Q-Witcher

Mehmet Fatih Gülakar

June 2019

1 Setup

11x11 maze is generated by hand with unpassable mountains. Geralt of Rivia's initial position will be (0,0) and basilisk's position is 11,11. For the rewards;

- Each step is -1 point.
- Trying to go mountains is -5 points (although it is not possible.)
- Slaying the basilisk is 100 points.
- Entering the areas with poison mist is -100 points.

For the Q-learning parameters;

- Learning rate α is chosen as 0.2, 0.7 and 0.95.
- Discount rate γ is chosen as 0.3, 0.5 and 0.9.
- Exploration rate ϵ is chosen as 0.2, 0.55 and 0.85

Also, maximum number of steps and epochs are chosen as 100.

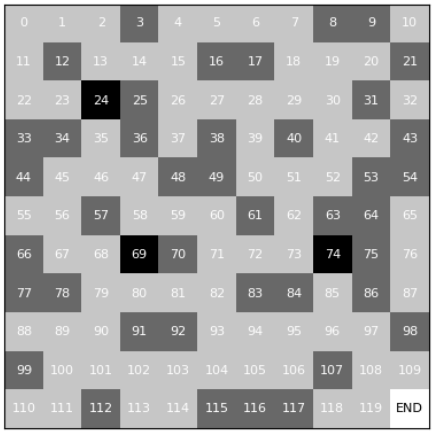
This report does not include probability density function of finding Geralt in a place and SARSA implementation(you can find script I tried, but it does not work), due to lack of my coding skills.

2 Results

For the path to the basilisk, there is a real-time visualization of it when Python script is run on terminal.

2.1 With poisonous mist

Figure 1: World map with toxic mist



2.1.1 Effect of various learning rates

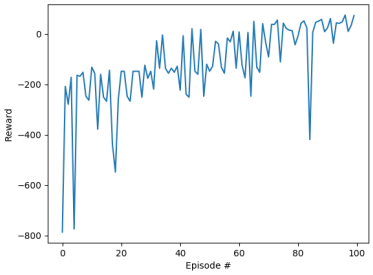
Discount rate = 0.9 and exploration rate = 0.2 is chosen. Interpreting the results, one should say low learning rate is not optimal with $\gamma = 0.9$ and $\epsilon = 0.2$. Other than that, 0.7 and 0.95 as learning rate gave similar results.

2.1.2 Effect of various discount rates

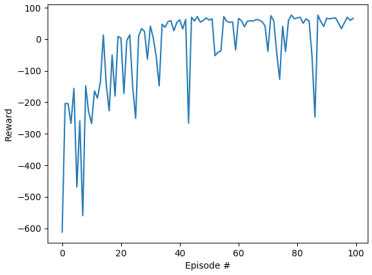
Learning rate = 0.7 and exploration rate = 0.2 is chosen. By interpreting the results, one can say that higher γ resulted in reaching of reward to the non-negative value faster. This can be related with connection between long-time reward and discount factor.

2.1.3 Effect of various exploration rates

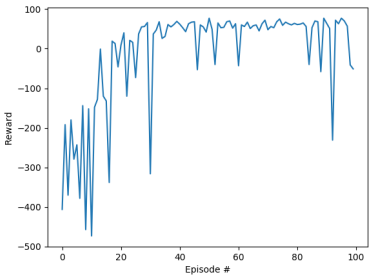
Learning rate = 0.7 and discount rate = 0.9 is chosen. By interpreting the results, one can say that higher exploration level results in more random trial. Therefore, for the results ϵ =0.2, reward is changing frequently. Also, for the ϵ=0.85, total reward cannot reach 0.



