

Stochastic Shortest Path Maze

CB.EN.P2AIE22001

ANIRUDHAN K S

ENVIRONMENT USED

- FrozenLake is a navigation problem in the Gym library.
- The goal is to reach the end of the lake while avoiding falling into holes.
- Action Space: Discrete(4) (LEFT, DOWN, RIGHT, UP)
- Observation Space: Discrete(16)
- Enabling Slipper It will move in intended direction with probability of 1/3 else will move in either perpendicular direction with equal probability of 1/3 in both directions.



Q learning

- Q-learning is a model-free reinforcement learning algorithm
- It learns the value of an action in a particular state
- It does not require a model of the environment

Q learning

- Q learning uses Q table to find the best action
- Q table is updated using bellman equation
- Q table consist of State X Action table
- State :- all the states the agent can reach
- Action:- all the the actions that an agent can take
- Q table is updated by using the bellman equation

Q learning

$$Q^{new}(s_t, a_t) \leftarrow \underbrace{Q(s_t, a_t)}_{\text{current value}} + \underbrace{\alpha}_{\text{learning rate}} \cdot \underbrace{\left(\underbrace{r_t}_{\text{reward}} + \underbrace{\gamma}_{\text{discount factor}} \cdot \underbrace{\max_a Q(s_{t+1}, a)}_{\text{estimate of optimal future value}} - \underbrace{Q(s_t, a_t)}_{\text{current value}} \right)}_{\text{temporal difference}}$$

new value (temporal difference target)

r_t is the reward received when moving from the state S_t to the state S_{t+1}

$\max Q(s_{t+1}, \mathbf{a})$:-the maximum reward that can be obtained from state

Q learning

- We have 15 states and 4 actions , we have a 15 X 4 Q table
- All the values are initialized to zero

	LEFT	DOWN	RIGHT	UP
0	0	0	0	0
1	0	0	0	0
.....
.....
14	0	0	0	0
15	0	0	0	0

Q learning

- Initially the agent randomly explores the maze
- After 1000 iterations the agent starts to use the Bellman equation to update Q table
- The updated Q table after 100000 iterations

	LEFT	DOWN	RIGHT	UP
0	1.59e-04	5.41e-04	2.35e-04	1.82e-04
1	5.21e-04	6.16e-04	1.48e-03	4.12e-04
.....
.....
14	7.29e-03	1.65e-01	3.58e-01	1.83e-01
15	0	0	0	0

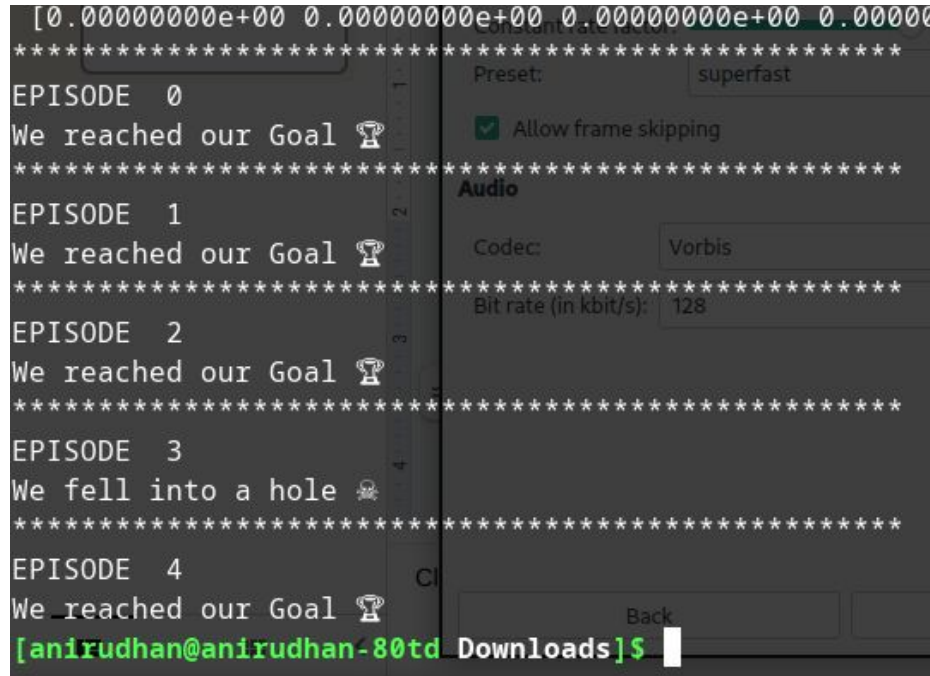
Q learning

- After updating the Q table the agent uses it to find the path to goal
- It choose the best action at each state from the Q table

Result

- By running 5 episodes the agent fails once and successfully win's 4 times

```
[0.00000000e+00 0.00000000e+00 0.00000000e+00 0.00000000e+00]
*****
EPISODE 0
We reached our Goal 🏆
*****
EPISODE 1
We reached our Goal 🏆
*****
EPISODE 2
We reached our Goal 🏆
*****
EPISODE 3
We fell into a hole 🕒
*****
EPISODE 4
We reached our Goal 🏆
[anirudhan@anirudhan-80td Downloads]$
```



Result



Reference

- <https://en.wikipedia.org/wiki/Q-learning>
- <https://ojs.aaai.org/index.php/AAAI/article/view/11170>
- <https://www.youtube.com/watch?v=9g32v7bK3Co>
- <https://www.youtube.com/watch?v=HpaHTfY52RQ>