人工智能基础

Homework 7

2.

（1）

状态空间：$\Omega=\{2,1,0,-1,-2\}$

状态转移矩阵：

$$
\begin{array}{c|ccccc}
 & -2 & -1 & 0 & 1 & 2 \\
\hline
-2 & 1 & 0 & 0 & 0 & 0 \\
-1 & q & r & p & 0 & 0 \\
0 & 0 & q & r & p & 0 \\
1 & 0 & 0 & q & r & p \\
2 & 0 & 0 & 0 & 0 & 1
\end{array}
$$

或简写为：

$$
P=\begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
q & r & p & 0 & 0 \\
0 & q & r & p & 0 \\
0 & 0 & q & r & p \\
0 & 0 & 0 & 0 & 1
\end{bmatrix} \tag{1}
$$

（2）

$$
P^2=\begin{bmatrix}
1 & 0 & 0 & 0 & 0 \\
q+qr & r^2+pq & 2pr & p^2 & 0 \\
q^2 & 2qr & r^2+2pq & 2pr & p^2 \\
0 & q^2 & 2qr & r^2+pq & p+pr \\
0 & 0 & 0 & 0 & 1
\end{bmatrix} \tag{2}
$$

可见从状态1再赛两局（包括一局）结束游戏（进入状态2或-2）的概率为：

$$
p_2=0+p+pr=p(1+r) \tag{3}
$$

（如果必须要赛两局，一局不包括在内，则为$p_2'=pr$。）

3.

（1）

用矩阵形式，令$\boldsymbol{v}=\begin{bmatrix}v_A\\v_B\\v_C\end{bmatrix}$，$\boldsymbol{r}=\begin{bmatrix}-1\\-1\\0\end{bmatrix}$则，由贝尔曼期望方程，有：

$$
\boldsymbol{v}=\boldsymbol{r}+\gamma\boldsymbol{P}\boldsymbol{v} \tag{4}
$$

其中状态转移矩阵为：

$$P=\begin{bmatrix}0 & 0.5 & 0.5\\0.5 & 0 & 0.5\\0 & 0 & 1\end{bmatrix} \tag{5}$$

求解线性方程组，得：

$$\boldsymbol{v}=(\boldsymbol{I}-\gamma\boldsymbol{P})^{-1}\boldsymbol{r}=\begin{bmatrix}\frac{16}{15} & \frac{4}{15} & \frac{2}{3}\\\frac{4}{15} & \frac{16}{15} & \frac{2}{3}\\0 & 0 & 2\end{bmatrix}\cdot\begin{bmatrix}-1\\-1\\0\end{bmatrix}=\begin{bmatrix}-\frac{4}{3}\\-\frac{4}{3}\\0\end{bmatrix} \tag{6}$$

即状态A,B,C的状态价值分别为$-\frac{4}{3},-\frac{4}{3},0$。

(2)

在简单情况下，可以直接由（6）式求解。但问题较为复杂时，矩阵维数很高，求逆十分困难甚至很可能不存在逆矩阵，这时可以采用迭代方法求解。

根据以下式子：

$$\boldsymbol{v}^{(t+1)}=\boldsymbol{r}+\gamma\boldsymbol{P}\boldsymbol{v}^{t} \tag{7}$$

先赋予$\boldsymbol{v}$一个随机的（或大致的估计值）初值，然后使用计算机反复对（7）式进行迭代，得到$\boldsymbol{v}$的序列，直到$\boldsymbol{v}$的各个分量收敛，则最终收敛到的极限值即为真实的状态价值$\boldsymbol{v}$。需要注意，当$\gamma\boldsymbol{P}$的谱半径$\rho>1$时，迭代是不收敛的。

另外，还可采用加入松弛因子、共轭梯度方法等对迭代进行改良和加速。