MAS: Hide and Seek (Part 1)

Author: Andrei Gabriel Popescu, IA-B

How I trained the entities

There are two possibilities when creating the training for this type of environment:

- Using two different Q helper structures (one per hider)
- Using only one and train them together

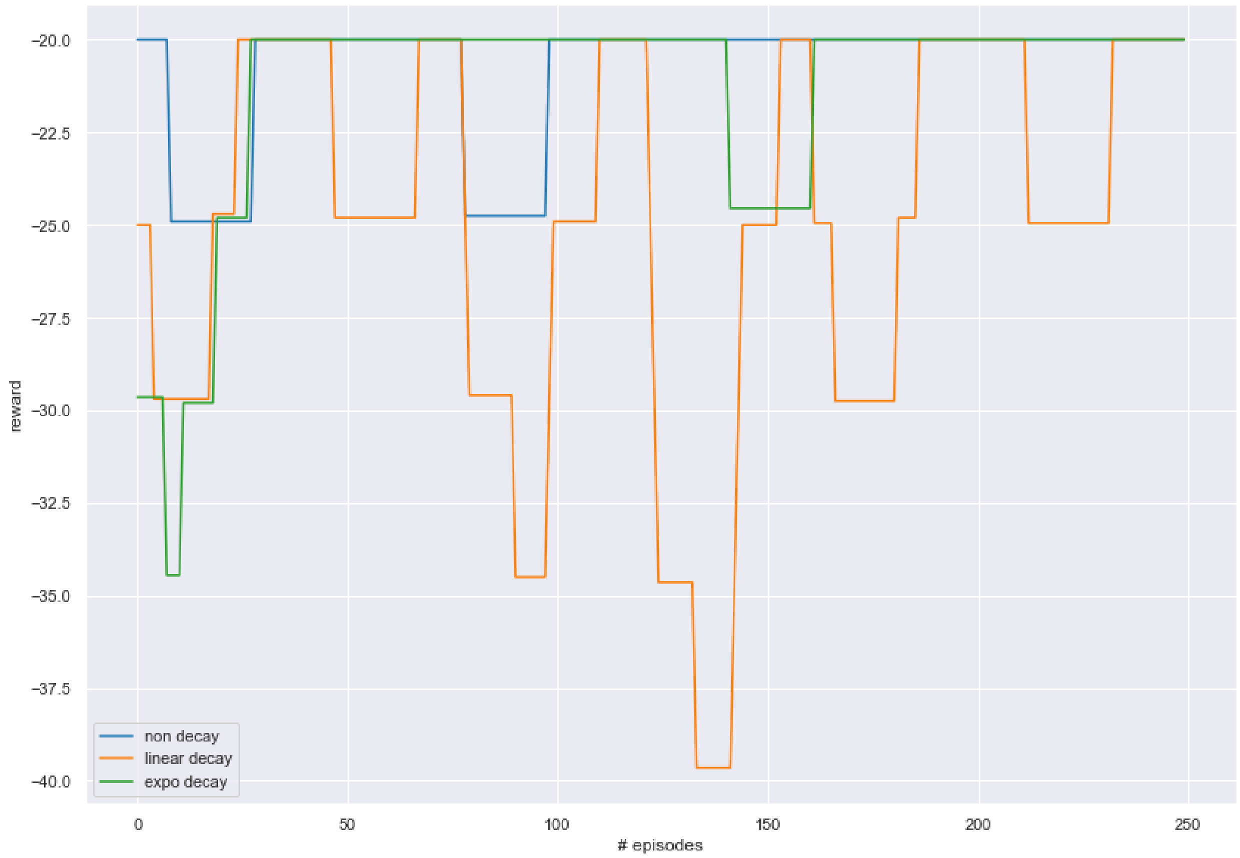
Both methods take into consideration a possible collaboration between the two agents, and I explored both of them, but the final version of the code implements only the last method.

