AI Assignment 2 Report

Team 48

Grid world:-

4.800 0.000 0.000 48.000
0.000 0.000 0.000 0.000
0.000 0.000 0.000 0.000
0.000 0.000 -9.600 0.000

Part B.)
1 a.) Step cost = -4.8, discount factor = 0.1.

Final utilities:-
4.800 0.000 -1.020 48.000
-4.514 -5.267 -4.941 -1.020
-5.267 -5.327 0.000 -4.981
-5.327 -5.333 -9.600 -5.326

Policy:-
T - E T
N W N N
N N - N
N N T N

(T marks terminal states)

Iterations till convergence = 3

The algorithm converges rapidly, and the final utilities are all mostly negative. This is because our discount factor is much less, and as such, the farther rewards are given much less favour. So, the immediate negative step cost dominates the utility values.

1 b.) Step cost = -4.8, discount factor = 0.99

4.800 0.000 40.703 48.000
18.937 27.621 34.924 40.703
14.114 20.501 0.000 34.210
8.312 11.310 -9.600 23.690

Final policy:-
T - E T
E E E N
E N - N
N N T N

Iterations till convergence = 10

Since the discount factor ≈ 1 , there is not much decay in the rewards of the farther states. As such, the policy always tends to direct us to the goal state at (0,3), as it has a very high value.

2 a.) Step Cost = 48, discount factor = 0.99
Final utilities:-

4.800 0.000 3222.557 48.000
3361.273 3374.841 3370.818 3350.118
3376.688 3390.494 0.000 3361.188
3389.144 3401.909 -9.600 3244.024

Policy:-
T - S T
S S W W
E S - -
E - T N

Iterations till convergence = 62

Since the step cost is very large, larger than the reward of the best goal state, the utilities are skewed and the policy sends us in circles. The cell at (1,3) has the highest utility, so it assumes that we cannot move from there.

2 b.) Step cost = -9.6, discount factor = 0.99

Final utilities:-
4.800 0.000 34.069 48.000
-3.282 9.135 23.036 34.069
-14.001 -4.164 0.000 21.675
-24.758 -16.299 -9.600 7.343

Policy:-
T - E T
E E E N
N N - N
N N T N

Iterations till convergence = 10

Here, the negative factor of the step cost initially outweighs all positive benefits, as they occur much farther away. However, as we approach the goal state at (0,3), the utilities become larger as it has a very high reward. So, the policy directs us towards that state.

2 c.) Step cost = -12, discount factor = 0.99

Final utilities:-
4.800 0.000 30.753 48.000
-9.104 -0.039 17.093 30.753
-23.067 -15.888 0.000 15.407
-36.177 -23.503 -9.600 -0.831

Policy:-
T - E T
N E E N
N N - N
N E T N

Iterations till convergence = 12

This case is also almost exact same as 2 b.) expect for the fact that, since the step cost is higher, the negative values are present till farther along. As such, for example at (0,1), the policy directs it to the north as opposed to the east as in the previous case.

2 d.) Step cost = -48,discount factor = 0.99

Final utilities:-
4.800 0.000 -18.984 48.000
-62.459 -122.780 -72.040 -18.995
-122.780 -133.833 0.000 -78.593
-133.832 -76.411 -9.600 -70.347

Policy:-
T - E T
N W N N
N S - N
E E T W

Iterations till convergence = 6

Here, the unit step cost is extremely high, and as such, even the negative termination state is a valid place for the agent. So, the agent directs itself towards it’s closest termination state, as the step cost is high for it to even consider the farther states.