Programming Assignment 4

CS 747

Aditya Jadhav

8 November, 2019

Configurations and conventions

I have used the following parameters to generate the graphs below -

● Grid size	:	7 x 10
● Exploration rate	:	0.5 / (numEpisodes + 1)
● Learning rate	:	0 . 5
● Start state	:	[3 , 0]
● Goal state	:	[3 , 7]
● Winds	:	[0 , 0 , 0 , 1 , 1 , 1 , 2 , 2 , 1 , 0]

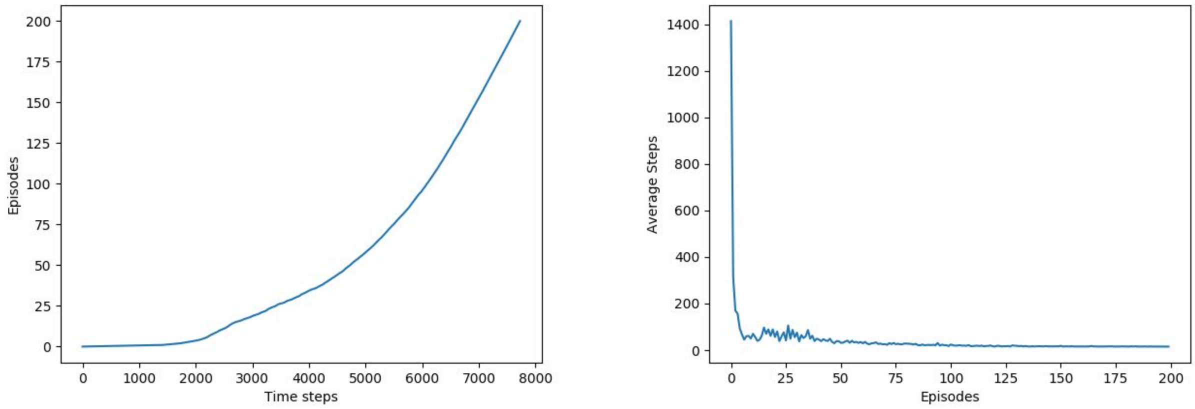
Some of the conventions I have used -

- **Epsilon** was decayed with the number of episodes as, **0.5 / (numEpisodes + 1)**
- If agent tries to get out of the boundary, it remains in that state and gets a reward of **-1**
- In the case of **stochastic winds**, the wind values will be **{ w_i + 1 , w_i , w_i - 1 }**

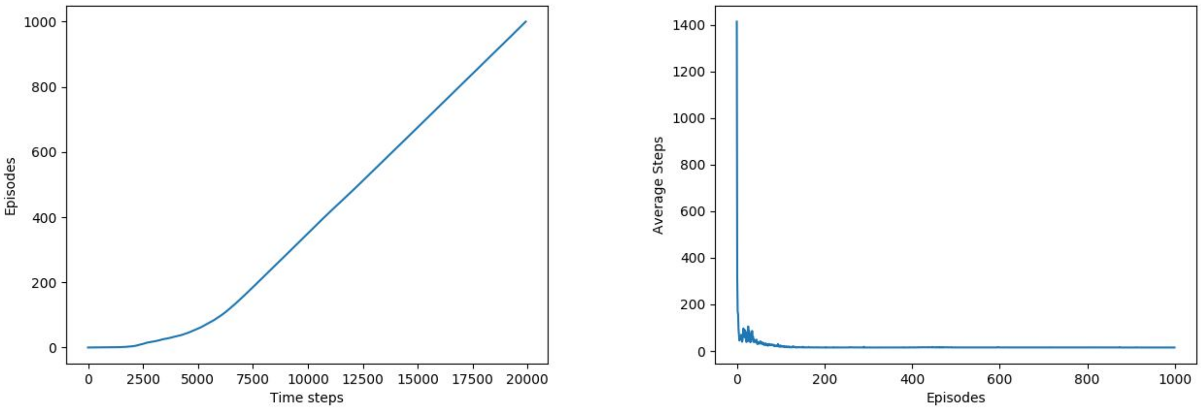
Finally the structure of code is as follows -

- **Class Gridworld** consists of model of "Gridworld" with the required functions.
- **getAction** tries exploration, exploitation and returns random, best action respectively.
- **getNextState** performs the action and returns next state. It also incorporates winds.
- **getEpisodeOutcme** executes one episode and returns number of steps to reach goal.
- **getOptimalPolicy** prints the QValues for each state.
- **Sarsa** performs 10 runs for seeds {0..9} and plots "Time Steps" v/s "Episodes" graph.

