MDL Assignment 3.2

Roll No 2022113005

we are given that on hitting a boundary then the agent remains on the same cell

constants are ~~dot~~ given as follows

step cost = -0.04

P(going directn chosen) : ~~0.7~~ p

P(going in ⊥ directn) : ~~0.15~~ $\frac{(1-p)}{2}$

γ (Discount factor) : 0.95

given a state the next utility is calculated as :

0	-1	+1
0	0	0
0	$\boxed{0}$ w	0
0	0	0

$$U_{t+1}(I) = \max_A [C(I,A) + \gamma \sum_j P(I \mid I,A) \cdot U_t(I)]$$

(Bellman update eqn)

A : action that we choose in state A

C (I.A) = cost of taking action

P(J | I,A) = probab. of reaching a state J given that agent chooses action A in state I

For all cells and for all directions we calculate and consider maximum in one episode

<u>1st Iteration</u> :

1. U (0,0) we can go up. down, left, right

a) up : $0.04 + 0.95(0.7\times 0 + 0.15\times(-1) + 0.15(0))$

$= -0.1825$

b) Down → $-0.04 + 0.95(0.7\times 0 + 0.15\times(-1) + 0.15(0))$

$= -0.1825$

c) Right $= -0.04 + 0.95(0.7(-1) + 0.15\times0 + 0.15\times0)$

$= -0.705$

d) Left $= -0.04 + 0.95(0.7\times0 + 0.15\times0 + 0.15\times0)$

$= -0.04$

Among these maximum is $\Rightarrow U_1(0,0) = -0.04$

Left

2. $U_1(1,0) =$

a) Up $\rightarrow -0.04 + 0.95(0.7\times0 + 0.15\times0 + 0.15\times0)$

$= -0.04$

b) Down $\rightarrow -0.04 + 0.95(0.7\times0 + 0.15\times0 + 0.15\times0)$

$= -0.04$

c) Right $\rightarrow -0.04 + 0.95(0.7\times0 + 0.15\times0 + 0.15\times0)$

$= -0.04$ Left is same

$U_1(1,0) = \text{argmax}(-0.04, -0.04, -0.04, -0.04)$

$U_1(1,0) = -0.04$

3. $U_1(1,1)$

a) Up $= -0.04 + 0.95(0.7\times(-1) + 0.15(0)\times2)$

$= -0.705$

b) Down $= -0.04 + 0.95(0 + 0 + 0)$

$= -0.04$

c) Right $= -0.04 + 0.95(0.7\times0 + 0.15\times(-1) + 0.15\times0)$

$= -0.1825$

d) Left $=$ similarly, -0.1825

$\Rightarrow U_1(1,1) = -0.04$

$U_1(1,2)$ $\quad U \rightarrow (-0.04) + 0.95(0.7 \times 1 + 0.15 \times 0 + 0.15 \times 0)$

$= +0.625$

$D \rightarrow (-0.04) + 0.95(0.9(0) + 0 + 0) = -0.04$

$R \rightarrow (-0.04) + 0.95(0.7(0) + 0.15(1) + 0)$

$= +0.1025$

$L \rightarrow 0.1025$ similarly

$\Rightarrow U_1(1,2) = 0.625$

$U_1(2,0)$ $\quad U \rightarrow -0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0)$

$= -0.04.$

$D \rightarrow -0.04 + 0.95(0.7 \times 0 + 0.15 \times 0 + 0.15 \times 0) = -0.04$

$L \rightarrow -0.04$ similarly

$R \rightarrow -0.04$ similarly

$\Rightarrow U_1(2,0) = -0.04$

$U_1(2,2)$ $\quad U \rightarrow -0.04 + 0.95(0+0+0)$

$= -0.04$

$D \rightarrow -0.04$ similarly

$R \rightarrow -0.04$ similarly

$L \rightarrow -0.04$ similarly

$\Rightarrow U_1(2,2) = -0.04.$

7\. $U_1(3,0) =$ $\quad U \rightarrow -0.04 + 0.95(0+0+0) = -0.04$

$D \rightarrow -0.04$

$L \rightarrow -0.04$

$R \rightarrow -0.04$

similarly

$\Rightarrow U_1(3,0) = -0.04$

8\. $U_1(3,1) =$ $\quad U \rightarrow -0.04 + 0.95(0+0+0)$

$= -0.04$

$D \rightarrow -0.04$

$L \rightarrow -0.04$

$R \rightarrow -0.04$

$\Rightarrow U_1(3,1) = -0.04$

9. $U_1(3,2)$

$U \rightarrow -0.04 + 0.95(0.7 \times 0 + 0.15 \ldots + 0.15 \times 0)$

$= -0.04$

D, L, R: -0.04 similarly

$\Rightarrow U_1(3,2) = -0.04$.

Now, the grid has been updated as follows.

-0.04	-1	+1
-0.04	-0.04	0.625
-0.04	0	-0.04
-0.04	-0.04	-0.04

We see that wall and final states remain unchanged.

