CAP6635 – Advanced AI

3/29/2024

Aaron Goldstein

Assignment 3

```
PS C:\Users\Aaron\OneDrive\Desktop\School\Advanced AI Class\CAP6635 - Assignment 3> python main.py
MDP #1
Default R Value Per State: -1
Varying R Value in Top Left Corner: -100
Terminal State Reward: 10
Reward Grid for Traveling to Each Cell
-100 -1 10
-1 -1 -1
-1 -1 -1
Environment Before Value Iteration
Utility of Each Cell
0 0 0
0 0 0
0 0 0
Initial Random Policy
↓ ↓ *
↓ ↑ ↓
→ ← ←
Environment After Value Iteration
Max Utility of Each Cell Given R
6.505720792612547 19.41769062442273 10
13.968127529555353 17.932720194668182 19.417690624422725
14.911550298646075 16.480384166397698 17.769554703676494
Optimal Policy Given R
→ → *
↓ → ↑
→ → ↑
----------------------------------------
```

1. For r of -100, we can see the policy definitely trying to move away from the top upper left-hand corner while simultaneously moving toward the terminal state of 10 in the top right hand corner. The policy balances safety with speed, as can be seen in the policy moving down in row 2, column 1 to avoid any chance of accidentally falling in the top upper left.

```
----------------------------------------
MDP #2
Default R Value Per State: -1
Varying R Value in Top Left Corner: -3
Terminal State Reward: 10
Reward Grid for Traveling to Each Cell
-3 -1 10
-1 -1 -1
-1 -1 -1
Environment Before Value Iteration
Utility of Each Cell
0 0 0
0 0 0
0 0 0
Initial Random Policy
↓ ↓ *
↓ ↑ ←
← ↓ →
Environment After Value Iteration
Max Utility of Each Cell Given R
17.520969932292555 19.417690334436468 10
16.238223933593297 17.93271905030548 19.41769033443647
15.160978083722462 16.480381117494666 17.769553342833216
Optimal Policy Given R
→ → *
→ → ↑
→ → ↑
```

2. For r of -3, we can see the policy is like that where r is -100, however in this case it's a bit more reckless. Where before, the policy said when at the state of row 2, column 1, go down to avoid the top upper left at all costs, it now takes more risk of falling into the top upper left corner for a chance to reach the middle more quickly.

```
---------------------------------------
MDP #3
Default R Value Per State: -1
Varying R Value in Top Left Corner: 0
Terminal State Reward: 10
Reward Grid for Traveling to Each Cell
0 -1 10
-1 -1 -1
-1 -1 -1
Environment Before Value Iteration
Utility of Each Cell
0 0 0
0 0 0
0 0 0
Initial Random Policy
↓ ← *
↓ ↓ ↓
← ← ←
Environment After Value Iteration
Max Utility of Each Cell Given R
18.01514538128446 19.431162703267315 10
17.597636913236084 18.055331243564606 19.431162703267315
16.189964358151563 16.66499274233772 17.80168049122213
Optimal Policy Given R
→ → *
↑ ↑ ↑
↑ ↑ ↑
---------------------------------------
```

3. For r of 0, we can that the policy aims to cross the top left corner if already in the first column because this allows it to approach the terminal state while avoiding a negative one penalty in one of the other transition states.

```
---------------------------------------
MDP #4
Default R Value Per State: -1
Varying R Value in Top Left Corner: 3
Terminal State Reward: 10
Reward Grid for Traveling to Each Cell
3 -1 10
-1 -1 -1
-1 -1 -1
Environment Before Value Iteration
Utility of Each Cell
0 0 0
0 0 0
0 0 0
Initial Random Policy
→ ← *
← → ←
← → ↓
Environment After Value Iteration
Max Utility of Each Cell Given R
250.87020535013193 249.94810399021742 10
249.94810399021742 245.5964542398843 233.97838634207278
245.1183040453416 241.3404393582597 236.74303918728054
Optimal Policy Given R
↑ ← *
↑ ← ↓
↑ ← ←
---------------------------------------
```

4. For r of 3, the top left corner seems to confuse the agent and cause it to head for the top left corner. Combined with the high gamma, it causes rapid growth in the utilities of the transition states while the terminal state's utility remains fixed to its reward state. Lowering the gamma seemed to lead to slightly better policy results for this r value. This can be seen in the image below when MDP #4 was run with a Gamma of 0.50, this may possibly be because the lower gamma caused the utilities to grow slower.

```
---------------------------------------
MDP #4
Default R Value Per State: -1
Varying R Value in Top Left Corner: 3
Terminal State Reward: 10
Reward Grid for Traveling to Each Cell
3 -1 10
-1 -1 -1
-1 -1 -1
Environment Before Value Iteration
Utility of Each Cell
0 0 0
0 0 0
0 0 0
Initial Random Policy
→ ↑ *
↑ ↑ ↑
→ ← →
Environment After Value Iteration
Max Utility of Each Cell Given R
5.87955606766715 12.682072547445333 10
5.05204658902264 4.959490245658074 12.682072547445333
1.1403964454154063 1.258404380911031 4.353403556198187
Optimal Policy Given R
↑ → *
↑ ↑ ↑
↑ ↑ ↑
---------------------------------------
```