

NANYANG
TECHNOLOGICAL
UNIVERSITY

CZ3005: Artificial Intelligence
Reinforcement Learning

Lab 2

Submitted By:
Royce Ang Jia Jie
U1840416D
TSP2

Problem Statement

We are to solve/ find the optimal value for the given MDP formulation:

- States: a 3D coordinate, which indicates the current position where the agent is. The initial state is (0, 0, 0) and there is only one terminal state: (3,3,3).
- Action: The action space is (forward, backward, left, right, up, down). The agent needs to select one of them to navigate in the environment.
- Reward: The agent will receive 1 reward when it arrives at the terminal states, or otherwise receive -0.1 reward.
- Transition: The intended movement happens with probability 0.6. With probability 0.1, the agent ends up in one of the states perpendicular to the intended direction. The happened movement "forward"/"backward" will add/subtract one to the 1st element of state.

Key Features of Q-Learning

To solve the given MDP formulation, I have decided to use Q-Learning.

- Q-Learning is a values-based learning algorithm.
- Q-Learning is an off-policy learner. Meaning it learns the value of the optimal policy independent of the agent's actions.
- $Q^*(s,a)$, where, s , represents state, and, a , representing action, is the expected value(cumulative discounted reward) of doing a in state s and then following some optimal policy.
- Q-Learning estimates Q^* out of the best Q values (as shown in given diagram below).

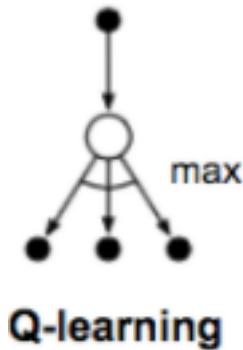


Figure 1 Q-Learning selecting the best Q values. [1]

- Q-Learning uses Temporal Difference (TD) to estimate the value of $Q^*(s,a)$. Note: *TD is an agent learning from an environment through episodes with no prior knowledge of the environment.*
- Definition of Temporal Difference:

$$TD(a, s) = R(s, a) + \gamma \max_{a'} Q(s', a') - Q_{t-1}(s, a)$$

Figure 2 Temporal Difference

Implementation

This will be how our Q-learning Algorithm process will look like:

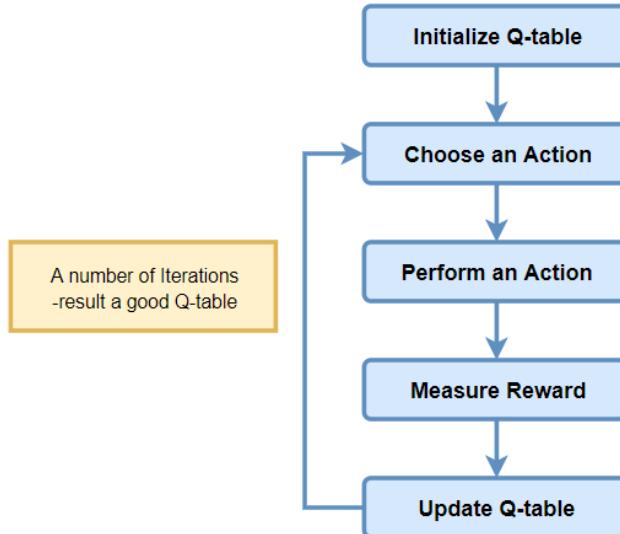


Figure 3 Q-Learning Process

Limitations Involved:

- Gamma (Discount Factor) = 0.99
- Alpha (Learning-Rate) = 0.5
- Epsilon (Exploration rate) = 0.01

Step 1: Initialize the Q-Table:

Given that, we have a 3x3x3 state, each with a total of 6 different actions (forward, backward, left, right, up, down).

1.1: `def createAllPossibleStates(self):` (Agent Class)

```

def createAllPossibleStates(self):
    """
    This creates all 64 permutations of states from [0,0,0] to [3,3,3]
    """
    possibleStates = []
    for i in range(4):
        for j in range(4):
            for k in range(4):
                possibleStates.append((i,j,k))
    return possibleStates
  
```

1.2: Initialize Q-Table in a Dictionary format to produce a Q(s,a).

```

# set all Q-values
Q = {}
#states = grid.all_states()
for s in possible_states:
    Q[s] = {}
    for a in agent.action_space:
        Q[s][a] = 0 # set initial Q-value to 0
print(Q)
agent.setQ(Q) #sets Q for easier computation
  
```

Step 2: Choosing an action

Epsilon greedy strategy concept comes into play here. The agent takes greedy action for probability $< (1 - \epsilon)$, else a random action will be taken from the given action space. Note: $\epsilon = 0.01$.

