Ketan Karnakota
3 October 2019

MDP Techniques

Introduction:

Our central research pursuit is to compare the speed of different techniques for Markov Decision Processes (MDPs) in the context of solving mazes. Two basic algorithms for solving MDPs are value iteration and policy iteration.

The MDP selected is to solve mazes. There are 4 choices the agent in the maze can take viz. Up, Down, Left, Right. And the states are the positions in the maze. We are considering deterministic actions. The Reward is in the first block that is the (0,0) which is denoted by red. This has a reward of 10. The agent or the mouse gets in from the bottom right and needs to go up to the first block that has the reward or cheese.

Setup

There is a grid that has '#'s in them as walls and the other parts are the paths to the goal. We have policies that point to the next state we take. We need to come-up/improve a policy such that we reach the target reward state in the least number of steps that is without wandering around and repeating the states in the grid.

To visualize the algorithm there is a LOGO kind of a plugin turtle that is used to visualize actions that are taken at each state to understand what the policy looks like.

Policy Iteration

Starting with a random policy $\pi 0$, this approach consists of two steps:

I. Policy Evaluation

$$V(s) = \sum_{s' \in S} T(s, \pi(s), s')[R(s, \pi(s), s') + \gamma V(s')], \forall s \in S$$

II. Policy Improvement

$$\pi(s) \leftarrow \underset{a}{\operatorname{argmax}} \sum_{s' \in S} T(s,a,s')[R(s,a,s') + \gamma V(s')], \forall s \in S$$

These steps are repeated until π converges to π*

Value Iteration

The value iteration approach finds the optimal policy π* by calculating the optimal value function, V*. Starting with V(s) = 0 for all states s, the values of each state are iteratively updated to get the next value function V, which converges towards V*.

$$V_{i+1}^{*} \leftarrow \underset{a}{\operatorname{argmax}} \sum_{s' \in S} T(s,a,s')[R(s,a,s') + V_i^{*}(s')]$$

Files List:

AnimationFile.py - Has the turtle Animation to visualize the policy and it's updates

maze.py - This creates a solvable maze for the size given in the calling classes

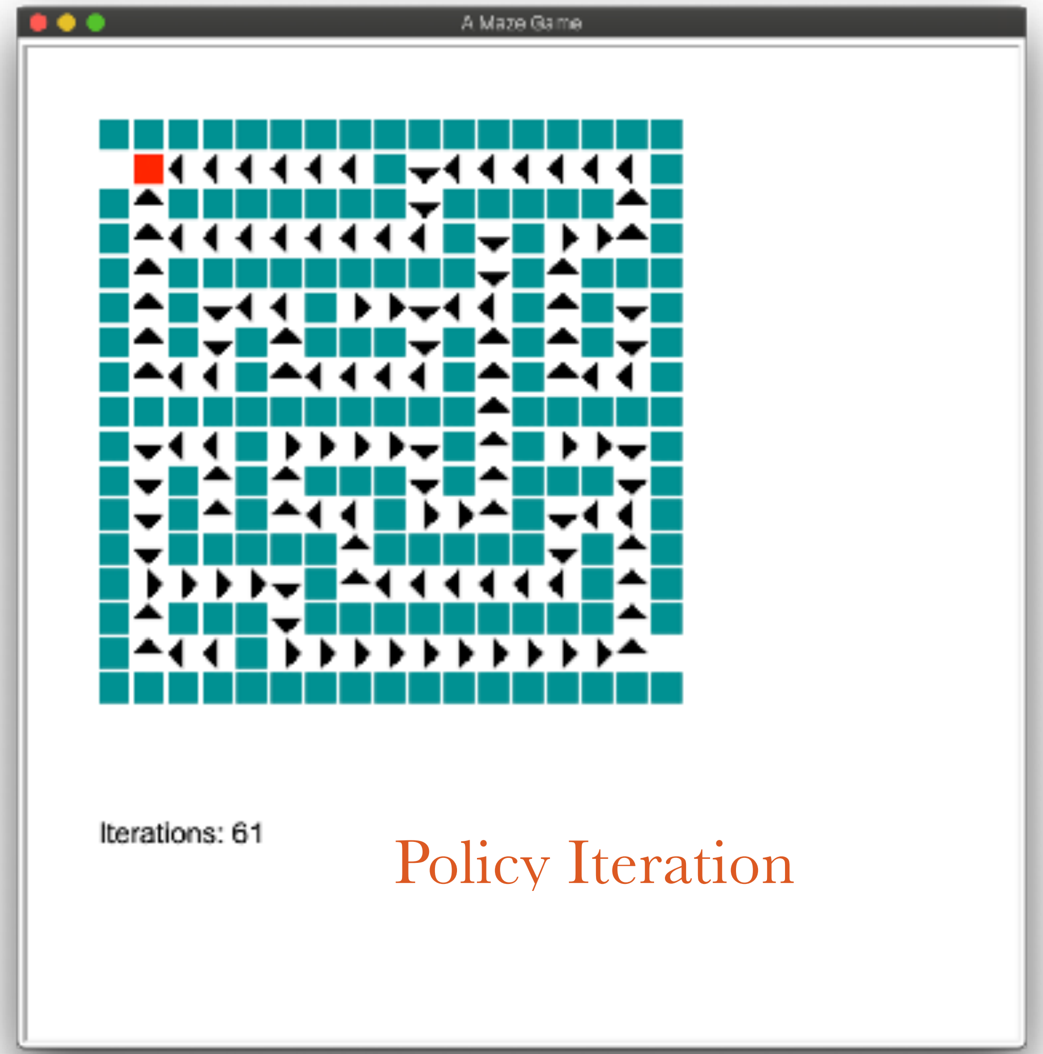
maze2mdp.py - Converts the maze matrix to an MDP problem with actions and states

