SAiDL-2019-Summer Assignment-2019(Markov Decision Process)

Ayush Aaryan

August 11, 2019

1 Question

You receive the following letter:
Dear Friend, Some time ago, I bought this old house, but found it to be haunted by ghostly sardonic laughter. As a result it is hardly habitable. There is hope, however, for by actual testing I have found that this haunting is subject to certain laws, obscure but infallible, and that the laughter can be affected by my playing the organ or burning incense. In each minute, the laughter occurs or not, it shows no degree. What it will do during the ensuing minute depends, in the following exact way, on what has been happening during the preceding minute: Whenever there is laughter, it will continue in the succeeding minute unless I play the organ, in which case it will stop. But continuing to play the organ does not keep the house quiet. I notice, however, that whenever I burn incense when the house is quiet and do not play the organ it remains quiet for the next minute. At this minute of writing, the laughter is going on. Please tell me what manipulations of incense and organ I should make to get that house quiet, and to keep it so.
Sincerely,
At Wits End

1.1 **Formulate this problem as an MDP (for the sake of uniformity , formulate it as a continuing discounted problem with $\gamma = 0.9$. Let the reward be +1 on any transition into the silent state, and -1 on any transition into the laughing state). Explicitly give the state set, action sets, state transition and reward function.**

1.2 **Starting with the policy $\pi(laughing) = \pi(silent) = \pi(incense, noorgan)$, perform a couple of policy iterations(by hand) until you find an optimal policy(Clearly show and label each step. If you are taking a lot of iterations, stop and reconsider your formulation). Do a couple of value iterations as well.**

1.3 **What are the resulting optimal state-action values for all state-action pairs?**

1.4 **What is your advice to At Wits End?**

2 Answers

2.1 a.State Set={L,S},where L refers to Laughter and S refers to Silent.

b.Action Set=$\{\neg I \wedge O, \neg I \wedge \neg O, I \wedge O, I \wedge \neg O\}$,where I refers to burning incense sticks and O refers to playing organ.

c.State Transition=$\{P_{L\to L}$,$P_{L\to S}$,$P_{S\to L}$, $P_{S\to S}\}$

d.R_t=$r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + ..$,where γ is the discount factor equal to 0.9, taking into account the possible errors.

For the MDP above,

Reward={+1 for Silent $\to$ Silent and Laughter $\to$ Silent}

{-1 for Silent $\to$ Laughter and Laughter $\to$ Laughter}

2.2 Finding an Optimal Policy

2.2.1 Policy Iteration

Step1:Arbritary Assignment of Policy: $\pi(L) = \pi(S) = I \wedge \neg O$

Step2: $\gamma = 0.9$

$$V(L) = P^{\pi(L)}_{L\to L}(R^{\pi(L)}_{L\to L} + \gamma V(L)) + P^{\pi(L)}_{L\to S}(R^{\pi(L)}_{L\to S} + \gamma V(S)) = -1$$

$$V(S) = P^{\pi(L)}_{S\to L}(R^{\pi(S)}_{S\to L} + \gamma V(L)) + P^{\pi(S)}_{S\to S}(R^{\pi(S)}_{S\to S} + \gamma V(S)) = 1$$

Step3: $\gamma = 0.9$

$$V(L) = P^{\pi(L)}_{L\to L}(R^{\pi(L)}_{L\to L} + \gamma V(L)) + P^{\pi(L)}_{L\to S}(R^{\pi(L)}_{L\to S} + \gamma V(S)) = -1.9$$

$$V(S) = P^{\pi(L)}_{S\to L}(R^{\pi(S)}_{S\to L} + \gamma V(L)) + P^{\pi(S)}_{S\to S}(R^{\pi(S)}_{S\to S} + \gamma V(S)) = 1.9$$