2.1: def random_greedy_action(self, state_key, eps=0.01)

```
def random_greedy_action(self, state_key, eps=0.01):
    """
    Note: greedy_epsilon function also note that state is an array
    The agent takes random actions for probability  $\epsilon$  and greedy action for probability  $(1-\epsilon)$ .
    """
    a,_ = self.max_dict(state_key)
    p = np.random.random()
    if p < (1 - eps):
        return a
    else:
        return np.random.choice(self.action_space)
```

Step 3: Perform an action

Here, we would make use of the given environment method, def step(action), to return us the *reward*, *terminate* and *next_state* values.

```
while not terminate:
    #action = agent.take_action(state)
    action = agent.random_greedy_action(Q[state], eps=0.01)
    reward, terminate, next_state = env.step(action)
```

Step 4: Measuring the reward

Every action taken would produce an observed outcome and reward. As such, the returned result would then be added to our overall *episode_reward* bank, using a simple cumulative method.

```
episode_reward += reward
# you can comment the following two lines, if the output is too much
env.render() # comment
print(f'step: {t}, action: {action}, reward: {reward}') # comment
t += 1
```

Step 5: Evaluate

We need to update the function $Q(s,a)$, via the following equation to give us the New $Q(s,a)$.

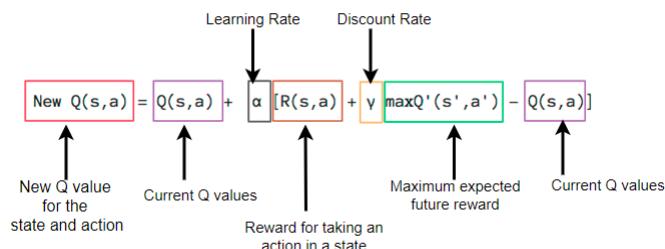


Figure 4 Equation Explanation

5.1: `def max_dict(self, d)`: this method would be used to return the maximum expected future reward.

```
def max_dict(self,d):
    # returns the argmax (key) and max (value) from a dictionary
    # put this into a function since we are using it so often
    max_key = None
    max_val = float('-inf')
    for k, v in d.items():
        if v > max_val:
            max_val = v
            max_key = k
    return max_key, max_val
```

5.2: `def train(self, state, action, next_state, reward)`: this is a direct implementation of the given formula stated above to obtain the new Q value.

```
def train(self, state, action, next_state, reward):
    """
    Note: state == current state e.g. [0,0,0]
    """
    ALPHA = 0.5 #learning rate
    GAMMA = 0.99 #discount factor
    #implement the normal q_algorithm here
    old_qsa = self.Q[state][action]
    a2, max_q_s2a2 = self.max_dict(self.Q[next_state])
    self.Q[state][action] = old_qsa + ALPHA*(reward + GAMMA*max_q_s2a2 - old_qsa)
```

Additional Implementation

Additional 1: `def print_values(self, V)`: this would give a proper UI for printing Q-Values.

```
def print_values(self, V):
    for i in range(4):
        print("-----")
        for j in range(4):
            for k in range(4):
                v = V.get((i,j,k), 0)
                if v >= 0:
                    print(" %.4f|" % v, end="")
                else:
                    print("%.4f|" % v, end="") # -ve sign takes up an extra space
    print("")
```

Additional 2: `def print_policy(self, P)`: this would give a proper UI for printing the best policies.

```
def print_policy(self,P):
    for i in range(4):
        print("-----")
        for j in range(4):
            for k in range(4):
                if(i == 3 and j == 3 and k == 3):
                    a = 'GOAL'
                else:
                    a = P.get((i,j,k), ' ')
                print(" %s |" % a, end="")
        print("")
```

Additional 3: print method to denote the proportional of time spent on updating each part of Q:

```

        print("----Proportion of time spent updating each part of Q-----")
        total = np.sum(list(update_counts.values()))
        for k,v in update_counts.items():
            update_counts[k] = float(v)/total
        agent.print_values(update_counts)
    
```

Outcome/ Overall Results

In order to properly evaluate the final results, I have implemented 4 different print statements

1. *Q-Table Full*
2. *Q-Table-Max-Value*
3. *Policy Table*
4. *Proportion of time spent updating each part of Q);*
5. *Plot episode rewards vs episodes.*

These are evaluated against --max_episode 500 & --max_step 500.