mdpstate.py - The definition class to define the variables/parameters for MDP

policy_iteration.py - Actual code for PI having animations for the optimal policy.

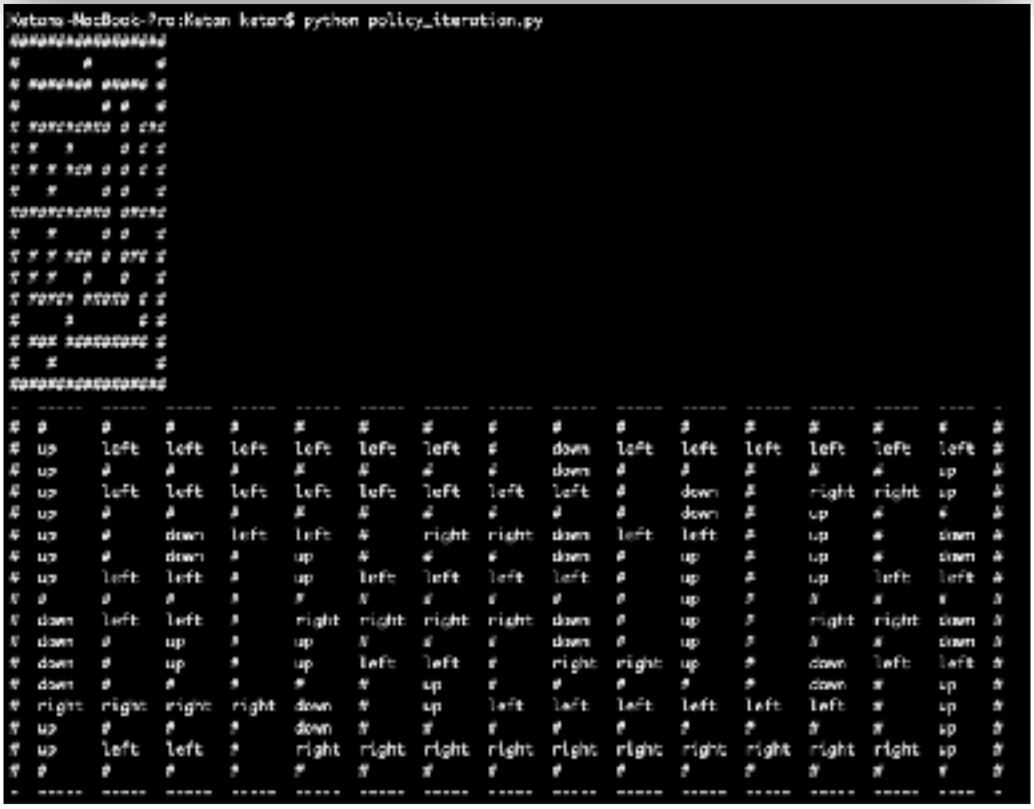
value_iteration.py - Code for VI without visualization that computes is less time.

vi_visualization.py - Code for VI with visualization and animation this takes way longer than the other version because it has to check the actions for all the stages to show it in the animation. This can be used to check the policy/(actions for each state) for every VI update for V(s) for the states in the matrix.

