Markov Decision Process Report

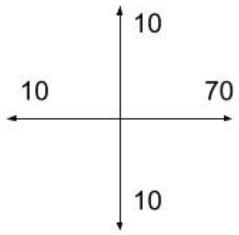
Dr. Carlos A. Lara-Álvarez

May 22, 2019

Problem #2
Collaborators: Andres Mitre.

(a) Given the next value function for a MDP (equation 1): consider the figure 1 and calculate the value function for every reward state, as well as the best policy. For s friendly understanding, suppose we start at S_0 and the goal is to get to the green box (figure 1) so we get a big prize. The red box takes money out of your pocket. Note: in Reinforcement Learning environments we start from any state not at S_0

$$V_{i+1}(s) = \max_a \left\{ \sum_{s'} P_a\left(S, S'\right)\left(R_a\left(S, S'\right) + \gamma V_i\left(S'\right)\right) \right\} \tag{1}$$



S0	S1	S2	S3
S4		S5	W:+3
S6		S7	L:-1
S8	S9	S10	S11

Figure 1: MDP exercise

Solution: In order to get the function value for the MDP. We proceed to define the transitions matrix for every state S_i to the next state S_{i+n}. for every action a_i taken (up, right, down, left).

The probabilities and rewards matrices according to the actions are defined as:

$$P_{up} = \begin{bmatrix}
S0 & S1 & S2 & S3 & S4 & S5 & S6 & S7 & S8 & S9 & S10 & S11 & W & L & \\
0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S0 \\
0.5 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S1 \\
0 & 0.33 & 0 & 0.33 & 0 & 0.33 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S2 \\
0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & S3 \\
0.7 & 0 & 0 & 0 & 0 & 0 & 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S4 \\
0 & 0 & 0.15 & 0 & 0 & 0.7 & 0 & 0.15 & 0 & 0 & 0 & 0 & 0.15 & 0 & S5 \\
0 & 0 & 0 & 0 & 0.7 & 0 & 0 & 0 & 0.3 & 0 & 0 & 0 & 0 & 0 & S6 \\
0 & 0 & 0 & 0 & 0 & 0.7 & 0 & 0 & 0 & 0 & 0.15 & 0 & 0 & 0.15 & S7 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.7 & 0 & 0 & 0.3 & 0 & 0 & 0 & 0 & S8 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 & 0 & 0 & S9 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7 & 0 & 0.15 & 0 & 0.15 & 0 & 0 & S10 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.3 & 0 & 0 & 0.7 & S11 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & W \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & L
\end{bmatrix}$$

Figure 2: Probability matrix for action taken up

$$P_{right} = \begin{bmatrix}
S0 & S1 & S2 & S3 & S4 & S5 & S6 & S7 & S8 & S9 & S10 & S11 & W & L & \\
0 & 0.7 & 0 & 0 & 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S0 \\
0.3 & 0 & 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S1 \\
0 & 0.15 & 0 & 0.7 & 0 & 0.15 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S2 \\
0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & S3 \\
0.5 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S4 \\
0 & 0 & 0.15 & 0 & 0 & 0 & 0 & 0.15 & 0 & 0 & 0 & 0 & 0.7 & 0 & S5 \\
0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & S6 \\
0 & 0 & 0 & 0 & 0 & 0.15 & 0 & 0 & 0 & 0 & 0.15 & 0 & 0 & 0.7 & S7 \\
0 & 0 & 0 & 0 & 0 & 0 & 0.3 & 0 & 0 & 0.7 & 0 & 0 & 0 & 0 & S8 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.3 & 0 & 0.7 & 0 & 0 & 0 & S9 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.3 & 0 & 0 & 0 & 0.7 & 0 & 0 & S10 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 & S11 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & W \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & L
\end{bmatrix}$$

Figure 3: Probability matrix for action taken right

$$P_{down} = \begin{bmatrix} S0 & S1 & S2 & S3 & S4 & S5 & S6 & S7 & S8 & S9 & S10 & S11 & W & L & \\ 0 & 0.3 & 0 & 0 & 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S0 \\ 0.5 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S1 \\ 0 & 0.15 & 0 & 0.15 & 0 & 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S2 \\ 0 & 0 & 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7 & 0 & S3 \\ 0.3 & 0 & 0 & 0 & 0 & 0 & 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S4 \\ 0 & 0 & 0.15 & 0 & 0 & 0 & 0 & 0.7 & 0 & 0 & 0 & 0 & 0.15 & 0 & S5 \\ 0 & 0 & 0 & 0 & 0.3 & 0 & 0 & 0 & 0.7 & 0 & 0 & 0 & 0 & 0 & S6 \\ 0 & 0 & 0 & 0 & 0 & 0.15 & 0 & 0 & 0 & 0 & 0.7 & 0 & 0 & 0.15 & S7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & S8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0.5 & 0 & 0 & 0 & S9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 & 0 & S10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 & S11 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & W \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & L \end{bmatrix}$$

