MDL Assignment 3

- Arghya Roy (2021115008)

Question 2

Initially, let the utility values of all the cells of the ~~gi~~ grid ~~except~~ be 0 except for the reward ~~state~~ cell (utility = 1), penalty cell (utility = -1) and the wall cell.

$U_{ij} \rightarrow$ utility of cell at row i column j

Iteration 1:

Row 1:

Cell 1

utility value of itself since out of bands

reward cell

- up: $0.7(0) + 0.15(0) + 0.15(1) = 0.15$

  direction of action

  perpendicular to direction of action

- down: $0.7(0) + 0.15(0) + 0.15(1) = 0.15$
- left: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- right: $0.7(1) + 0.15(0) + 0.15(0) = 0.7$

$\therefore$ max ~~Sprob~~ summation of weighted probabilities $= 0.7$

So,

$$U_{11} = \cancel{-0.04+} \text{ stepcost} + \gamma(0.7)$$
$$= -0.04 + 0.95 \times 0.7$$
$$= -0.04 + 0.665$$
$$= 0.625$$

Cell 2:

- ~~Utell~~ Reward state. $\therefore U_{12} = 1$

Cell 3

Penalty state. $\therefore U_{13} = -1$

Row 2 :

Cell 1

- up : $0.7(0) + 0.15(0) + 0.15(0) = 0$
- down : $0.7(0) + 0.15(0) + 0.15(0) = 0$
- left : $0.7(0) + 0.15(0) + 0.15(0) = 0$
- right : $0.7(0) + 0.15(0) + 0.15(0) = 0$

$\therefore$ max summation of weighted probabilities $= 0$

So,

$$U_{21} = \text{stepcost} + \gamma(0)$$
$$= -0.04$$

Cell 2

- up : $0.7(1) + 0.15(0) + 0.15(0) = 0.7$
- down : $0.7(0) + 0.15(0) + 0.15(0) = 0$

  value of itself since wall
- left : $0.7(0) + 0.15(0) + 0.15(1) = 0.15$
- right : $0.7(0) + 0.15(1) + 0.15(0) = 0.15$

$\therefore$ max summation of weighted probabilities $= 0.7$

So,

$$U_{22} = \text{stepcost} + \gamma(0.7)$$
$$= -0.04 + 0.95(0.7)$$
$$= -0.04 + 0.665$$
$$= 0.625$$

Cell 3

penalty

- up : $0.7(-1) + 0.15(0) + 0.15(0) = -0.7$
- down : $0.7(0) + 0.15(0) + 0.15(0) = 0$
- left : $0.7(0) + 0.15(0) + 0.15(-1) = -0.15$
- right : $0.7(0) + 0.15(-1) + 0.15(0) = -0.15$

$\therefore$ max summation of weighted probabilities $= 0$

So,

U_{23} = stepcost + $\gamma(0)$

= -0.04

Row 3:

Cell 1

- up: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- down: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- left: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- right: $0.7(0) + 0.15(0) + 0.15(0) = 0$

$\therefore$ max summation of weighted probabilities $= 0$

So,

U_{31} = stepcost + $\gamma(0)$

= -0.04

Cell 2

Wall

Cell 3

- up: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- down: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- left: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- right: $0.7(0) + 0.15(0) + 0.15(0) = 0$

$\therefore$ max summation of weighted probabilities $= 0$

So,

U_{33} = stepcost + $\gamma(0)$

= -0.04

Row 4:

Cell 1

- up: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- down: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- left: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- right $= 0.7(0) + 0.15(0) + 0.15(0) = 0$

$\therefore$ max summation of weighted probabilities $= 0$

So,

$U_{41} = \text{stepcost} + \gamma(0)$
$= -0.04$

Cell 2

- up: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- down: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- left: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- right: $0.7(0) + 0.15(0) + 0.15(0) = 0$

$\therefore$ max summation of ~~weighte~~ weighted probabilities $= 0$

So,

$U_{42} = \text{stepcost} + \gamma(0)$
$= -0.04$

Cell 3

- up: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- down: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- left: $0.7(0) + 0.15(0) + 0.15(0) = 0$
- right: $0.7(0) + 0.15(0) + 0.15(0) = 0$

$\therefore$ max summation of weighted probabilities $= 0$

So,

$U_{43} = \text{stepcost} + \gamma(0)$
$= -0.04$

So, after one iteration, the utilities of each cell are:

$$\begin{bmatrix} 0.625 & 1 & -1 \\ -0.04 & 0.625 & -0.04 \\ -0.04 & (\text{wall}) & -0.04 \\ -0.04 & -0.04 & -0.04 \end{bmatrix}$$

Iteration 2

Row 1:

Cell 1

- up: $0.7(0.625) + 0.15(0.625) + 0.15(\cancel{0.625}\ 1) = 0.68125$
- down: $0.7(-0.04) + 0.15(0.625) + 0.15(1) = 0.21575$
- left: $0.7(0.625) + 0.15(0.625) + 0.15(-0.04) = 0.52525$
- right: $0.7(1) + 0.15(0.625) + 0.15(-0.04) = 0.78775$

