.1579 \\
P(q^-|\neg Pass) &= 1 - 0.4179 = 0.5821
\end{aligned}
$$

- 当通过测试时，买车与不买车的期望效用为：

$$
\begin{aligned}
\sum_q P(q|Pass)U(q, b, t) &= 0.8421 \times (2000 - 1500 - 50) + 0.1579 \times (2000 - 1500 - 700 - 50) \\
&= 339.47 \\
\sum_q P(q|Pass)U(q, \neg b, t) &= 0.8421 \times (-50) + 0.1579 \times (-50) \\
&= -50
\end{aligned}
$$

当没通过测试时，买车与不买车的期望效用为：

$$
\begin{aligned}
\sum_q P(q|\neg Pass)U(q, b, t) &= 0.4179 \times (2000 - 1500 - 50) + 0.5821 \times (2000 - 1500 - 700 - 50) \\
&= 42.055 \\
\sum_q P(q|\neg Pass)U(q, \neg b, t) &= 0.4179 \times (-50) + 0.5821 \times (-50) \\
&= -50
\end{aligned}
$$

不管是否通过测试，买车的期望效用都大于不买车的，所以最优决策是买车。

- 不进行测试时，买车的期望效用为290大于不买车的期望效用0。所以测试之后最优决策并未改变，测试的信息价值为0。最优条件规划就是不进行测试直接买车。

3.5

效用矩阵：

	揭发	**沉默**
揭发	-5 : -5	0 : -4
沉默	-4 : 0	-1 : -1

- 当B揭发时，A的最优反应为沉默；当B沉默时，A的最优反应为揭发。所以不存在占优策略均衡。
- 有两个纳什均衡：（揭发，沉默）、（沉默，揭发）。