Iteratn #2

We will use the Bellman update eq^n again

$U_2(a,b) = \operatorname{argmax} [\text{step cost} + \gamma \sum_J P(J|(a,b),A) \cdot U_1(J)]$

1. $U_2(0,0) \rightarrow Up = -0.04 + 0.95((-0.04 \times 0) + (-1)(0.15) +$

$= -0.2148$

$Down = (-0.04) + 0.95((-0.04 \times 0.7) + -1 \times 0.15 + (-0.04)(0.15))$

$= -0.2148$.

$Right = (-0.04) + 0.95((-0.7 \times (-1)) + (-0.04 \times 0.15) + (-0.04 \times 0.15))$

$= -0.71604$

$Left = (-0.04) + 0.95(-0.7 \times (-0.04) + -0.04 \times 0.15 + -0.04 \times 0.15$

$= -0.078$.

$U_2(0,0)$: Max of all these $= -0.078$

$U_2(1,0)$.

$U \rightarrow -0.04 + 0.95(0.7 \times -0.04 + (-0.04 \times 0.15) + (-0.04 \times 0.15))$

$= -0.078$

$D \rightarrow -0.04 + 0.95[(0.7 \times -0.04) + (-0.04 \times 0.15) + (-0.04 \times 0.15)]$.

$= -0.078$

$R \rightarrow -0.04 \times 0.95[(0.7 \times 0.4) + (0.15 \times (-0.04)) + (0.15 \times (0.04))]$.

$= -0.078$

$L \rightarrow$ ~~similarly~~ $= -0.04 \times 0.95[(0.7 \times (-0.04) + 0.15 \times (-0.04) + 0.15 \times -0.04)]$

$= -0.078$

Considering max of all these, $U_2(1,0) = -0.078$.

3\. $U_2(1,1) =$

$U \rightarrow -0.04 + 0.95[(0.7 \times (-1)) + (0.15 \times -0.04) + (0.15 \times 0.625)]$.

$= -0.6216375$

$D \rightarrow -0.04 + 0.95[(0.625 \times 0.7) + (-1 \times 0.15) + (-0.04 \times 0.15)]$

$= 0.227425$

Right, similarly $= 0.227425$.

$L \rightarrow -0.04 + 0.95[(-0.04 \times 0.7) + (-1 \times 0.15) + (-0.04 \times 0.15)]$

$= -0.2148$.

$U_2(1,1) = \max\{-0.6216, 0.227425, -0.2148\}$

$= 0.227425$

4\. $U_2(1,2)$. $U \rightarrow -0.04 + 0.95[(0.7 \times 1) + (0.15 \times 0.625) + (0.15 \times -0.04)]$

$= 0.783625$

$D \rightarrow -0.04 + 0.95[(0.7 \times -0.04) + (0.625 \times 0.15) + 0.15$

$= 0.0167625$

$R \rightarrow -0.04 + 0.95[(0.7 \times 0.625) + (0.15 \times 1) + (0.15 \times -0$

$= 0.512425$

$L \rightarrow -0.04 + 0.95[(0.7 \times -0.04) + (0.15 \times 1) + (0.15 \times -0$

$= 0.0902$

$\Rightarrow U_2(1,2) = 0.7083625$

5. $U_2(2,0)$

$U \rightarrow -0.04 + 0.95(-0.04 \times 0.7 + 0 \times 0.15 \times 0.150)$

$= -0.04 + 0.95(-0.04)$

$= -0.078$

$D \rightarrow -0.04 + 0.95(-0.04) = -0.078$

Similarly $L, R = -0.078$

$\Rightarrow U_2(2,0) = -0.078$

6. $U_2(2,2)$

$U \rightarrow -0.04 + 0.95[(0.625 \times 0.7) + (0.04 \times 0.15) + (0.04 \times 0.15)]$

$= 0.364225$

$D \rightarrow -0.04 + 0.95[(-0.04 \times 0.7) + (0.93 \times -0.04) + (-0.04 \times$

$= -0.078$

$L \rightarrow -0.04 + 0.95[(0.7 \times -0.04) + (0.15 \times 0.625) + (0.15 \times 0.625)]$

$= 0.0167625$

Similarly, $R \rightarrow 0.0167625$

So $U_2(2,2) =$ Max over all directions

$= 0.364225$

$U_2(3,0)$

$U \rightarrow -0.04 + 0.95[(0.7 \times -0.04) + (0.15 \times -0.04) + (0.15 \times -0.04)]$

$= -0.078$

Similarly

$R \rightarrow -0.078$

same spot ↑

$L, D \rightarrow -0.04 + 0.95((-0.7 \times -0.04) + (0.04 \times 0.15) + (0.04 \times 0.15))$

$\Rightarrow U_2(3,0) = -0.078.$

8. $U_2(3,1)$, $U_2(3,2)$

$U, L, D, R = -0.078.$

$-0.04 + 0.95[(-0.7 \times -0.04) + (-0.04 \times 0.15) + (0.04 \times 0.15))$

(Because on bumping into walls, it returns to same place which has same value as adjacent cells $\Rightarrow U_2(3,1), U_2(3,2) = -0.078.$

<u>After iteratn 2</u>

-0.078	-1	+1
-0.078	0.2274	0.768
-0.078	0	0.364215
-0.078	-0.078	-0.078

As we can see, wall and reward/punishment final states remain unchanged. On comparison with the result generated in the second iteration with the computer is the same as what we calculated with the computer program