Results

Both algorithms were trained: 5k episodes

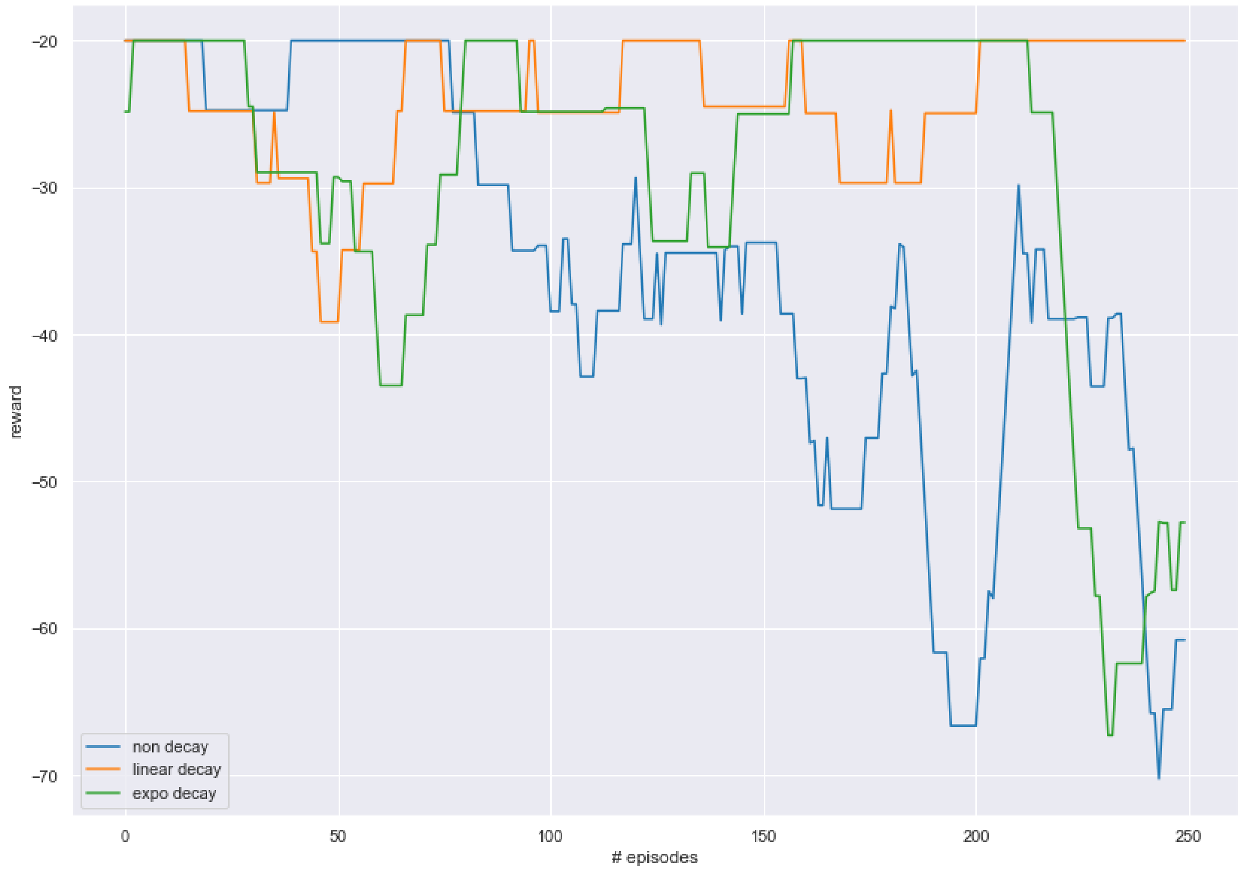
Q learning agents:

The parameters used are: α = 0.75, γ = 0.95.



SARSA agents:

The parameters used are: α = 0.75, γ = 0.95.



Part 2: POMDP

How I modeled the problem:

We have

- 5 states: from A0 to A4
- 2 actions: LEFT, RIGHT (STAY)
- 2 observations: O_2, O_3 (for 2 and 3 walls)

The next matrices are computed taking into account the data from the problem formulation.

Task 1:

- Compute state action transition probabilities (I have used this the environment implementation)
- The main idea is that being a two actions env we have either full (1.0) or none (0.0) probabilities

```
left_action = np.array([[
            [1.0, 0.0, 0.0, 0.0, 0.0],
            [1.0, 0.0, 0.0, 0.0, 0.0],
            [0.0, 1.0, 0.0, 0.0, 0.0],
            [0.0, 0.0, 1.0, 0.0, 0.0],
            [0.0, 0.0, 0.0, 1.0, 0.0]
        ]])
```

```
right_action =  np.array([[
            [0.0, 1.0, 0.0, 0.0, 0.0],
            [0.0, 0.0, 1.0, 0.0, 0.0],
            [0.0, 0.0, 0.0, 1.0, 0.0],
            [0.0, 0.0, 0.0, 0.0, 1.0],
            [0.0, 0.0, 0.0, 0.0, 1.0]
        ]])
```

Observations

```
obs_left = np.array([[
            [0.0, 1.0],
            [1.0, 0.0],
            [1.0, 0.0],
            [1.0, 0.0],
            [0.0, 1.0]
        ]])

obs_right = np.array([[
    [0.0, 1.0],
    [1.0, 0.0],
    [1.0, 0.0],
    [1.0, 0.0],
    [0.0, 1.0]
]])
```

Rewards:

```
R_left = np.array([[-1, -1, -1, 0, -1]])
R_right = np.array([[-1, 0, -1, -1, -1]])
```

What is the observed best policy of the agent? How do you justify the resulting actions?

- Present in the notebook