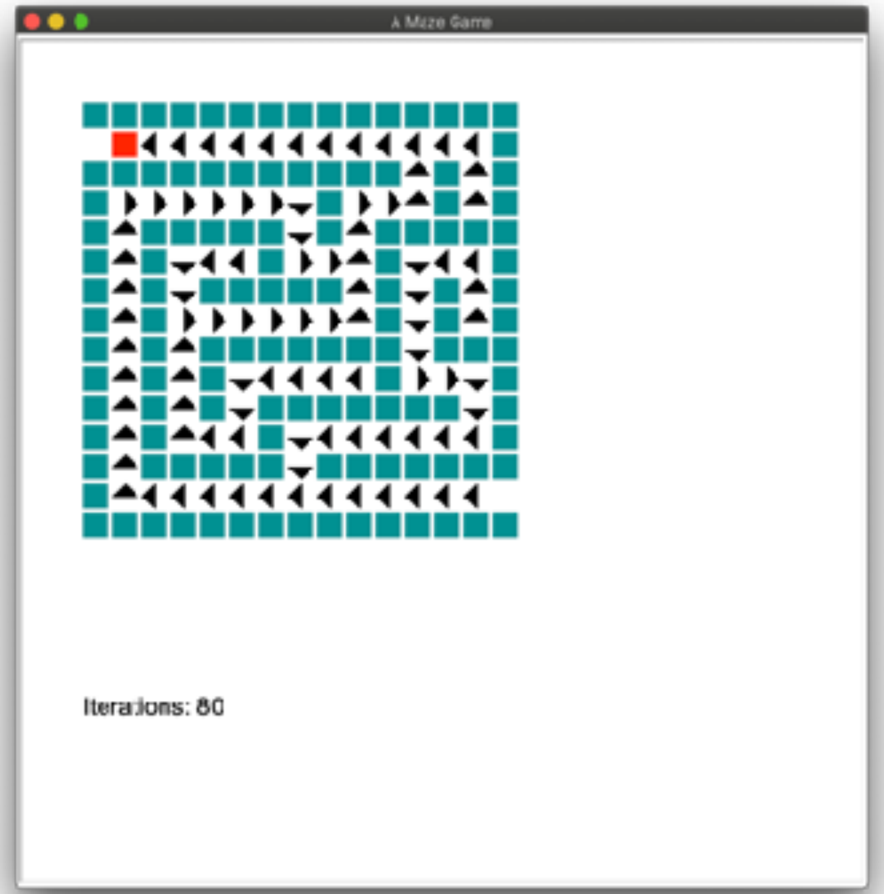


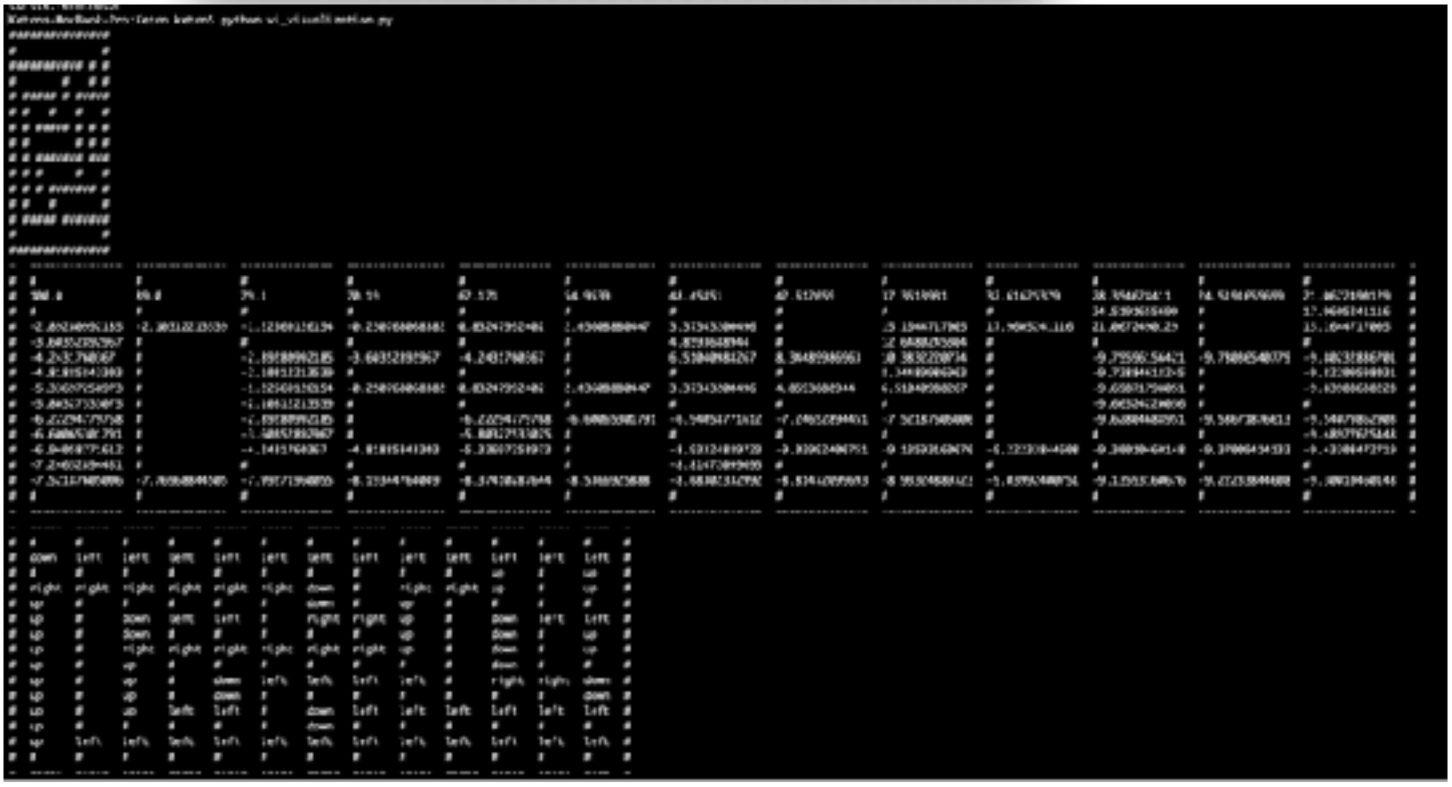
Iterations: 61

Policy Iteration



```
Ketans-MacBook-Pro:Ketan ketan$ python policy_iteration.py
#################
#       #       #
# ####### ##### #
#         # #   #
# ######### # ###
# #   #     # # #
# # # ### # # # #
#   #     # #   #
########### #####
#   #     # #   #
# # # ### # ### #
# # #   #   #   #
# ##### ##### # #
#     #       # #
# ### ######### #
#   #           #
#################
-  -----  -----  -----  -----  -----  -----  -----  -----  -----  -----  -----  -----  -----  -----  -----  -
#  #      #      #      #      #      #      #      #      #      #      #      #      #      #      #      #
#  up     left   left   left   left   left   left   #      down   left   left   left   left   left   left   #
#  up     #      #      #      #      #      #      #      down   #      #      #      #      #      up     #
#  up     left   left   left   left   left   left   left   left   #      down   #      right  right  up     #
#  up     #      #      #      #      #      #      #      #      #      down   #      up     #      #      #
#  up     #      down   left   left   #      right  right  down   left   left   #      up     #      down   #
#  up     #      down   #      up     #      #      #      down   #      up     #      up     #      down   #
#  up     left   left   #      up     left   left   left   left   #      up     #      up     left   left   #
#  #      #      #      #      #      #      #      #      #      #      up     #      #      #      #      #
#  down   left   left   #      right  right  right  right  down   #      up     #      right  right  down   #
#  down   #      up     #      up     #      #      #      down   #      up     #      #      #      down   #
#  down   #      up     #      up     left   left   #      right  right  up     #      down   left   left   #
#  down   #      #      #      #      #      up     #      #      #      #      #      down   #      up     #
#  right  right  right  right  down   #      up     left   left   left   left   left   left   #      up     #
#  up     #      #      #      down   #      #      #      #      #      #      #      #      #      up     #
#  up     left   left   #      right  right  right  right  right  right  right  right  right  right  up     #
#  #      #      #      #      #      #      #      #      #      #      #      #      #      #      #      #
-  -----  -----  -----  -----  -----  -----  -----  -----  -----  -----  -----  -----  -----  -----  -----  -