Note: Q-Table-Max-Value and Policy Tables are states ordered in rows from (0,0,0) → (0,0,3), etc. with each value representing the max of values for each of the states respectively.

Case 1 (Best-Case)

Total Steps: 9

```

step: 0, action: up, reward: 1
episode: 500, total_steps: 9 episode reward: 0.20000000000000007

-Q-Table-FULL-
(0, 0, 0) {'left': -0.451421741474582, 'right': -0.5096692529127174, 'forward': -0.8144452838184857, 'backward': -0.7980235947928812, 'up': -0.7616488775689985, 'down': -0.8038198900878262}
(0, 0, 1) {'left': -0.7340549191256498, 'right': -0.53932692125748378, 'backward': -0.72815791915624411, 'up': -0.726859370788191, 'down': -0.732889586429643}
(0, 0, 2) {'left': -0.630897661331322, 'right': -0.6831318946865731, 'forward': -0.6013218385840833, 'backward': -0.5837719388652439, 'up': -0.34794264227259844, 'down': -0.4681203638183367}
(0, 0, 3) {'left': -0.63222884397545122, 'right': -0.43408515262524627, 'forward': -0.89587727285799259, 'backward': -0.48598781741764534, 'up': -0.440821687199508384, 'down': -0.49699697908843803}
(0, 1, 0) {'left': -0.63222884397545122, 'right': -0.3984983865382219, 'forward': -0.50007705887467376, 'backward': -0.605758986029966, 'up': -0.541993586536422, 'down': -0.5919527959593631}
(0, 1, 1) {'left': -0.63222884397545122, 'right': -0.23629773833994985, 'forward': -0.5407259316500654, 'backward': -0.4966348383538808, 'up': -0.4881141327847368, 'down': -0.5323618840363393}
(0, 1, 2) {'left': -0.63222884397545122, 'right': -0.188644777383399498, 'forward': -0.5407259316500654, 'backward': -0.4966348383538808, 'up': -0.4881141327847368, 'down': -0.5323618840363393}
(0, 1, 3) {'left': -0.373392115018722, 'right': -0.3358803788478465, 'forward': -0.348665780414823, 'backward': -0.3794686112727084, 'up': -0.397236876878647, 'down': -0.3277467872588673}
(0, 2, 0) {'left': -0.43167408568958932, 'right': -0.15746227658866932, 'forward': -0.4281874646464625, 'backward': -0.46377194127367243, 'up': -0.45986488745312423}
(0, 2, 1) {'left': -0.43167408568958932, 'right': -0.15746227658866932, 'forward': -0.428805212624384, 'backward': -0.399549852114463784, 'up': -0.399549852114463784, 'down': -0.4424258835772447}
(0, 2, 2) {'left': -0.26977515955775864, 'right': -0.2426664049461376, 'forward': -0.24751246878152, 'backward': -0.24751246878152, 'up': -0.2825684249103766}
(0, 2, 3) {'left': -0.23832293930585288, 'right': -0.2426664049461376, 'forward': -0.11759736568754985, 'backward': -0.24751246878152, 'up': -0.2767219524651965, 'down': -0.2424276239322124}
(0, 3, 0) {'left': -0.3990414768972083, 'right': -0.33469922857306333, 'forward': -0.07542419889717089, 'backward': -0.3983646950875633, 'up': -0.3566133316792346, 'down': -0.24994656452864705}
(0, 3, 1) {'left': -0.3990414768972083, 'right': -0.33469922857306333, 'forward': -0.07542419889717089, 'backward': -0.3983646950875633, 'up': -0.3566133316792346, 'down': -0.24994656452864705}
(0, 3, 2) {'left': -0.165514895142692, 'right': -0.108804935765498, 'forward': -0.165514895142692, 'backward': -0.248412068712524087, 'up': -0.108804935765498, 'down': -0.258862003493857}
(0, 3, 3) {'left': -0.581497903713374, 'right': -0.6447700818425658, 'forward': -0.581497903713374, 'backward': -0.518568734365249, 'up': -0.5888031289519881, 'down': -0.6952103276569298}
(1, 0, 0) {'left': -0.581497903713374, 'right': -0.5077533823286414, 'forward': -0.581497903713374, 'backward': -0.518568734365249, 'up': -0.5888031289519881, 'down': -0.581497903713374}
