鲁东大学2022-2023学年第一学期期末考试
《人工智能基础》答题纸

学号：2020220202823　　姓名：朱相颐　　班级：电信2001

一、搜索

1)

Search Algorithm	A-B-D-G	A-C-D-G	A-B-A-B-C-D-G	A-B-C-D-F-G
Depth first search	√	√	√	√
Breadth first search	√	√	X	X
UCS	X	X	X	X √
A* with h_1	X	X	X	X √
A* with h_2	X	X	X	√

2) ① [0,12]

② [9,10]

③ ~~[12,12]~~ [12.5,13]

(2)

1) $(S_1, S_2, S_3, \ldots, S_n)$ $S \in \{1,2,\ldots,m\}$

m^n

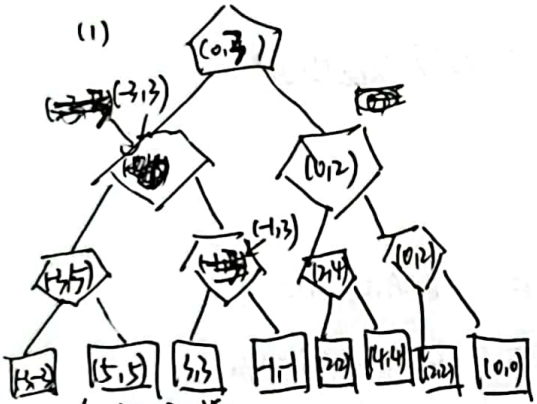
2) $5^{\frac{mn}{2}}$

3) 是不可接受的　eg.

在左图中，有3对pacman没有成对，但只需至少一步便可成对，预估值大于实际值，所以是不可接受的

二、博弈

(1)



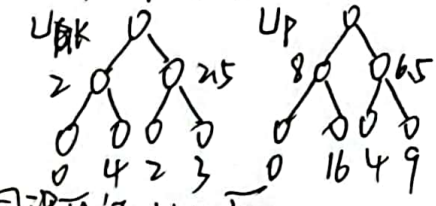
(2) $u \leftarrow$ ~~max~~ min(uList)

$v \leftarrow \max(\max(uList), \min(vList))$

三、效用函数

(1) 选择 $U_p = U_{GK}^2$ 和 $U_p = \sqrt{U_{GK}}$

解释：eg



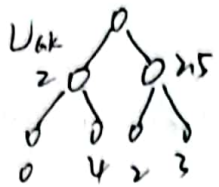
⇒有相同的outcome偏好，但对lottery偏好不同

由此同理可证 $U_p = \sqrt{U_{GK}}$

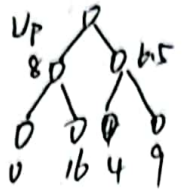
学号：2020222000913　　姓名：朱相颐　　班级：电信2001

(2) 选择 $U_P = \mathbb{E}U_{Gk}^2$

解释：



此时 Gk 会选择 outcome 为 2 有收益的右边路



此时，Pacman 会选择能得到最大收益的右边路径
但也同时会获得到 0 收益的风险
∴ 更也冒险

同理可证 $U_P = \sqrt{U_{Gk}}$ 时，也Gk更冒险.

四、马尔科夫决策过程.

由题意可知 1、2 点时可能会继续 Roll
因此最终状态下 $V(s_1) = V(s_2)$

$V(s_1) = -1 + \frac{1}{6}(V(s_1) + V(s_2) + 3 + 4 + 5 + 6)$
$V(s_2) = -1 + \frac{1}{6}(V(s_1) + V(s_2) + 3 + 4 + 5 + 6)$
可解得 $V(s_1) = V(s_2) = 3$

(1)

State	s_1	s_2	s_3	s_4	s_5	s_6
$\pi(s)$	Roll	Roll	Stop	Stop	Stop	Stop
$V_\pi(s)$	3	3	3	4	5	6

(2)

State	s_1	s_2	s_3	s_4	s_5	s_6
$\pi(s)$	Roll	Roll	Stop	Stop	Stop	Stop
$\pi'(s)$	Roll	Roll	Roll/Stop	Stop	Stop	Stop

old policy average reward: $-1 + \frac{3+3+3+4+5+6}{6} = 3$.

new policy：若 1 点 stop，则 $R(s_1) = 1 < 3$ ✗
若 2 点 stop，则 $R(s_2) = 2 < 3$
若 3 点 stop，则 $R(s_3) = 3 = 3$.

