Reinforcement Learning for Maze solving

NISHANTH. VANIPENTA

UID:116409605

UNIVERSITY OF MARYLAND 18 56

A. JAMES CLARK SCHOOL OF ENGINEERING

Outline

- Introduction
- Problem Overview
- Reinforcement learning problem
  - Algorithm
  - Demonstration and Analysis
  - Conclusion

Introduction

Q-Learning is a popular method to solve many problems by training an AI system. In our approach, a maze solving learning was developed using Q-learning algorithm. The problem attends to fundamental robotic concepts of path planning.

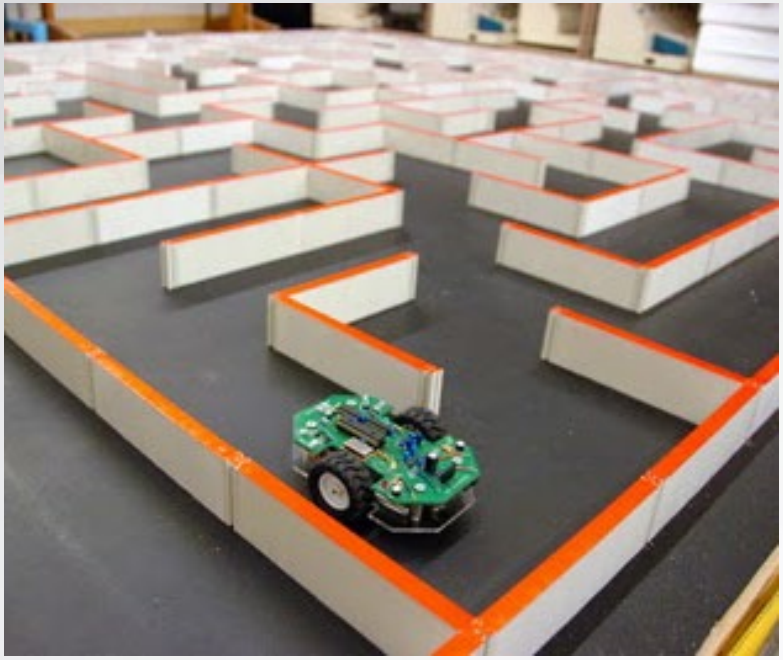


Image*- IEDProjects2015@IIITD

A. JAMES CLARK SCHOOL OF ENGINEERING

Q-Learning

- Q-Learning is a reinforcement learning technique that involves the training of a Q-table to evaluate an optimal path for a given RL problem.
- Q-table that has the details of expected future reward for each action based on the current state.
- A reward is given for each action based on the correspondence to the goal.

Q Learning

- In our model, the agent is encouraged to find the optimal path to the target cell by rewarding the actions.

Cell	Reward
Goal	+5
Pits	-10
Walls	-1
Free cells	-0.1

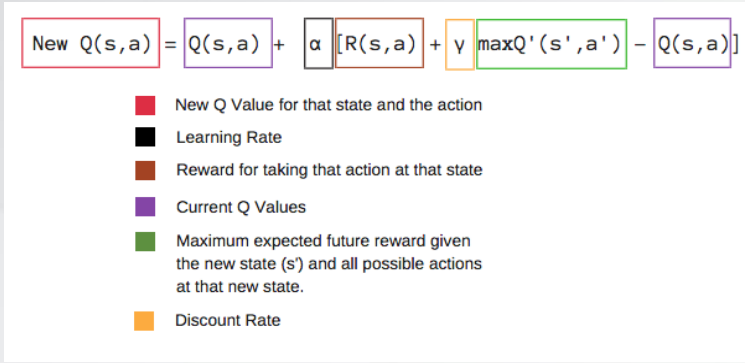
Actions	Encoded
UP	0
Down	1
Left	2
Right	3

- For each action, the agent must take an action, which transits agent from current state to new state

