Artificial Intelligence – 2

Assignment-2

Name: Umang Barbhaya

Roll No: M20CS017

Policy Design

Reinforcement Learning the agent don't get any reward until the agent reaches any state and in simple way we can say agent gets the reward the once the action has been taken on the state. The agent chooses the action that leads to the maximum total reward. The agent don't have the information about the transition model and the reward function.

The utility of state depends on the reward of that state plus the utility of the successor states. This statement has been defined by the bellman's equation.

To make the policy dynamic and adaptive we consider the transition model and rewards to be known and place the same in bellman's equation to get the utility of the current state.

$\pi(s) = \max \hat{Q}\theta(s, a)$

Reward Function Design

The Agent will only take the shortest and proper path only if its reward function is properly given

We have given the reward for each state and since as per the concept of Reinforcement learning the agent will know the reward value when it has been landed into new state.

We have given the Goal state as the reward of +100

The empty white square has the reward of -0.4

The Wall has the reward of -9999

The Goto Power position has a value of +1

The Restart position has the value of -1

Initially the Agent knowledge base has value 0 and it gets updated with the Help of QLearning and Temporal Difference

Knowledge Base Format

In the output below have given the agent Knowledge Base and the Real World state value matrices are given

QLearning is used to update the agent Knowledge base

The Knowledge Base Matrix or QLearningMatrix is a 8*8*4 matrix where there are 8 rows and 8 columns of the Grid and in each state there can be at most 4 action (Up, down, right, left)

Below are the formula used in the code to update the Qlearning Values as the Agent moves to the new state.

epsilon, discount_factor, learning_rate = 0.9, 0.9, 0.9

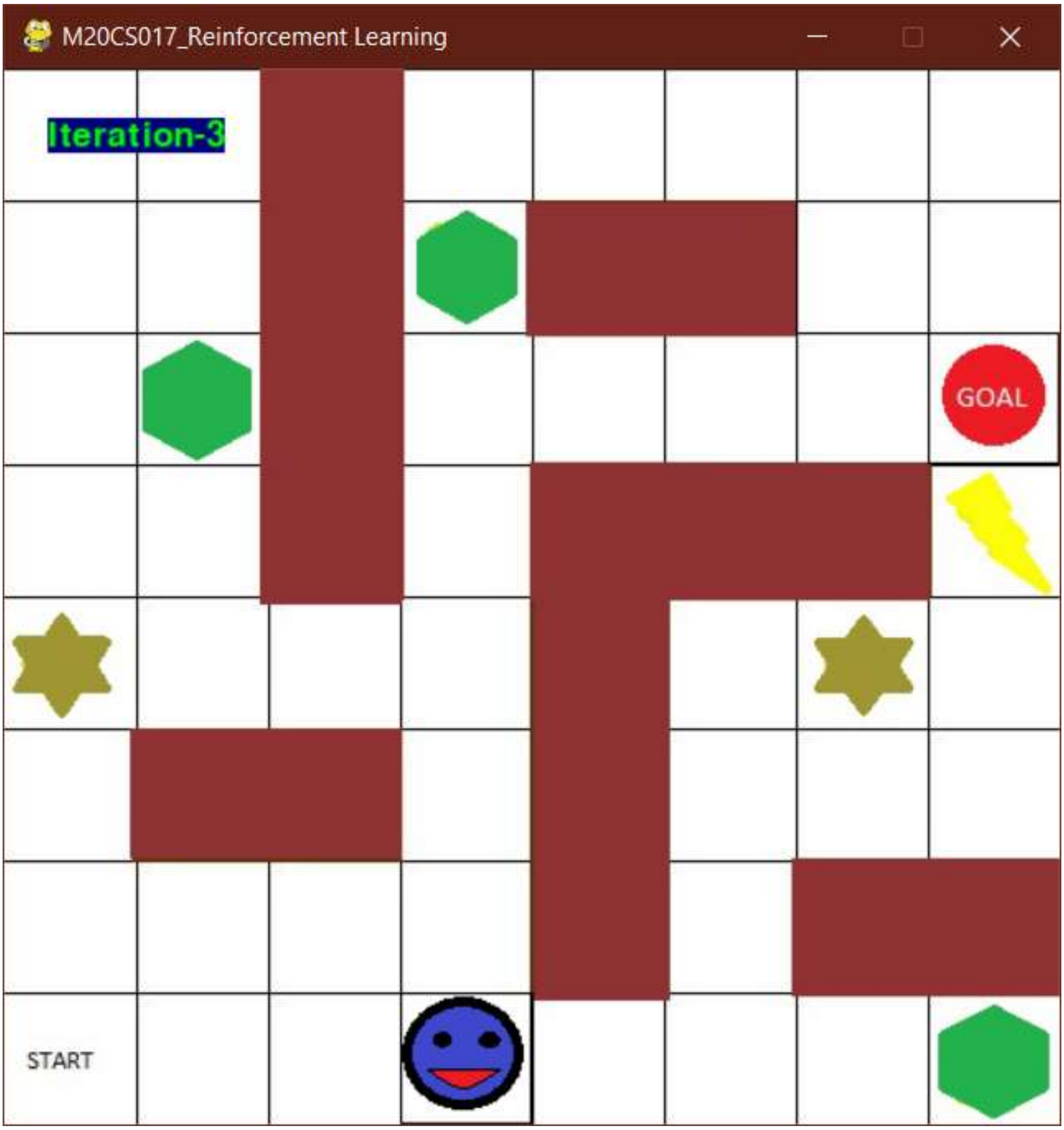
OldQLearningValue = QLearningMatrix[row_ith, column_jth, ACTION]