$\therefore$ max summation of weighted probabilities $= 0.78775$

So,

$$U_{11} = \text{stepcost} + \gamma(0.78775)$$
$$= -0.04 + 0.95(0.78775)$$
$$= 0.7083625$$

Cell 2: reward

Cell 3: penalty

Row 2:

Cell 1

- up: $0.7(0.625) + 0.15(-0.04) + 0.15(0.625) = 0.52525$
- down: $0.7(-0.04) + 0.15(-0.04) + 0.15(0.625) = 0.05975$
- left: $0.7(-0.04) + 0.15(\underset{-0.04}{\cancel{0}}) + 0.15(0.625) = 0.05975$
- right: $0.7(0.625) + 0.15(-0.04) + 0.15(0.625) = 0.52525$

$\therefore$ max summation of weighted probabilities $= 0.52525$

So,

$$U_{21} = \text{stepcost} + \gamma(0.52525)$$
$$= -0.04 + 0.95(0.52525)$$
$$= -0.04 + 0.4989875$$
$$= 0.4589875$$

Cell 2

- up: $0.7(1) + 0.15(-0.04) + 0.15(-0.04) = 0.688$
- down: $0.7(0.625) + 0.15(-0.04) + 0.15(-0.04) = 0.4255$
- left: $0.7(-0.04) + 0.15(0.625) + 0.15(1) = 0.21575$
- right: $0.7(-0.04) + 0.15(0.625) + 0.15(1) = 0.21575$

$\therefore$ max summation of weighted probabilities $= 0.688$

So,

$$U_{22} = \text{stepcost} + \gamma(0.688)$$
$$= -0.04 + 0.95(0.688)$$
$$= -0.04 + 0.6536$$
$$= 0.6136$$

Cell 3

- up: $0.7(-1) + 0.15(0.625) + 0.15(-0.04) = -0.61225$
- down: $0.7(-0.04) + 0.15(0.625) + 0.15(-0.04) = 0.05975$
- left: $0.7(0.625) + 0.15(-0.04) + 0.15(-1) = 0.2815$
- right: $0.7(-0.04) + 0.15(-0.04) + 0.15(-1) = -0.184$

$\therefore$ max summation of weighted probabilities $= 0.2815$

So,

$$U_{22} = \text{stepcost} + \gamma(0.2815)$$
$$= -0.04 + 0.95 \times 0.2815$$
$$= -0.04 + 0.267425$$
$$= 0.227425$$

Row 3:

Cell 1

- up: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- down: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- left: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- right: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$

$\therefore$ max summation of weighted probabilities $= -0.04$

So,

$$U_{31} = \text{stepcost} + \gamma(-0.04)$$
$$= -0.04 - 0.95(0.04)$$
$$= -0.04 - 0.038$$
$$= -0.078$$

Cell 2: wall

Cell 3:

- up: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- down: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- left: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- right: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$

$\therefore$ max summation of weighted probabilities $= -0.04$

So,

$$U_{33} = \text{stepcost} + \gamma(-0.04)$$
$$= -0.04 - 0.95(0.04)$$
$$= -0.04 - 0.038$$
$$= -0.078$$

Row 4:

Cell 1

- up: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- ~~left~~ down: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- left: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- right: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$

$\therefore$ max summation ~~of~~ ~~all~~ weighted probabilities $= -0.04$

So,

$$U_{41} = \text{stepcost} + \gamma(-0.04)$$
$$= -0.04 + 0.95(-0.04)$$
$$= -0.04 - 0.038$$
$$= -0.078$$

Cell 2

- up: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- down: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- left: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- right: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$

$\therefore$ max summation ~~of~~ weighted probabilities $= -0.04$

So,

$$U_{42} = \text{stepcost} + \gamma(-0.04)$$
$$= -0.04 + 0.95(-0.04)$$
$$= -0.04 - 0.038$$
$$= -0.078$$

Cell 3

- up: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- down: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- left: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$
- right: $0.7(-0.04)+0.15(-0.04)+0.15(-0.04) = -0.04$

$\therefore$ max summation of weighted probabilities $= -0.04$

So,

$$U_{43} = \text{Stepcost} + \gamma(-0.04)$$
$$= -0.04 + 0.95(-0.04)$$
$$= -0.04 - 0.038$$
$$= -0.078$$

So, after two iterations, the utilities of each cell are :

$$\begin{bmatrix} 0.7083625 & 1 & -1 \\ 0.4589875 & 0.6136 & 0.227425 \\ -0.078 & \text{(wall)} & -0.078 \\ -0.078 & -0.078 & -0.078 \end{bmatrix}$$

We thus see that these values exactly match the output from the code for the ~~second iteration~~ first and second iterations.