CSC_5RO11_TA

Reinforcement Learning

Élaboré par :
Shikun Wei

Encadré par :
Adriana TAPUS

3A robotique

Année universitaire : 2024/2025

1 Question1

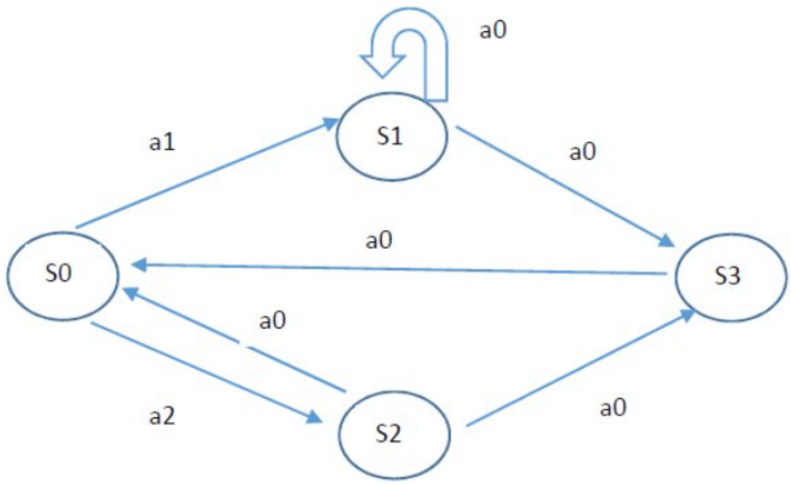

FIGURE 1 – Trasition of states

To find the total number of policies, we calculate the combinations based on the available actions in each state :

- State $S0$: Actions available are $a1$ and $a2$.
- State $S1$: Only action available is $a0$.
- State $S2$: Only action available is $a0$.
- State $S3$: Only action available is $a0$.

2 Question2

To write the equation for each optimal value function $V^*(s)$ for each state, we use the formula :

$$V^*(S) = R(S) + \max_a \gamma \sum_{S'} T(S, a, S')V^*(S')$$

1. For state $S0$:

$$V^*(S0) = R(S0) + \max\left(\gamma \sum_{S'} T(S0, a1, S')V^*(S'), \gamma \sum_{S'} T(S0, a2, S')V^*(S')\right)$$

Replace the transition probabilities for $a1$ and $a2$:

$$V^*(S0) = \max(\gamma \cdot V^*(S1), \gamma \cdot V^*(S2))$$

2. For state $S1$:

$$V^*(S1) = R(S1) + \max\left(\gamma \sum_{S'} T(S1, a0, S')V^*(S')\right)$$

Replace the transition probabilities for $a0$:

$$V^*(S1) = \gamma\left[(1-x)V^*(S1) + x \cdot V^*(S3)\right]$$

3. For state $S2$:

$$V^*(S2) = R(S2) + \max\left(\gamma \sum_{S'} T(S2, a0, S')V^*(S')\right)$$

Replace the transition probabilities for $a0$:

$$V^*(S2) = 1 + \gamma\left[(1-y)V^*(S0) + y \cdot V^*(S3)\right]$$

4. For state $S3$:

$$V^*(S3) = R(S3) + \max\left(\gamma \sum_{S'} T(S3, a0, S')V^*(S')\right)$$

Replace the transition probabilities for $a0$:

$$V^*(S3) = 10 + \gamma \cdot V^*(S0)$$

3 Question3

We need to determine if there exists a value x such that, for all $\gamma \in [0, 1)$ and $y \in [0, 1]$, the optimal policy $\pi^*(S_0)$ always chooses $a2$.
The optimal policy for a given state S is defined as :

$$\pi^*(S) = \arg\max_a \sum_{S'} T(S, a, S')V^*(S')$$

For $\pi^*(S_0) = a2$, we need :

$$\sum_{S'} T(S_0, a2, S')V^*(S') > \sum_{S'} T(S_0, a1, S')V^*(S')$$

we have :

$$V^*(S_1) = \gamma\left[(1-x)V^*(S_1) + x \cdot V^*(S_3)\right]$$

$$V^*(S_2) = 1 + \gamma\left[(1-y)V^*(S_0) + y \cdot (10 + \gamma \cdot V^*(S_0))\right]$$

By setting $x = 0$, we have :

$$V^*(S_1) = \gamma V^*(S_1)$$

Simplification with $x = 0$ Solving the equation :

$$V^*(S_1)(1-\gamma) = 0$$

This implies :

$$V^*(S_1) = 0 \quad \text{for } \gamma \neq 1$$

To ensure $\pi^*(S_0) = a2$, we need :

$$V^*(S_2) > V^*(S_1)$$

Substitute the results :

$$1 + \gamma \left[(1-y)V^*(S_0) + y \cdot (10 + \gamma \cdot V^*(S_0))\right] > 0$$

Since $V^*(S_1) = 0$, the condition simplifies to checking if $V^*(S_2) > 0$. Given the positive terms in $V^*(S_2)$, this condition holds as long as $\gamma \in [0, 1)$ and $y \in [0, 1]$. So, the answer is yes, $x = 0$ satisfies the condition for $\pi^*(S_0)$ to choose $a2$.

4 Question4

Expression for $V^*(S_1)$:

$$V^*(S_1) = \gamma \left[(1-x)V^*(S_1) + x \cdot V^*(S_3)\right]$$

So we have :

$$V^*(S_1)(1 - \gamma + \gamma x) = \gamma x \cdot V^*(S_3)$$

$$V^*(S_1) = \frac{\gamma x \cdot V^*(S_3)}{1 - \gamma + \gamma x}$$

Expression for $V^*(S_2)$:

$$V^*(S_2) = 1 + \gamma \left[(1-y)V^*(S_0) + y \cdot (10 + \gamma \cdot V^*(S_0))\right]$$

To prove whether $V^*(S_1) > V^*(S_2)$ for all $x > 0$ and $\gamma \in (0, 1)$, we need :

$$\frac{\gamma x \cdot V^*(S_3)}{1 - \gamma + \gamma x} > 1 + \gamma \left[(1-y)V^*(S_0) + y \cdot (10 + \gamma \cdot V^*(S_0))\right]$$

As $x \to 0$:

$$V^*(S_1) \to 0$$

$$V^*(S_2) = 1 + \gamma \left[(1-y)V^*(S_0) + y \cdot (10 + \gamma \cdot V^*(S_0))\right]$$

When x is very small, $V^*(S_1)$ approaches 0, while $V^*(S_2)$ includes a constant term 1, ensuring $V^*(S_2) > V^*(S_1)$.

From this analysis, we see that there is no fixed value of y such that for all $x > 0$ and $\gamma \in (0, 1)$, $V^*(S_1) > V^*(S_2)$. Therefore, we cannot prove the existence of a single y such that $\pi^*(S_0)$ always chooses $a1$ for all $x > 0$ and $\gamma \in (0, 1)$.

5 Question5

By running the implementation of python code, we have obtained the Optimal Policy π^* and Value Function V^* :

— **Optimal actions** π^* :

$$\begin{cases} \Pi^*(S_0) = a_1 \\ \Pi^*(S_1) = a_0 \\ \Pi^*(S_2) = a_0 \\ \Pi^*(S_3) = a_0 \end{cases}$$

— **Optimal value function** V^* :

$$V^* = \begin{bmatrix} 14.1852 \\ 15.7614 \\ 15.6975 \\ 22.7666 \end{bmatrix}$$