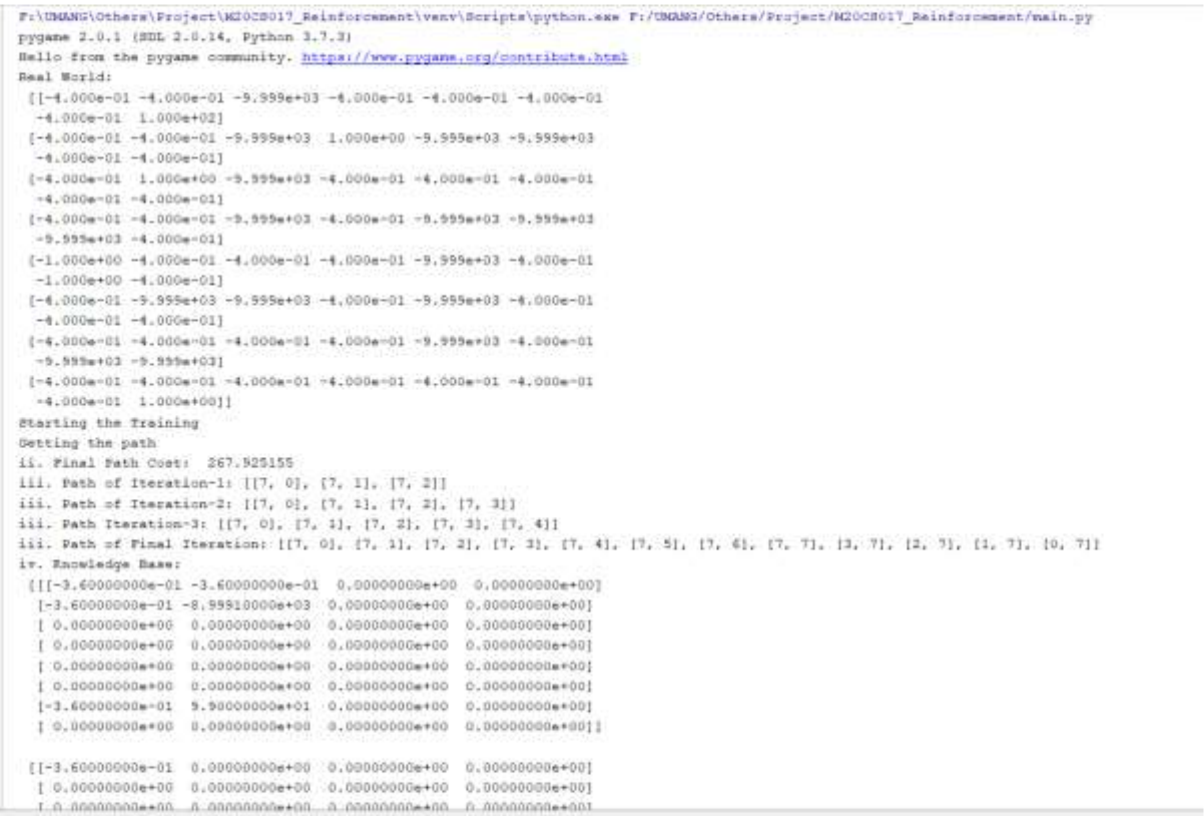
TD = reward + (discount_factor * np.max(QLearningMatrix[row_ith, column_jth])) - QLearningMatrix

