Assignment 5

Pranav Kasela 846965

Introduction to the problem

The objective of the assignment is to maximize the reward in following condition:

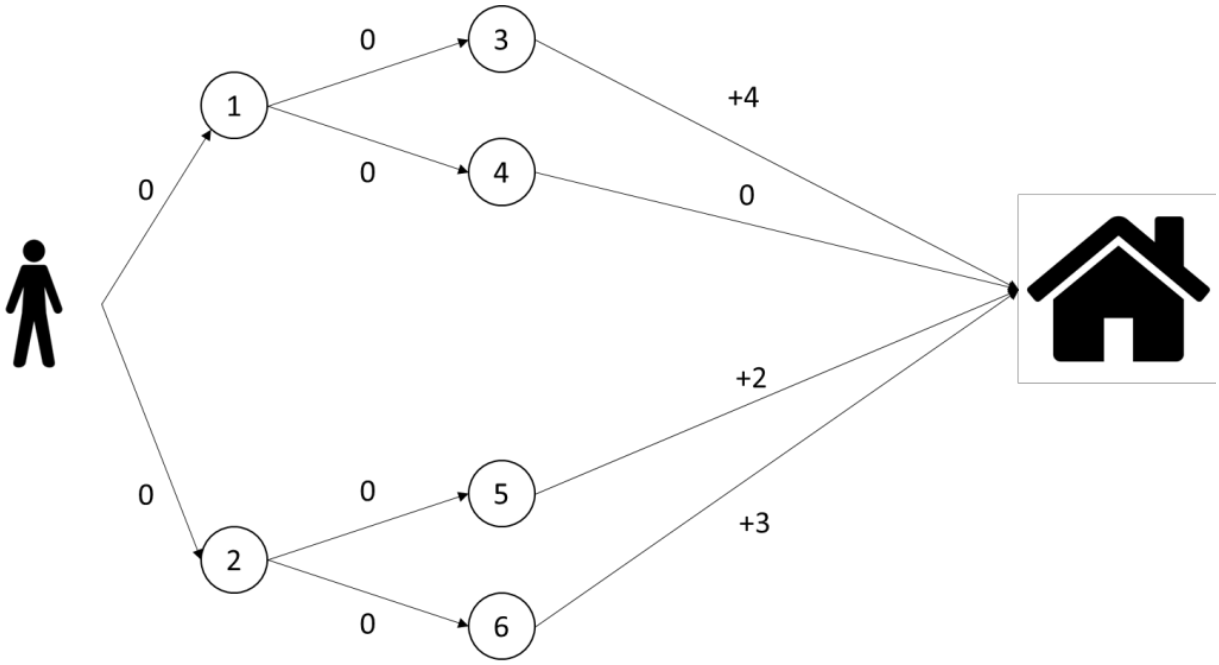

Figure 1: Representation of the problem

The initial polict π_0 is:

- $0 \to 2$
- $2 \to 5$
- $1 \to 4$

In this particular problem, the node home can be neglected and the nodes $3, 4, 5, 6$ can be considered as final nodes with the reward given by the node connecting these nodes to the home node. This way in the state function these is no need to write the the value of these nodes which remains costant.

The state function for the initial policy is:

$$V^{\pi_0}(0) = max\{0 + \gamma \cdot 0, 0 + \gamma \cdot 2\} = 2$$
$$V^{\pi_0}(1) = max\{0 + \gamma \cdot 4, 0 + \gamma \cdot 0\} = 4$$
$$V^{\pi_0}(2) = max\{0 + \gamma \cdot 2, 0 + \gamma \cdot 3\} = 3$$

where the first argument indicates the path to the left and the second argument indicates the path to the right. $\gamma = 1$ has been decided.

The new policy π_1 now is:

- $0 \rightarrow 2$
- $2 \rightarrow 6$
- $1 \rightarrow 3$

The state function in π_1 is:

$$V^{\pi_1}(0) = max\{0 + \gamma \cdot 4, 0 + \gamma \cdot 3\} = 4$$
$$V^{\pi_1}(1) = max\{0 + \gamma \cdot 4, 0 + \gamma \cdot 0\} = 4$$
$$V^{\pi_1}(2) = max\{0 + \gamma \cdot 2, 0 + \gamma \cdot 3\} = 3$$

The new policy derived from the state function is π_2:

- $0 \rightarrow 1$
- $2 \rightarrow 6$
- $1 \rightarrow 3$

Actually this is the best path, but to confirm the convergence the next step is needed.

It can be verified immediatly that the state function does not change and that $\pi_3 = \pi_2$, which is the condition for the convergence.