Task 1



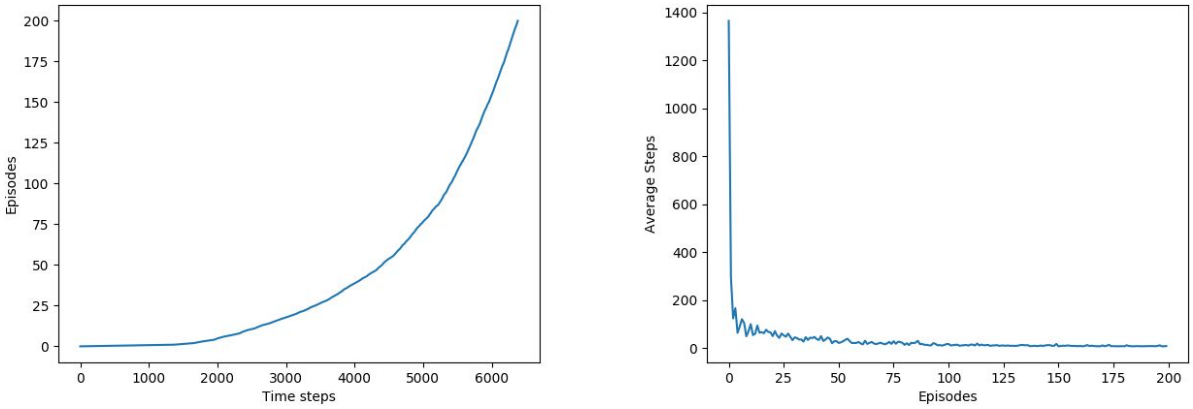
(Figure 1 : Without king's moves and stochasticity - 200 episodes)



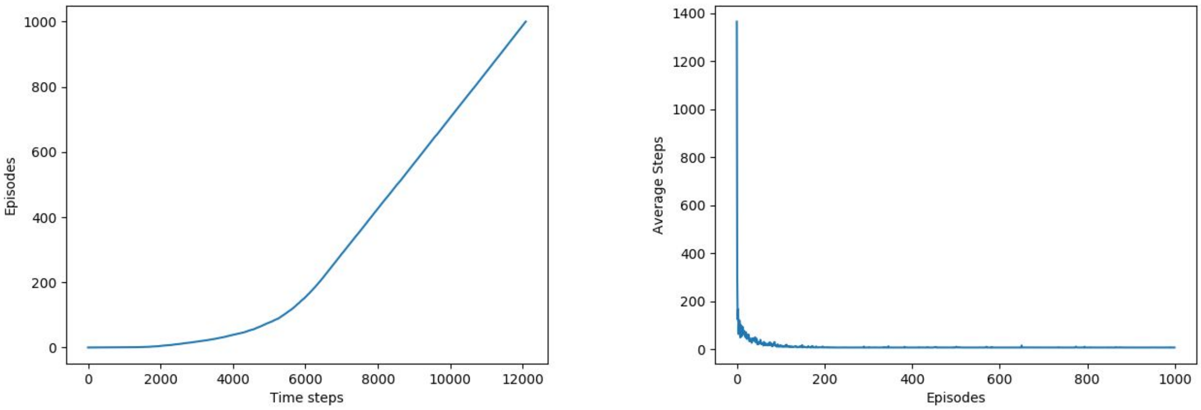
(Figure 2 : Without king's moves and stochasticity - 1000 episodes)

1. I was able to reach a minimum of 15 steps/episode after iterating over 200 episodes.
2. Initially the agent does not have a good estimate of the optimal policy hence takes longer to reach the goal state. This justifies the sudden increase in the number of steps to reach the goal state for the first few episodes.
3. The increasing slope of the graph shows that the goal state is reached more and more quickly over time. This is expected as the number of episodes increase the agent starts to get a good estimate of the optimal policy. Hence takes lesser moves.