STATE	0	1	2	3
[5, 45, 5, 45]	0	0	0	0

Method of Q-Learning

- After transition to new state, a reward is generated, and the objective is to maximize the value of this reward over time



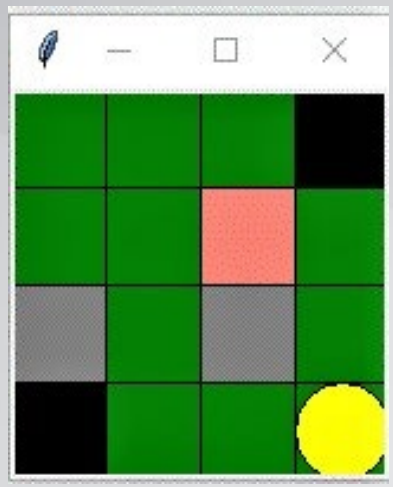
- The next action is predicted based on the max q value for that state. Then the Q table is updated for each iteration append all the values in the Q matrix

$$Action = \begin{cases} \max_a Q(s,a), & R > \varepsilon \\ Random\ a, & R \leq \varepsilon \end{cases}$$

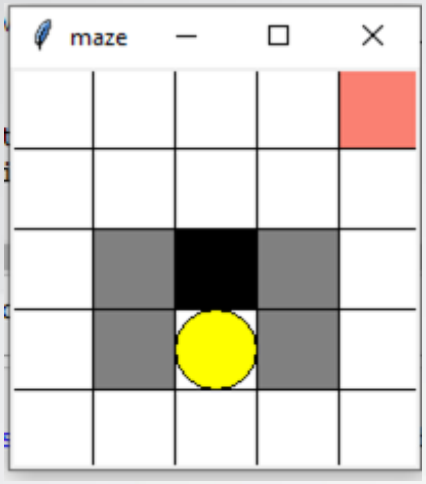
R is random number between 0 and 1, ε is the exploration factor between 0 and 1. If ε is 0.1, then 10% of the times, the algorithm will select a random action to explore corresponding rewards.

Environment

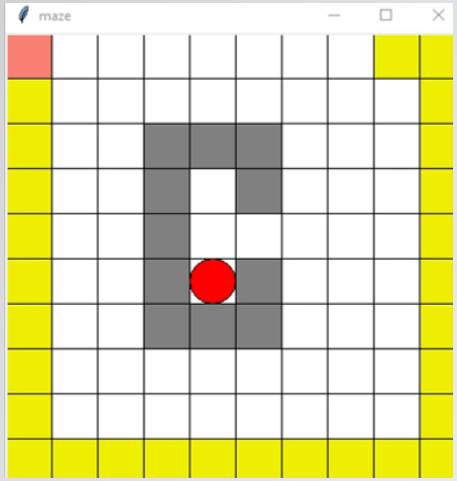
The environment created is split into 3 different grid sizes:



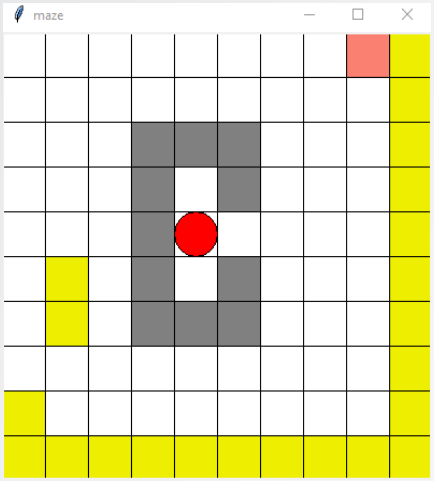
4x4 grid



5x5 grid



10x10 complex grid



10x10 standard grid

A. JAMES CLARK SCHOOL OF ENGINEERING

Parameters and Q-Table

Sample Q-table for 10x10 grid:

States ↓ / Actions →	0	1	2	3
[0.0, 0.0, 40.0, 40.0]	-0.02	-0.02	-0.02	-0.02
[40.0, 0.0, 80.0, 40.0]	0	-0.02	0	0
[40.0, 40.0, 80.0, 80.0]	-0.0392	-0.05936	-0.032	-0.05504
[80.0, 40.0, 120.0, 80.0]	-0.0392	-0.04	-0.02	-0.0392
[120.0, 40.0, 160.0, 80.0]	0	0	0	-0.02
[40.0, 80.0, 80.0, 120.0]	-0.04352	-0.04352	-0.04	-0.0392
[40.0, 120.0, 80.0, 160.0]	-0.04611	-0.032	7.168	-0.02
[80.0, 120.0, 120.0, 160.0]	-0.04	0	32	0
[120.0, 120.0, 160.0, 160.0]	0	0	0	0
[0.0, 40.0, 40.0, 80.0]	-0.02	-0.04	-0.03152	0

Values after tuning:

Parameters	**Threshold**
Epsilon	0.9
Learning rate	0.7
Discount factor	0.9

Values before tuning:

Parameters	**Threshold**
Epsilon	0.6
Learning rate	0.4
Discount factor	0.9

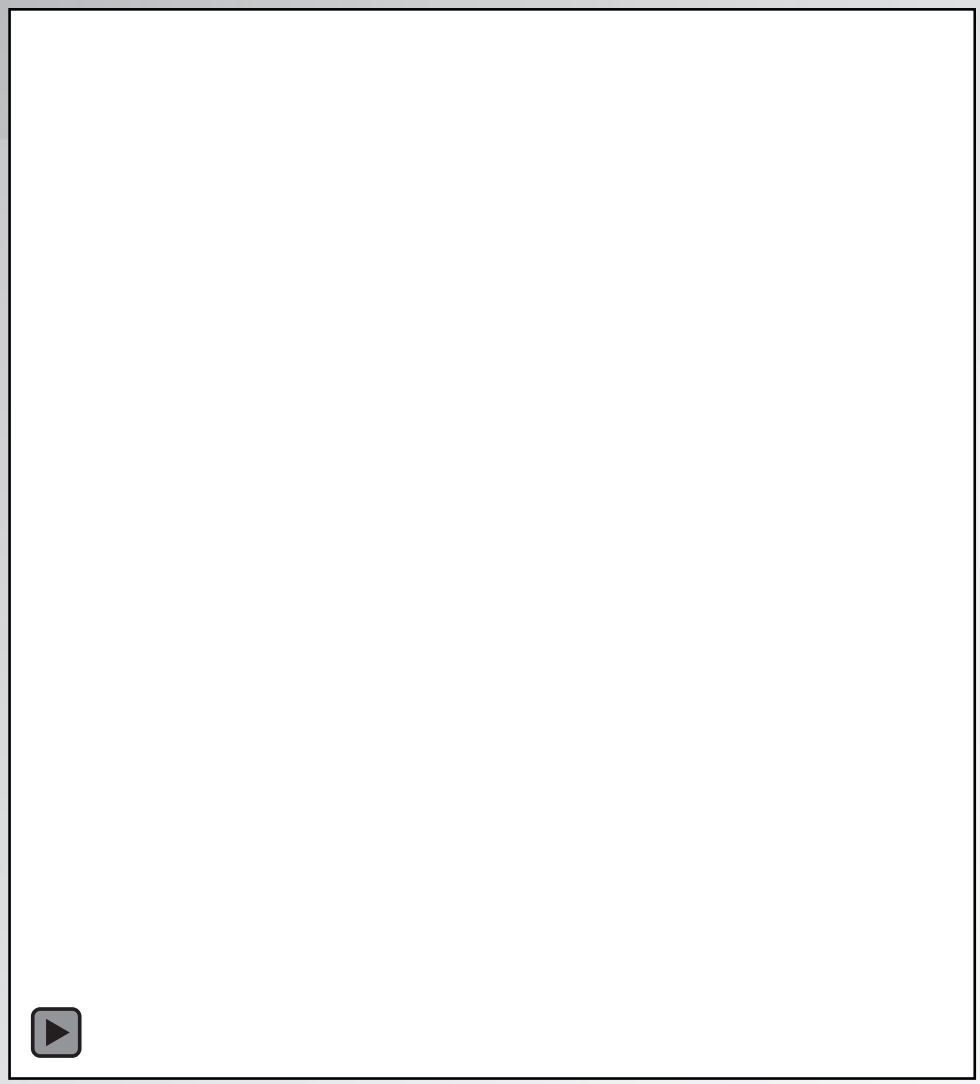
A. JAMES CLARK
SCHOOL OF ENGINEERING

4x4 grid output

Maze Size (4X4)	Number
Average time of execution	0.487 sec
Average Number of iteration	15episodes