Step4: $\gamma = 0.9$

$$V(L) = P^{\pi(L)}_{L\to L}(R^{\pi(L)}_{L\to L} + \gamma V(L)) + P^{\pi(L)}_{L\to S}(R^{\pi(L)}_{L\to S} + \gamma V(S)) = -2.71$$

$$V(S) = P^{\pi(L)}_{S\to L}(R^{\pi(S)}_{S\to L} + \gamma V(L)) + P^{\pi(S)}_{S\to S}(R^{\pi(S)}_{S\to S} + \gamma V(S)) = +2.71$$

Hence,we see that V(S) is improving whereas V(L) is decreasing.So,let's try a new policy-:

$$\pi(L) = O \wedge I$$

$$\pi(S) = \neg O \wedge I$$

Step5: $\gamma = 0.9$

$V(L) = P^{\pi(L)}_{L\to L}(R^{\pi(L)}_{L\to L} + \gamma V(L)) + P^{\pi(L)}_{L\to S}(R^{\pi(L)}_{L\to S} + \gamma V(S)) = +3.44$

$V(S) = P^{\pi(L)}_{S\to L}(R^{\pi(S)}_{S\to L} + \gamma V(L)) + P^{\pi(S)}_{S\to S}(R^{\pi(S)}_{S\to S} + \gamma V(S)) = +3.44$

This looks like a good policy but lets perform another iteration to be sure that this is the optimal policy.

Step5: $\gamma = 0.9$

$V(L) = P^{\pi(L)}_{L\to L}(R^{\pi(L)}_{L\to L} + \gamma V(L)) + P^{\pi(L)}_{L\to S}(R^{\pi(L)}_{L\to S} + \gamma V(S)) = +4.10$

$V(S) = P^{\pi(L)}_{S\to L}(R^{\pi(S)}_{S\to L} + \gamma V(L)) + P^{\pi(S)}_{S\to S}(R^{\pi(S)}_{S\to S} + \gamma V(S)) = +4.10$

Hence,there is still improvement.Hence,this is the optimal policy.

2.2.2 Value Iteration

In value iteration,we first set the values to 0 and then find the values for different actions, and choose the one which gives the maximum q reward.

Step1: V(L)=V(S)=0

Step2:

$V(L) = P^{action}_{L\to L}(R^{action}_{L\to L} + \gamma V(L)) + P^{action}_{L\to S}(R^{action}_{L\to S} + \gamma V(S))$

We find that V(L) is maximum for action=$\{O \wedge I, O \wedge \neg I\}$

$V(S) = P^{action}_{S\to L}(R^{action}_{S\to L} + \gamma V(L)) + P^{action}_{S\to S}(R^{action}_{S\to S} + \gamma V(S))$

We find that V(L) is maximum for action=$\{\neg O \wedge I\}$

Step3: Iterating again following the same set of actions.

V(L)=+1.9

V(S)=+1.9

Step4: Iterating again following the same set of actions.

V(L)=+2.71

V(S)=+2.71

So,we see that the improvement over each iteration is decreasing/converging to an optimum to which we set our values,this convergence follows a GP with common ratio γ=0.9.So,now the final value can be found out after discounting rewards.

2.3 State-Action Table

Whenever we take the optimal action first and continue taking optimal actions,we get the full reward(+10) and when we take a random action and then take optimal actions we get 0.8 of the reward.$(-1(for the first action) + \gamma * total - reward)$.

$L \Longrightarrow LaughterState$
$S \Longrightarrow SilentState$
$O \Longrightarrow PlayingOrgan$
$I \Longrightarrow BurningIncenseStick$

CurrentState	Action	NewState	Optimal State-Action values
S	$O \wedge I$	L	+8
S	$O \wedge \neg I$	L	+8
S	$\neg O \wedge I$	S	+10
S	$\neg O \wedge \neg I$	L	+8
L	$O \wedge I$	S	+10
L	$O \wedge \neg I$	S	+10
L	$\neg O \wedge \neg I$	L	+8
L	$\neg O \wedge I$	L	+8

2.4 My advice to "At Wits End" would be if the room is silent, do not play the organ and burn incense and if there is laughter,play the organ