NewQLearningValue = OldQLearningValue + (alpha * TD)

QLearningMatrix[row_ith, column_jth, ACTION] = new_q_value

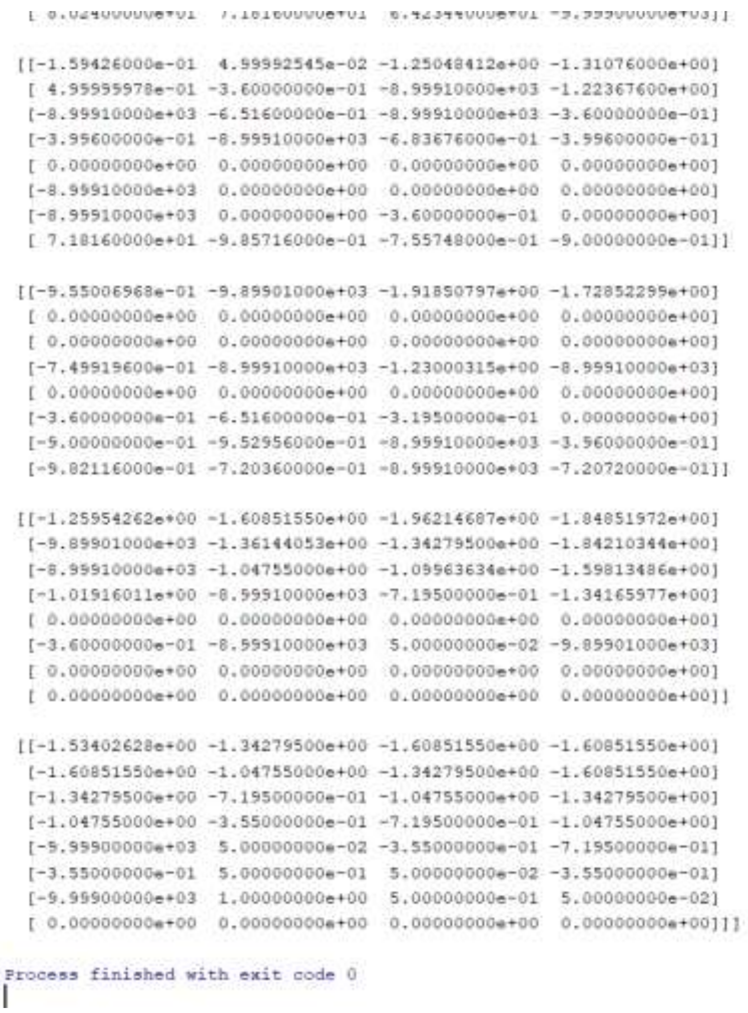
Output





```
F:\UMANG\Others\Project\M20CS017_Reinforcement\venv\Scripts\python.exe F:/UMANG/Others/Project/M20CS017_Reinforcement/main.py
pygame 2.0.1 (SDL 2.0.14, Python 3.7.3)
Hello from the pygame community. https://www.pygame.org/contribute.html
Real World:
 [[-4.000e-01 -4.000e-01 -9.999e+03 -4.000e-01 -4.000e-01 -4.000e-01
  -4.000e-01  1.000e+02]
 [-4.000e-01 -4.000e-01 -9.999e+03  1.000e+00 -9.999e+03 -9.999e+03
  -4.000e-01 -4.000e-01]
 [-4.000e-01  1.000e+00 -9.999e+03 -4.000e-01 -4.000e-01 -4.000e-01
  -4.000e-01 -4.000e-01]
 [-4.000e-01 -4.000e-01 -9.999e+03 -4.000e-01 -9.999e+03 -9.999e+03
  -9.999e+03 -4.000e-01]
 [-1.000e+00 -4.000e-01 -4.000e-01 -4.000e-01 -9.999e+03 -4.000e-01
  -1.000e+00 -4.000e-01]
 [-4.000e-01 -9.999e+03 -9.999e+03 -4.000e-01 -9.999e+03 -4.000e-01
  -4.000e-01 -4.000e-01]
 [-4.000e-01 -4.000e-01 -4.000e-01 -4.000e-01 -9.999e+03 -4.000e-01
  -9.999e+03 -9.999e+03]
 [-4.000e-01 -4.000e-01 -4.000e-01 -4.000e-01 -4.000e-01 -4.000e-01
  -4.000e-01  1.000e+00]]
Starting the Training
Getting the path
ii. Final Path Cost:  267.925155
iii. Path of Iteration-1: [[7, 0], [7, 1], [7, 2]]
iii. Path of Iteration-2: [[7, 0], [7, 1], [7, 2], [7, 3]]
iii. Path Iteration-3: [[7, 0], [7, 1], [7, 2], [7, 3], [7, 4]]
iii. Path of Final Iteration: [[7, 0], [7, 1], [7, 2], [7, 3], [7, 4], [7, 5], [7, 6], [7, 7], [3, 7], [2, 7], [1, 7], [0, 7]]
iv. Knowledge Base:
 [[[-3.60000000e-01 -3.60000000e-01  0.00000000e+00  0.00000000e+00]
  [-3.60000000e-01 -8.99910000e+03  0.00000000e+00  0.00000000e+00]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
  [-3.60000000e-01  5.90000000e+01  0.00000000e+00  0.00000000e+00]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]]

 [[-3.60000000e-01  0.00000000e+00  0.00000000e+00  0.00000000e+00]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
```



```
  [ 6.02400000e+01  7.18160000e+01  6.42344000e+01 -9.99900000e+03]]