(a) α=0.2

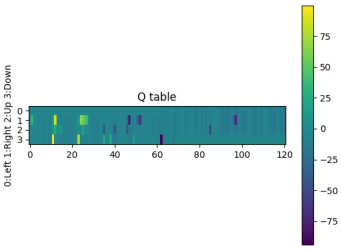


(b) α=0.7

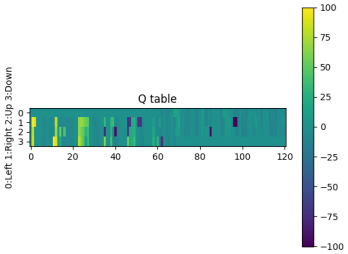


(c) α=0.95

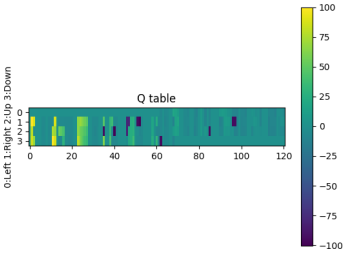
Figure 2: Reward-episode graphs



(a) α=0.2

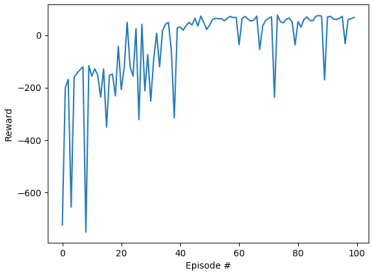


(b) α=0.7

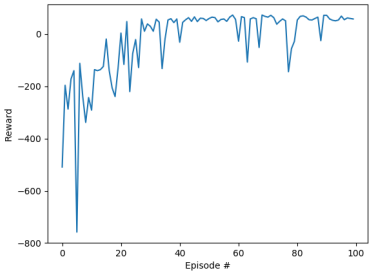


(c) α=0.95

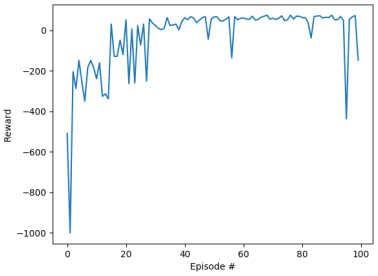
Figure 3: Q-tables



(a) γ=0.3

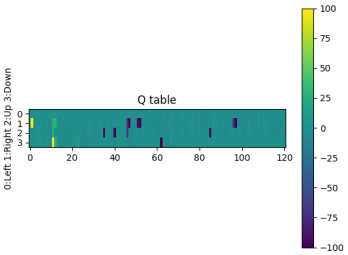


(b) γ=0.5

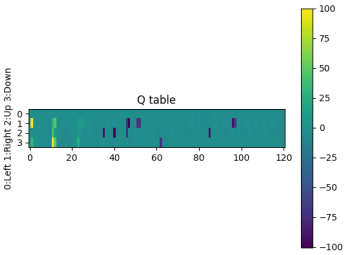


(c) γ=0.9

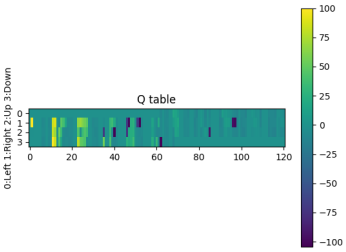
Figure 4: Reward-episode graphs



(a) γ=0.3

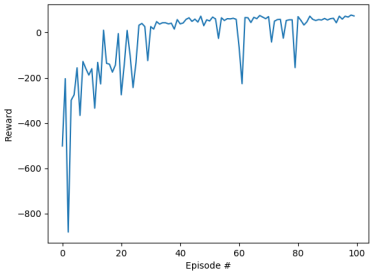


(b) γ=0.5

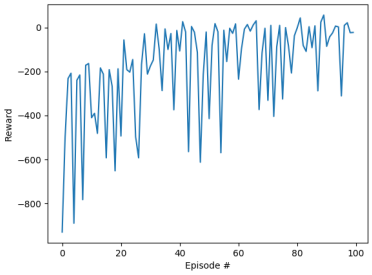


(c) γ=0.9

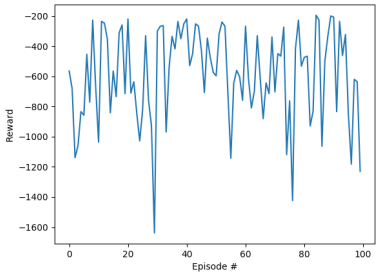
Figure 5: Q-tables



(a) ϵ=0.2

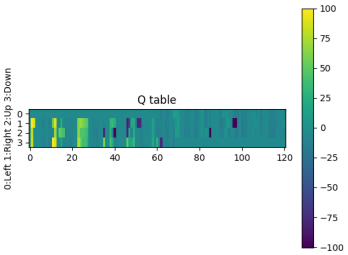


(b) ϵ=0.55

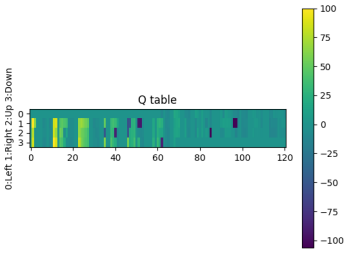


(c) ϵ=0.85

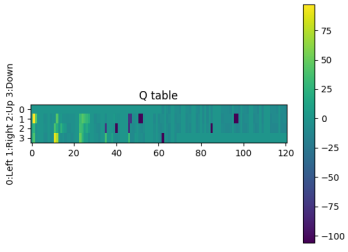
Figure 6: Reward-episode graphs



(a) ϵ=0.2



(b) ϵ=0.55

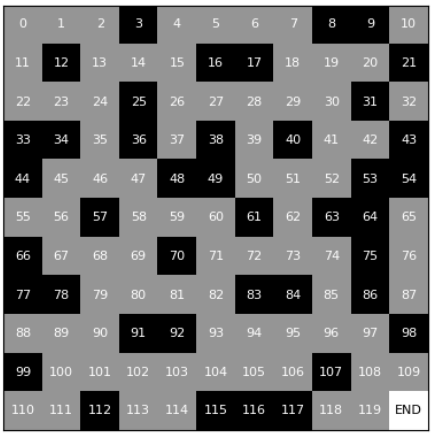


(c) ϵ=0.85

Figure 7: Q-tables

2.2 Without poisonous mist

