Assignment 4

DDL: 2023.12.31 23:55

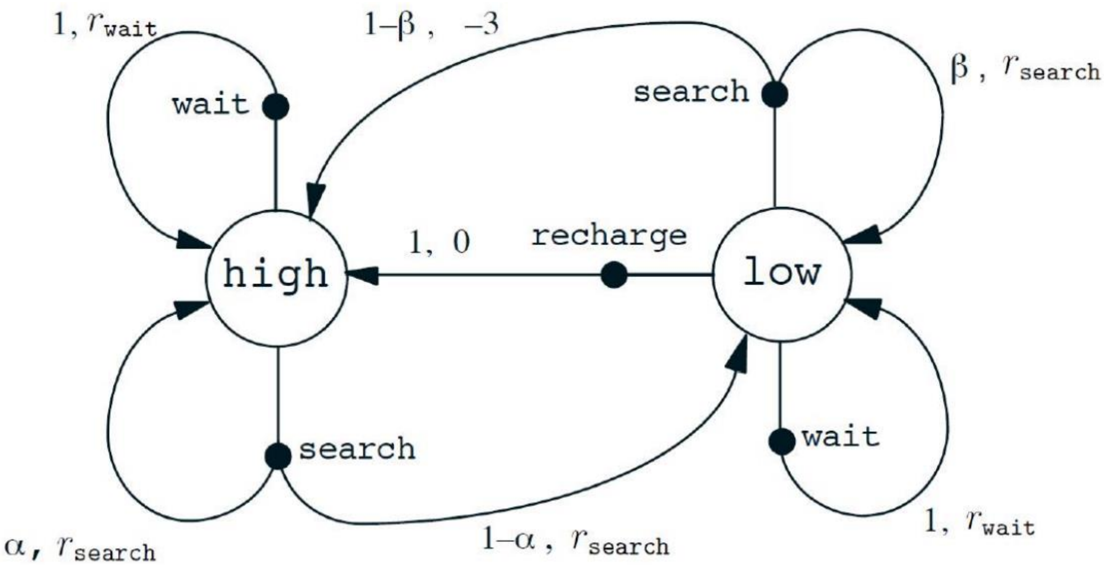
1. Given an arbitrary MDP with a reward function $R(s)$ and two given constants $\alpha > 0$ and β, consider a modified MDP where everything remains the same, except it has a new reward function $R'(s) = \alpha R(s) + \beta$. Prove that the modified MDP has the same optimal policy as the original MDP. (30 marks)

2. A mobile robot has the job of collecting empty cans in an environment. We only consider the high-level decisions about how to search cans by a RL agent based on the charge level of the battery. At each time step, the robot decides whether it should
   1) actively search a can for a certain period of time,
   2) remain stationary and wait for someone to bring it a can, or
   3) head back to its home base to recharge its battery.

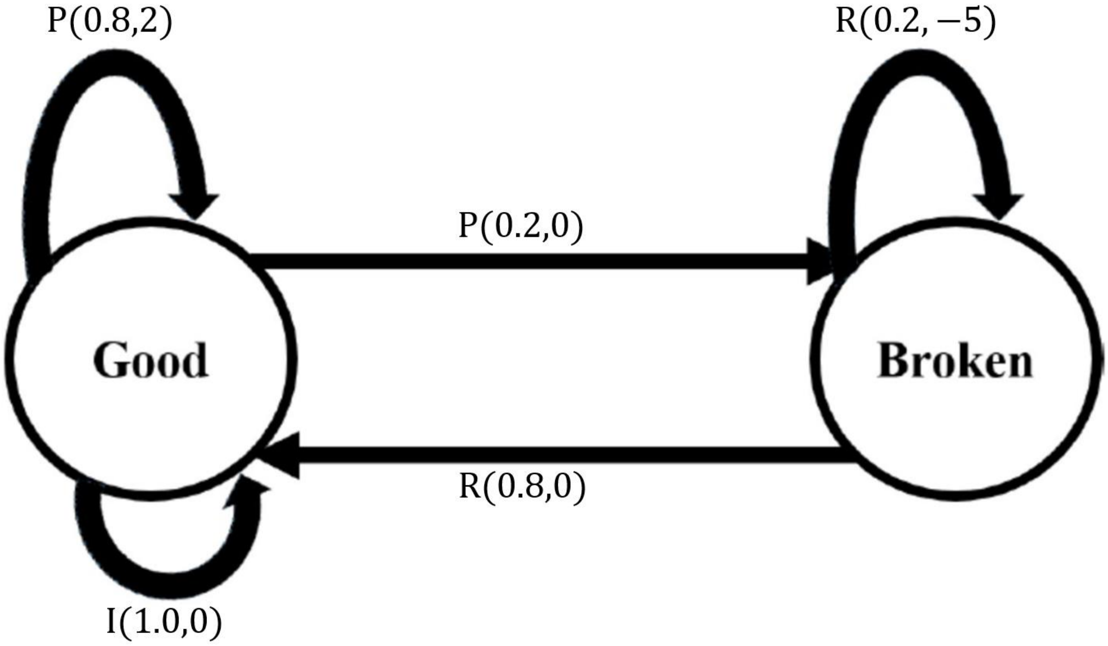
   Consider the MDP in the following figure, with $\alpha = 0.5, \beta = 0.3, \gamma = 0.8, r_{search} = 3, r_{wait} = 0$ and the following policy:

Probability \ Action / State	wait	search	recharge
low	0	0.4	0.6
high	0	1	0

   (1) Compute the state-value function where the MDP stops after two steps. (15 marks)
   (2) Compute the action-value function for each action value pair in the case the MDP stops after a single step. (15 marks)



3. A machine has two states: ‘Good’(G) and ‘Broken’(B). In the ‘Good’ state, there are two possible actions: ‘produce’ (P) and ‘inactive’ (I). Taking the ‘produce’ action in the ‘Good’ state has probability 0.8 to remain in the ‘Good’ state with the immediate reward 2 and probability 0.2 to reach the `Broken' state with the immediate reward 0. Taking the ‘inactive’ action in the ‘Good’ state will remain in the ‘Good’ state with probability 1 and the immediate reward 0. In the ‘Broken’ state, there is only one action: ‘Repair’ (R), which leads to the ‘Good’ state with probability 0.8 and the immediate reward 0 and otherwise remains in the ‘Broken’ state with the immediate reward -5. Such state transitions are shown in the following figure.



(1) For a policy that always takes the action ‘produce’ in the ‘Good’ state, determine the value function of the two states in terms of the discounted factor γ. (20 marks)

(2) Denote the optimal value function in the ‘Good’ state by $V^*(G)$ and that in the ‘Broken’ state by $V^*(B)$. In order to determine the optimal policy, specify the relations between $V^*(G)$ and $V^*(B)$ in two equations with the discounted factor γ. (20 marks)