```





```
[illegible]
[illegible] python vi_visualization.py
###############
#             #
########### # #
#       #   # #
# ##### # #####
# #   #   #   #
# # ##### # # #
# #       # # #
# # ####### ###
# # #     #   #
# # # ####### #
# #   #       #
# ##### #######
#             #
###############
- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -
# #              #              #              #              #              #              #              #              #              #              #              #              #              #
# 100.0          89.0           79.1           70.19          62.171         54.9539        48.45291       42.512[illegible]  37.5610[illegible]  33.6162[illegible]  30.3544[illegible]  26.5216[illegible]  2[illegible].0672[illegible]  #
# #              #              #              #              #              #              #              #              #              #              34.5[illegible]        #              17.9[illegible]        #
# -2.8[illegible]  -2.3031[illegible]  -1.1230[illegible]  -0.2507[illegible]  0.8324[illegible]  2.4500[illegible]  3.3134[illegible]  #              25.15[illegible]  17.9[illegible]  21.0[illegible]  #              15.1[illegible]        #
# -3.6[illegible]  #              #              #              #              #              4.85[illegible]  #              12.6[illegible]  #              #              #              #              #
# -4.2[illegible]  #              -2.0[illegible]  -3.6035[illegible]  -4.2431[illegible]  #              6.51[illegible]  8.3[illegible]  10.38[illegible]  #              -9.7[illegible]  -9.7[illegible]  -9.1[illegible]  #
# -4.8[illegible]  #              -2.1[illegible]  #              #              #              #              #              1.3[illegible]  #              -9.7[illegible]  #              -9.1[illegible]  #
# -5.3[illegible]  #              -1.1[illegible]  -0.2507[illegible]  0.8324[illegible]  2.4500[illegible]  3.3134[illegible]  4.65[illegible]  6.51[illegible]  #              -9.6[illegible]  #              -9.1[illegible]  #
# -5.8[illegible]  #              -2.1[illegible]  #              #              #              #              #              #              #              -9.6[illegible]  #              #              #
# -6.2[illegible]  #              -2.8[illegible]  #              -6.2[illegible]  -6.6[illegible]  -6.9[illegible]  -7.0[illegible]  -7.5[illegible]  #              -9.6[illegible]  -9.5[illegible]  -9.5[illegible]  #
# -6.6[illegible]  #              -3.6[illegible]  #              -5.8[illegible]  #              #              #              #              #              #              #              -9.4[illegible]  #
# -6.9[illegible]  #              -4.1[illegible]  -4.8[illegible]  -5.3[illegible]  #              -4.6[illegible]  -9.3[illegible]  -9.1[illegible]  -6.2[illegible]  -9.3[illegible]  -9.3[illegible]  -9.4[illegible]  #
# -7.2[illegible]  #              #              #              #              #              -4.4[illegible]  #              #              #              #              #              #              #
# -7.5[illegible]  -7.7[illegible]  -7.9[illegible]  -8.2[illegible]  -8.3[illegible]  -8.5[illegible]  -8.6[illegible]  -8.8[illegible]  -8.9[illegible]  -5.8[illegible]  -9.1[illegible]  -9.2[illegible]  -9.3[illegible]  #
# #              #              #              #              #              #              #              #              #              #              #              #              #              #
- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -------------- -

- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -
# #     #     #     #     #     #     #     #     #     #     #     #     #     #
# down  left  left  left  left  left  left  left  left  left  left  left  left  #
# #     #     #     #     #     #     #     #     #     #     up    #     up    #
# right right right right right right down  #     right right up    #     up    #
# up    #     #     #     #     #     down  #     up    #     #     #     #     #
# up    #     down  left  left  #     right right up    #     down  left  left  #
# up    #     down  #     #     #     #     #     up    #     down  #     up    #
# up    #     right right right right right right up    #     down  #     up    #
# up    #     up    #     #     #     #     #     #     #     down  #     #     #
# up    #     up    #     down  left  left  left  left  #     right right down  #
# up    #     up    #     down  #     #     #     #     #     #     #     down  #
# up    #     up    left  left  #     down  left  left  left  left  left  left  #
# up    #     #     #     #     #     down  #     #     #     #     #     #     #
# up    left  left  left  left  left  left  left  left  left  left  left  left  #
# #     #     #     #     #     #     #     #     #     #     #     #     #     #
- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- ----- -
```