Figure 8: World map without toxic mist



2.2.1 Effect of various learning rates

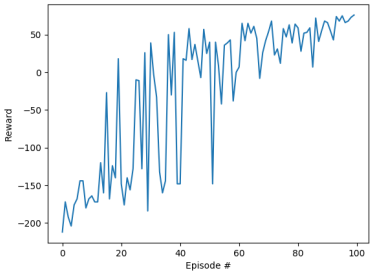
Discount rate = 0.9 and exploration rate = 0.2 is chosen. Low α value results in frequent change in result, and high value gives quickest reach to the high reward.

2.2.2 Effect of various discount rates

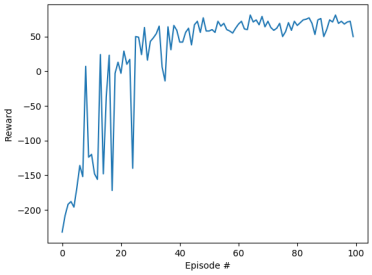
Learning rate = 0.7 and exploration rate = 0.2 is chosen. By looking to the Q-tables, one can say that future rewards are higher at higher γ values, which corrects the fact that discount rate determines the whether Geralt seeks long-term reward or not.

2.2.3 Effect of various exploration rates

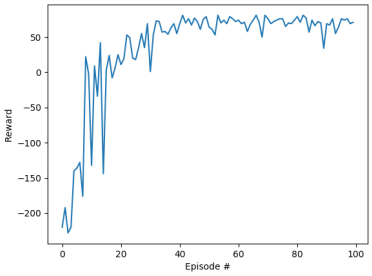
Learning rate = 0.7 and discount rate = 0.9 is chosen. Same situation with non-poisonous map, higher ϵ results in non-consistent reward change. Moreover we can say that, for ϵ=0.85, total reward is below -1, which is unique among others, since only thing that leads to losing points in the map is stepping, which means the Geralt took several random action.