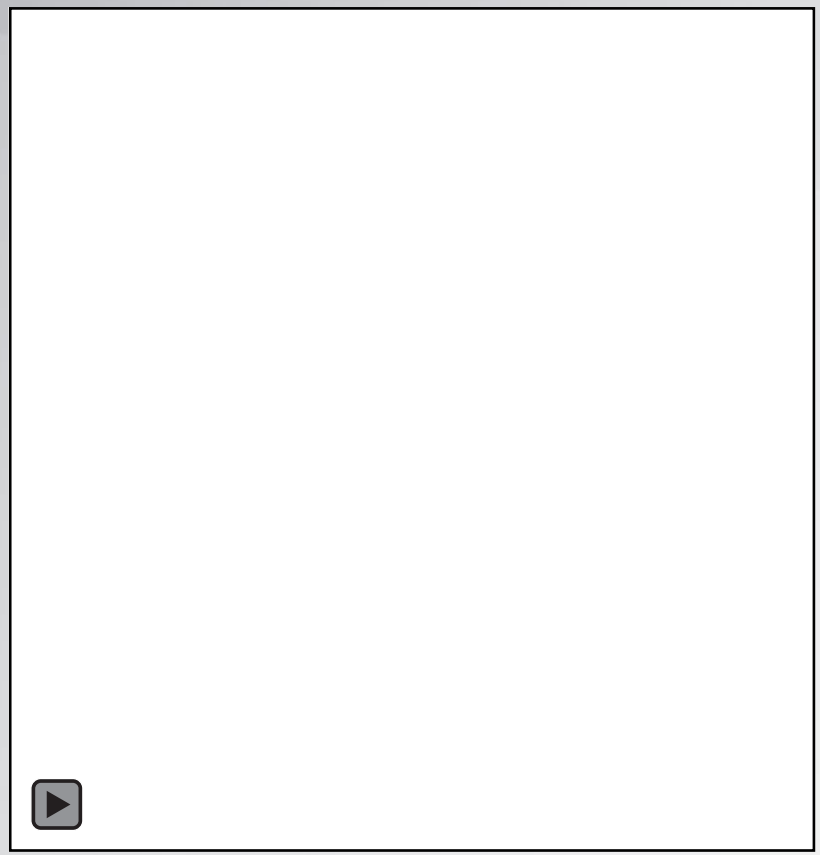
A. JAMES CLARK
SCHOOL OF ENGINEERING

5x5 grid output



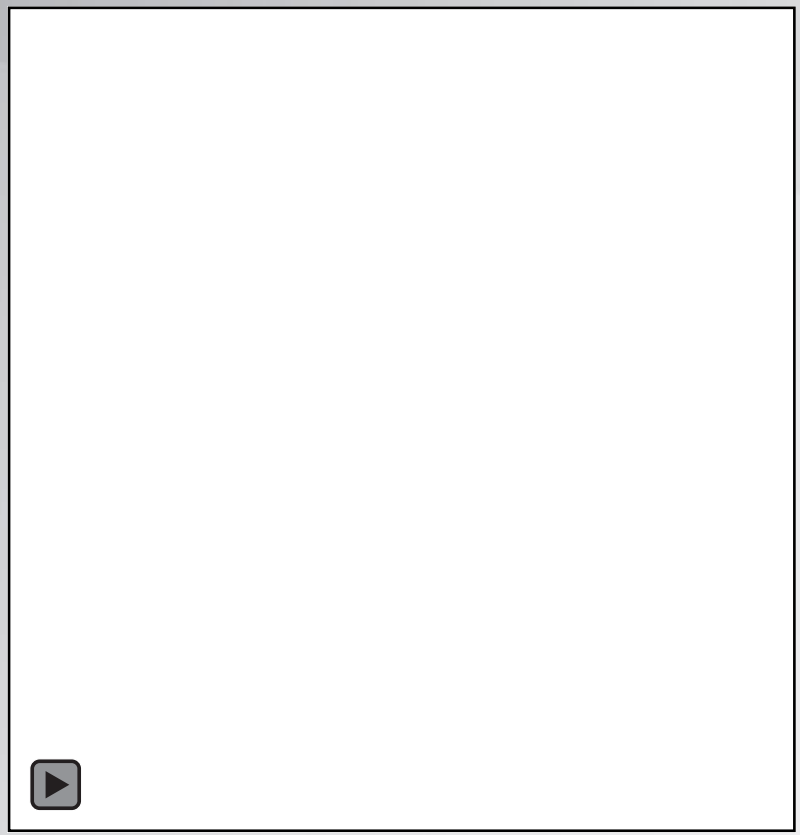
Maze Size (5X5)	Number
Average time of execution	1.2712 sec
Average Number of iteration	20 episodes

10x10 Standard space



Maze Size (10X10)	Number
Average time of execution	2.4 sec
Average Number of iteration	1250 episodes