Figure 4: Probability matrix for action taken down

$$P_{left} = \begin{bmatrix} S0 & S1 & S2 & S3 & S4 & S5 & S6 & S7 & S8 & S9 & S10 & S11 & W & L & \\ 0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S0 \\ 0.7 & 0 & 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S1 \\ 0 & 0.7 & 0 & 0.15 & 0 & 0.15 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S2 \\ 0 & 0 & 0.7 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.3 & 0 & S3 \\ 0.5 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S4 \\ 0 & 0 & 0.25 & 0 & 0 & 0.25 & 0 & 0.25 & 0 & 0 & 0 & 0 & 0.25 & 0 & S5 \\ 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & 0 & S6 \\ 0 & 0 & 0 & 0 & 0 & 0.25 & 0 & 0.25 & 0 & 0 & 0.25 & 0 & 0 & 0.25 & S7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.5 & 0 & 0 & 0.5 & 0 & 0 & 0 & 0 & S8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7 & 0 & 0.3 & 0 & 0 & 0 & S9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.15 & 0 & 0.7 & 0 & 0.15 & 0 & 0 & S10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.7 & 0 & 0 & 0.3 & S11 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & W \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & L \end{bmatrix}$$

Figure 5: Probability matrix for action taken left

$$R_{up} = \begin{bmatrix} S0 & S1 & S2 & S3 & S4 & S5 & S6 & S7 & S8 & S9 & S10 & S11 & W & L & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & S11 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & W \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & L \end{bmatrix}$$

Figure 6: Reward matrix for action taken up

$$R_{right} = \begin{bmatrix} S0 & S1 & S2 & S3 & S4 & S5 & S6 & S7 & S8 & S9 & S10 & S11 & W & L & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & S5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & S7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S11 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & W \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & L \end{bmatrix}$$

Figure 7: Reward matrix for action taken right

$$R_{down} = \begin{bmatrix} S0 & S1 & S2 & S3 & S4 & S5 & S6 & S7 & S8 & S9 & S10 & S11 & W & L & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & S3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S11 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & W \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & L \end{bmatrix}$$

Figure 8: Reward matrix for action taken down

$$R_{left} = \begin{bmatrix} S0 & S1 & S2 & S3 & S4 & S5 & S6 & S7 & S8 & S9 & S10 & S11 & W & L & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S3 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S5 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S6 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S7 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S8 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S9 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S10 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & S11 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & W \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & L \end{bmatrix}$$

Figure 9: Reward matrix for action taken left

Value functions: From the above matrices, we can focus only in the probability where we have a reward value (Reward matrices). Starting from the S_0 to S_n, we find that the first reward value is located from the probability (P=0.5) of state S_3 to state W is the action R_{down} with a reward value of 3. As stated in the value function formula(1), we get the following value function:

$$a_{down} = (0.5)\,(3 + 0.1(0)) = 2.4 \tag{2}$$

$$V_{S_3} = \max_a \left\{ \sum_{s'} P_a\,(S, S')\,(R_a\,(S, S') + \gamma V_i\,(S')) \right\} = 1.5 \tag{3}$$

The second reward value is located from the probability (P=0.7) of state S_5 to state W is the action R_{right} with a reward value of 3:

$$a_{right} = (0.7)\,(3 + 0.1(0)) = 2.1 \tag{4}$$

$$V_{S_3} = \max_a \left\{ \sum_{s'} P_a\,(S, S')\,(R_a\,(S, S') + \gamma V_i\,(S')) \right\} = 2.1 \tag{5}$$

For the next value functions we get a negative value, according to the MDP, we avoid negative rewards so we do not take in mind. The third reward value is located from the probability (P=0.7) of state S_7 to state L is the action R_{right} with a reward value of -1:

$$a_{right} = (0.7)\,(-1 + 0.1(0)) = -0.7 \tag{6}$$

Since we only take the maximum value for the function, then:

$$V_{S_7} = \max_a \left\{ \sum_{s'} P_a\,(S, S')\,(R_a\,(S, S') + \gamma V_i\,(S')) \right\} = 0 \tag{7}$$

The fourth reward value is located from the probability (P=0.7) of state S_{11} to state L is the action R_{up} with a reward value of -1:

$$a_{up} = (0.7)\,(-1 + 0.1(0)) = -0.7 \tag{8}$$

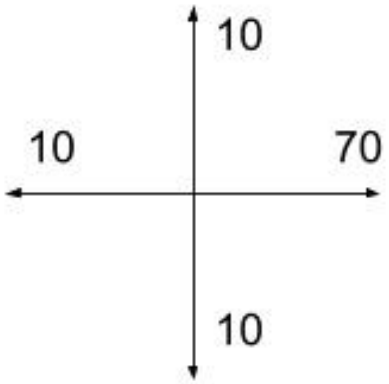
then:

$$V_{S_{11}} = \max_a \left\{ \sum_{s'} P_a\,(S, S')\,(R_a\,(S, S') + \gamma V_i\,(S')) \right\} = 0 \tag{9}$$

Policy: The policy from the proposed problem is visualized in the next sequence:

$$policy_{(S0:S9)} = \begin{bmatrix} S0 & S1 & S2 & S3 & S4 & S5 & S6 & S7 & S8 & S9 \\ 1 & 1 & 1 & 2 & 0 & 1 & 0 & 2 & 0 & 3 \\ right & right & right & down & up & right & up & down & up & left \end{bmatrix} \quad (10)$$

$$policy_{(S10:S11)} = \begin{bmatrix} S10 & S11 \\ 3 & 3 \\ left & left \end{bmatrix} \quad (11)$$



S0	S1	S2	S3
S4		S5	W:+3
S6		S7	L:-1
S8	S9	S10	S11

Figure 10: Environment