Task 2



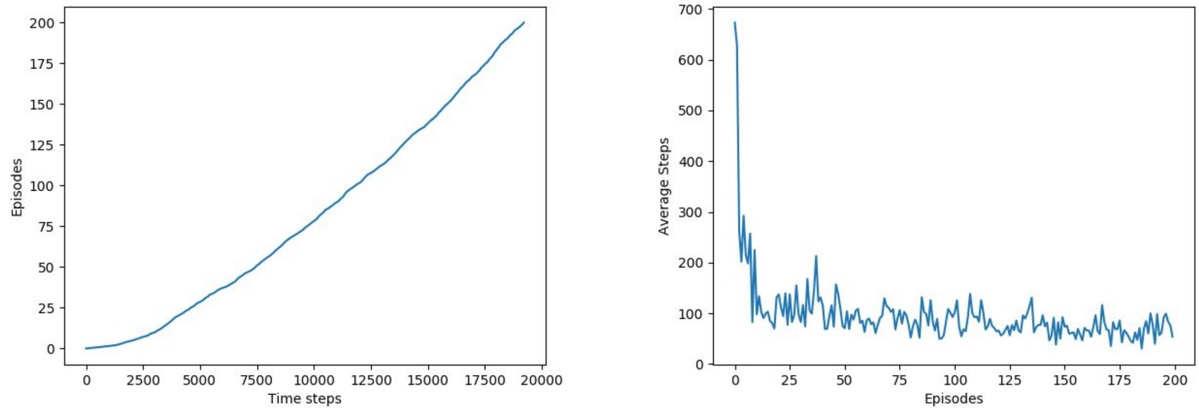
(Figure 3 : With king's moves but no stochasticity - 200 episodes)



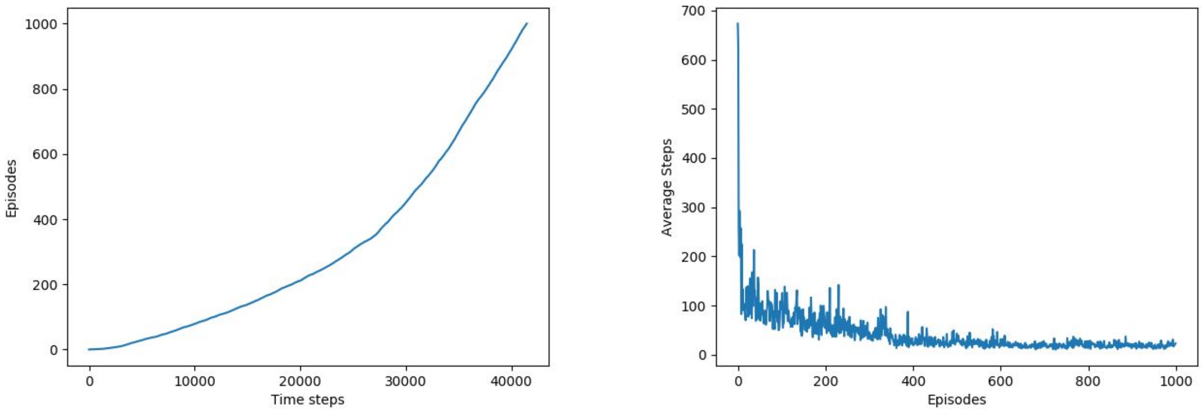
(Figure 4 : With king's moves but no stochasticity - 1000 episodes)

1. I was able to reach a minimum of 7 steps/episode after iterating over 140 episodes.
2. The 2nd and 3rd observations of Task 1 can also be made here.
3. Here we observe that the slope of the graph is higher than that in task 1. This suggests that the agent is now able to learn the optimal strategy faster. This is expected as the agent now converges on a shorter optimal path than before.
4. Due to this faster convergence to optimum path, many of the states in the top-right part of gridworld remained unexplored and had Qvalue of 0.

Task 3



(Figure 5 : With king’s moves and stochasticity - 200 episodes)



(Figure 6 : With king’s moves and stochasticity - 1000 episodes)

1. I was able to reach a minimum of 30 steps/episode after iterating over 1000 episodes.
2. It can be seen that the slope of the graph in this case is less than that of the earlier 2 cases. The lesser slope indicates that the agent is not necessarily reaching the goal faster as time passes. This is due to the added stochastic wind that the agent is now learning the optimal strategy slower, which is expected as the environment is stochastic and thus the optimal strategy isn’t the same for all cases.
3. We can see that even after running for 1000 episodes the “Average steps” do not converge and keep fluctuating. This is due to the stochastic nature of winds.