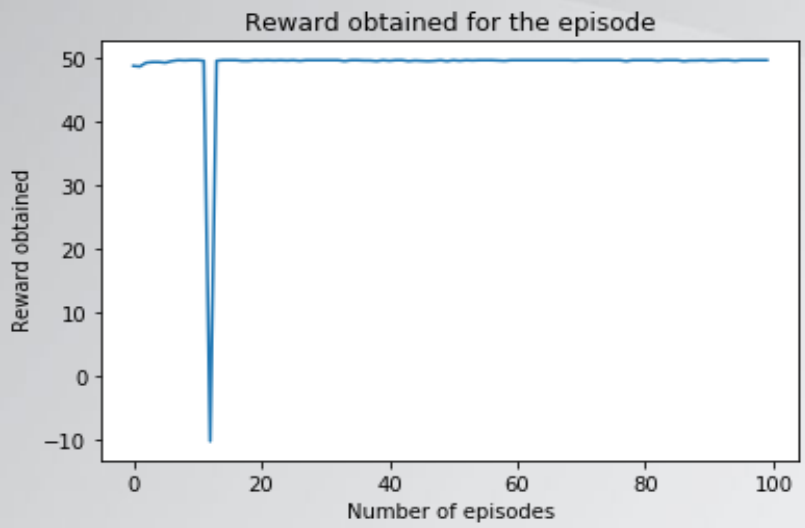
10x10 Complex space



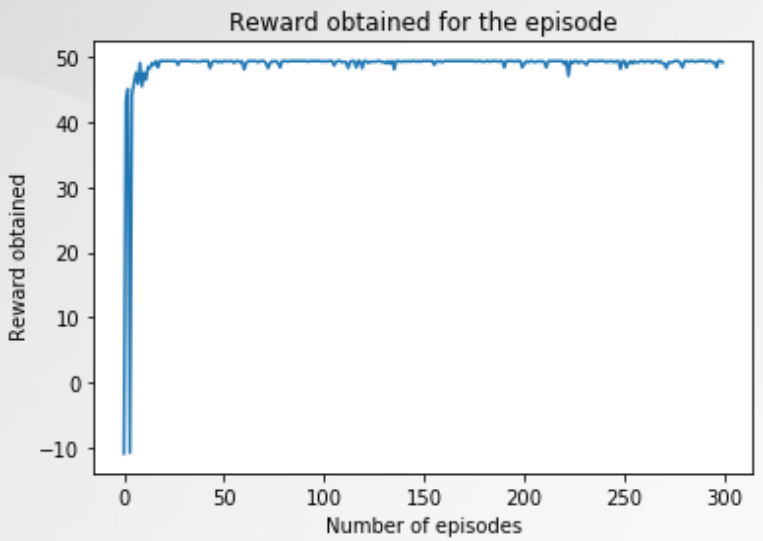
Maze Size (10X10)	Number
Average time of execution	3.7 sec
Average Number of iteration	1600 episodes

Plots for 4x4 and 5x5

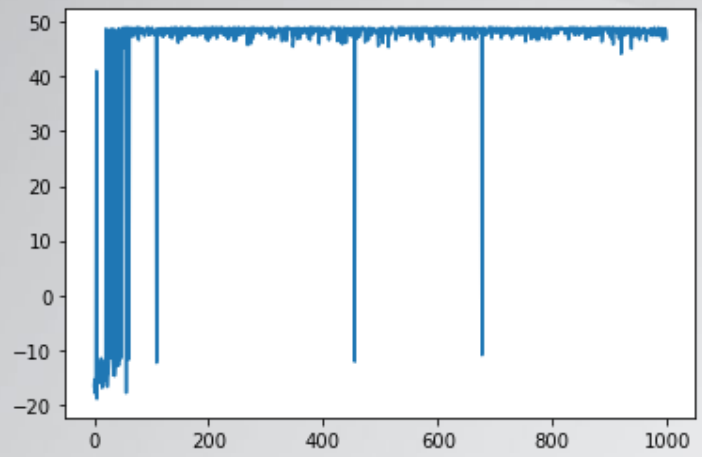
Plot for 4x4 grid



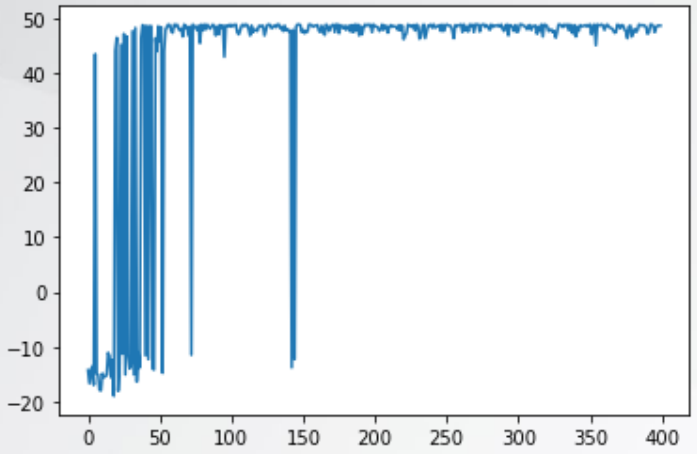
Plot for 5x5 grid



Plots for 10x10 grids



Complex 10x10



Standard 10x10

Analysis

- Upon Observation, for a 10x10 grid it is observed that tuning the learning rate has a lot of affect in the rate of convergence of the algorithm.
- For a complex tasks with huge grid space, Q learning isn’t the first choice because of the memory usage and time inefficient. Markov’s decision process or a Deep reinforcement would work better.
- Also, the agent always tries to take the optimal path during learning. This does allow the agents to adapt to new strategies.

References

1. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5967320

2. IEDProjects2015@IIITD

3. http://www.mitchellspryn.com/2017/10/28/Solving-A-Maze-With-Q-Learning.html

THANK YOU!

UNIVERSITY OF MARYLAND 1856

A. JAMES CLARK
SCHOOL OF ENGINEERING