(a) α=0.2

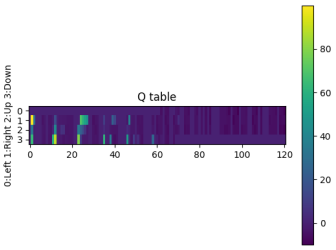


(b) α=0.7

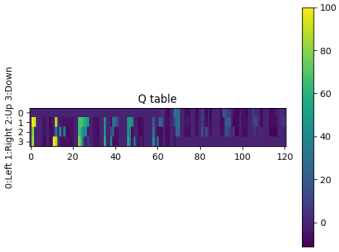


(c) α=0.95

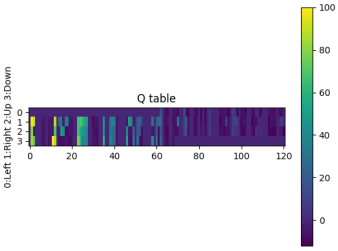
Figure 9: Reward-episode graphs



(a) α=0.2

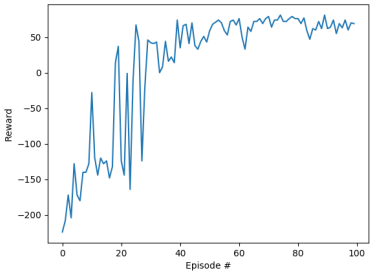


(b) α=0.7

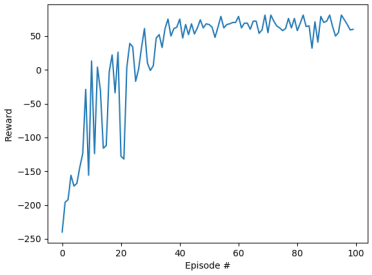


(c) α=0.95

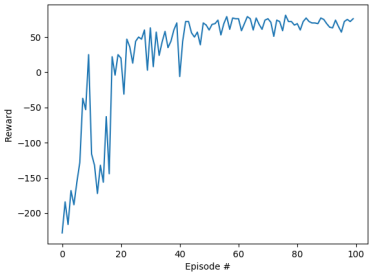
Figure 10: Q-tables



(a) γ=0.3

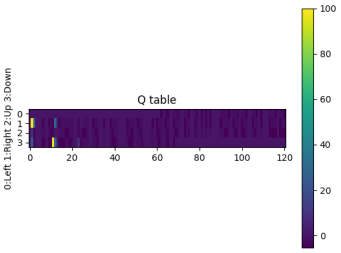


(b) γ=0.5

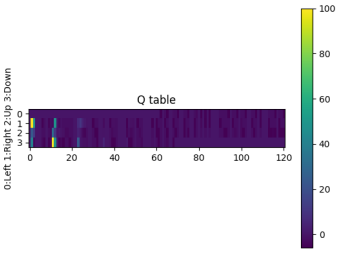


(c) γ=0.9

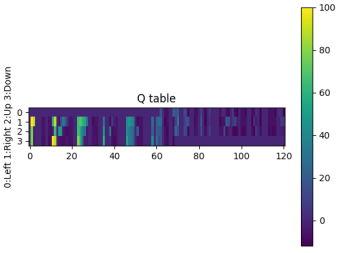
Figure 11: Reward-episode graphs



(a) γ=0.3

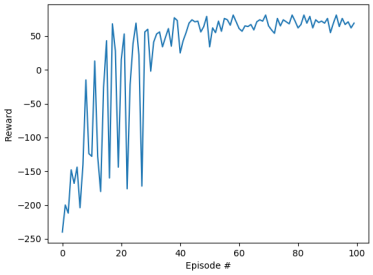


(b) γ=0.5

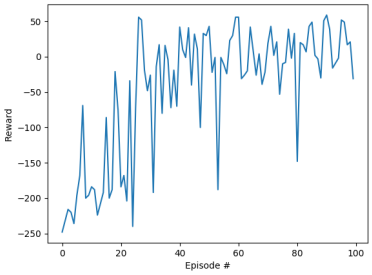


(c) γ=0.9

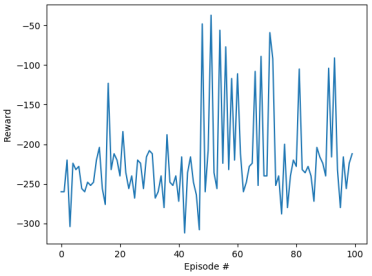
Figure 12: Q-tables



(a) ϵ=0.2

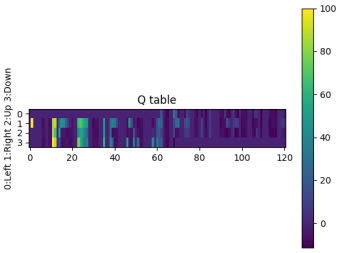


(b) ϵ=0.55

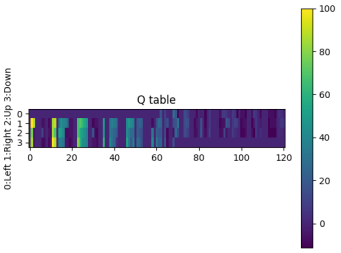


(c) ϵ=0.85

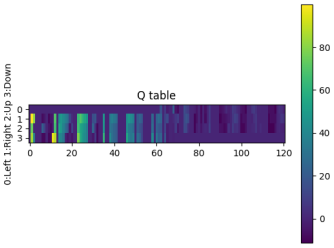
Figure 13: Reward-episode graphs



(a) ϵ=0.2



(b) ϵ=0.55



(c) ϵ=0.85

Figure 14: Q-tables