
Exercise 10 – Reinforcement Learning: Construction of Reward Matrix

Date: 24/11/2022

Aim:

To write functions for the following using Python:

1. Construct a state diagram for the given scenario.
2. Read and store the state diagram using appropriate data structure.
3. Construct a reward matrix and find the path using MDP sequence with passive RL.

Code:

```
#Reinforcement Learning - Construction of Reward Matrix

reward = [[-5, 0, 0, 0, 0],
        [0, -5, 0, 0, 0],
        [0, 0, -5, 0, 0],
        [0, 0, 0, -5, 0],
        [0, 0, 0, 0, -5]]

valid = [[0, 0, 1, 1, 0],
        [0, 0, 0, 0, 0],
        [0, 1, 0, 0, 1],
        [0, 1, 1, 0, 0],
        [0, 1, 0, 0, 0]]
pl = {0 : "Canteen", 1 :  "CSE", 2 : "ECE", 3: "Admin", 4 : "Auditorium"}

def find_reward(node, l, vst):
  for x in range(5):
    if valid[node][x] > 0:
      if reward[node][x] < pow(5, l):
        reward[node][x] = max(reward[node][x], pow(5, l))
        vst1 = vst + [x]
        find_reward(x, l + 1, vst1)

def find_path(node, vst):
  if node == 1:
    print("Visited: ", vst)
    reward_val = 0
```

```
    t = 1
    for (x, y) in vst:
      reward_val += y * t
      t *= 0.2
    print("\tReward value: ", reward_val)
    return

  for x in range(5):
    if reward[node][x] > 0:
      vst1 = vst + [(x, reward[node][x])]
      find_path(x, vst1)
find_reward(0, 1, [0])

print("Reward matrix: ")
for i in reward:
  print(i)

print("\n\nFinding path: \n")
find_path(0, [(0, 0)])
```

Output:

```
~/AIwork$ python ex10rl.py
Reward matrix:
[-5, 0, 5, 5, 0]
[0, -5, 0, 0, 0]
[0, 125, -5, 0, 125]
[0, 25, 25, -5, 0]
[0, 625, 0, 0, -5]


Finding path:

Visited:  [(0, 0), (2, 5), (1, 125)]
    Reward value:  6.000000000000001
Visited:  [(0, 0), (2, 5), (4, 125), (1, 625)]
    Reward value:  11.000000000000002
Visited:  [(0, 0), (3, 5), (1, 25)]
    Reward value:  2.0
Visited:  [(0, 0), (3, 5), (2, 25), (1, 125)]
    Reward value:  3.0
Visited:  [(0, 0), (3, 5), (2, 25), (4, 125), (1, 625)]
    Reward value:  4.0
~/AIwork$
```