 [[-1.59426000e-01  4.99992545e-02 -1.25048412e+00 -1.31076000e+00]
  [ 4.99999978e-01 -3.60000000e-01 -8.99910000e+03 -1.22367600e+00]
  [-8.99910000e+03 -6.51600000e-01 -8.99910000e+03 -3.60000000e-01]
  [-3.99600000e-01 -8.99910000e+03 -6.83676000e-01 -3.99600000e-01]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
  [-8.99910000e+03  0.00000000e+00  0.00000000e+00  0.00000000e+00]
  [-8.99910000e+03  0.00000000e+00 -3.60000000e-01  0.00000000e+00]
  [ 7.18160000e+01 -9.85716000e-01 -7.55748000e-01 -9.00000000e-01]]

 [[-9.55006968e-01 -9.89901000e+03 -1.91850797e+00 -1.72852299e+00]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
  [-7.49919600e-01 -8.99910000e+03 -1.23000315e+00 -8.99910000e+03]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
  [-3.60000000e-01 -6.51600000e-01 -3.19500000e-01  0.00000000e+00]
  [-9.00000000e-01 -9.52956000e-01 -8.99910000e+03 -3.96000000e-01]
  [-9.82116000e-01 -7.20360000e-01 -8.99910000e+03 -7.20720000e-01]]

 [[-1.25954262e+00 -1.60851550e+00 -1.96214687e+00 -1.84851972e+00]
  [-9.89901000e+03 -1.36144053e+00 -1.34279500e+00 -1.84210344e+00]
  [-8.99910000e+03 -1.04755000e+00 -1.09963634e+00 -1.59813486e+00]
  [-1.01916011e+00 -8.99910000e+03 -7.19500000e-01 -1.34165977e+00]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
  [-3.60000000e-01 -8.99910000e+03  5.00000000e-02 -9.89901000e+03]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]]

 [[-1.53402628e+00 -1.34279500e+00 -1.60851550e+00 -1.60851550e+00]
  [-1.60851550e+00 -1.04755000e+00 -1.34279500e+00 -1.60851550e+00]
  [-1.34279500e+00 -7.19500000e-01 -1.04755000e+00 -1.34279500e+00]
  [-1.04755000e+00 -3.55000000e-01 -7.19500000e-01 -1.04755000e+00]
  [-9.99900000e+03  5.00000000e-02 -3.55000000e-01 -7.19500000e-01]
  [-3.55000000e-01  5.00000000e-01  5.00000000e-02 -3.55000000e-01]
  [-9.99900000e+03  1.00000000e+00  5.00000000e-01  5.00000000e-02]
  [ 0.00000000e+00  0.00000000e+00  0.00000000e+00  0.00000000e+00]]]

Process finished with exit code 0
```