综上，1、2 点 Roll　3 点可 Roll 可 Stop
4 点 stop　5 点 stop
6 点 stop $R(s_4) = 4 > 3$
同理可证 5 点

(3) 是最优的，当前决策下平均收益为 3，1 点、2 点 stop 的话收益均小于 3
而 4、5、6 点 stop 的直接收益要大于 3.
∴ 是最优的

五、强化学习

(1) $Q(A, Down) = 2$　　$Q(B, Up) = 2$.

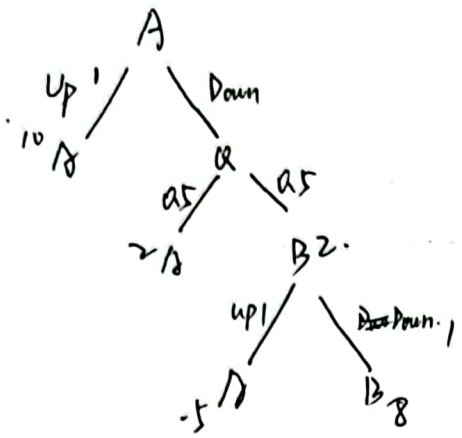
$Q(A, Down) = \frac{1}{2} \cdot 0 + \frac{1}{2}(4 + \frac{1}{2} \cdot 0) = 2$
$Q(B, Down) = \frac{1}{2} \cdot 0 + \frac{1}{2}(-4 + \frac{1}{2} \cdot 0) = -2$.
$Q(B, Up) = \frac{1}{2} \cdot 0 + \frac{1}{2}(0 + \frac{1}{2} \cdot 0) = 0$
$Q(B, Up) = \frac{1}{2} \cdot 0 + \frac{1}{2}(3 + \frac{1}{2} \cdot 2) = 2$

(2) $\hat{T}(A, Up, A) = 1$　　$\hat{R}(A, Up, A) = -1$
$\hat{T}(A, Up, B) = 0$　　$\hat{R}(A, Up, B) = n/a$
$\hat{T}(B, Up, A) = \frac{1}{2}$　　$\hat{R}(B, Up, A) = 3$
$\hat{T}(B, Up, B) = \frac{1}{2}$　　$\hat{R}(B, Up, B) = 0$

学号：2020220283　姓名：宋相颐　利秋次：电信2001

(3)

(i) $\pi^*(A)=$ Up　$\pi^*(B)=$ Down　$V^*(A)=20$　$V^*(B)=16$



B-Down-B : $2+V(B)$

A-Down-B : ~~$0+\frac{1}{2}(2+\frac{1}{2}\cdot 0)$~~ ⊗ $2+V(B)$ ×

A-Up-B : $10+V(B)$

A-Down-B-Up-B : $2+5+V(B)$ ~~$1+\frac{1}{2}\cdot 5+\frac{1}{2}\cdot 1+\frac{1}{2}\cdot 10=$~~ ×

A-Up-B-Up-B : $10+10+V(B)$ ✓

B-Down-B-Down-B : $2+8+V(B)$ ✓

~~$\therefore \pi^*(A)=$ Up~~

~~B-Down-B ⇒ 8 ✓~~

~~B-Up-B ⇒ 5 ×~~

$\therefore \pi^*(B)=$ Down · $\pi^*(B)=$ Up

由此代下去可得

$V^*(A)=10+10\times\frac{1}{2}+10\times(\frac{1}{2})^2+\cdots=20.$

$V^*(B)=8+8\times\frac{1}{2}+8\times(\frac{1}{2})^2+\cdots=16.$

(ii) 最终会收敛到上面所求得的 V^*

将其输送入 Q-learning 算法，可以找到局部最优解 V^* 但不能找到全局最优解 V^*，要想找到全局最优解，需要更多的关于这个世界的数据