(1, 0, 1) {'left': -0.362862636399862, 'right': -0.4083159378467497, 'forward': -0.39271133965910837, 'backward': -0.415108473785742, 'up': -0.2037576277796821, 'down': -0.3351092124249683}
(1, 0, 2) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(1, 0, 3) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(1, 1, 0) {'left': -0.47585657545519053, 'right': -0.3120494741764023, 'forward': -0.4642364161436588, 'backward': -0.48664994964986265, 'up': -0.471569387313687, 'down': -0.5260376968240243}
(1, 1, 1) {'left': -0.4499517753325257, 'right': -0.4699517753325257, 'forward': -0.4699517753325257, 'backward': -0.48979445257658835, 'up': -0.4787857484084876}
(1, 1, 2) {'left': -0.4499517753325257, 'right': -0.4499517753325257, 'forward': -0.4499517753325257, 'backward': -0.472651622210557, 'up': -0.4787857484084876}
(1, 1, 3) {'left': -0.4499517753325257, 'right': -0.4499517753325257, 'forward': -0.4499517753325257, 'backward': -0.472651622210557, 'up': -0.4787857484084876}
(1, 2, 0) {'left': -0.3564454242654262, 'right': -0.35992379844983, 'forward': -0.3564454242654262, 'backward': -0.35992379844983, 'up': -0.3744877456808319, 'down': -0.3744877456808319}
(1, 2, 1) {'left': -0.3564454242654262, 'right': -0.35992379844983, 'forward': -0.3564454242654262, 'backward': -0.35992379844983, 'up': -0.3744877456808319}
(1, 2, 2) {'left': -0.3564454242654262, 'right': -0.35992379844983, 'forward': -0.3564454242654262, 'backward': -0.35992379844983, 'up': -0.3744877456808319}
(1, 2, 3) {'left': -0.3564454242654262, 'right': -0.35992379844983, 'forward': -0.3564454242654262, 'backward': -0.35992379844983, 'up': -0.3744877456808319}
(1, 3, 0) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26288686335786743}
(1, 3, 1) {'left': -0.47585657545519053, 'right': -0.3120494741764023, 'forward': -0.4642364161436588, 'backward': -0.48664994964986265, 'up': -0.471569387313687, 'down': -0.5260376968240243}
(1, 3, 2) {'left': -0.4499517753325257, 'right': -0.4499517753325257, 'forward': -0.4499517753325257, 'backward': -0.48979445257658835, 'up': -0.4787857484084876}
(1, 3, 3) {'left': -0.4499517753325257, 'right': -0.4499517753325257, 'forward': -0.4499517753325257, 'backward': -0.472651622210557, 'up': -0.4787857484084876}
(2, 0, 0) {'left': -0.165514895142692, 'right': -0.108804935765498, 'forward': -0.165514895142692, 'backward': -0.248412068712524087, 'up': -0.108804935765498, 'down': -0.258862003493857}
(2, 0, 1) {'left': -0.118516273519851794, 'right': -0.120216273519851794, 'forward': -0.118516273519851794, 'backward': -0.120216273519851794, 'up': -0.120216273519851794}
(2, 0, 2) {'left': -0.118516273519851794, 'right': -0.120216273519851794, 'forward': -0.118516273519851794, 'backward': -0.120216273519851794, 'up': -0.120216273519851794}
(2, 0, 3) {'left': -0.118516273519851794, 'right': -0.120216273519851794, 'forward': -0.118516273519851794, 'backward': -0.120216273519851794, 'up': -0.120216273519851794}
(2, 1, 0) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(2, 1, 1) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(2, 1, 2) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(2, 1, 3) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(2, 2, 0) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(2, 2, 1) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(2, 2, 2) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(2, 2, 3) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(2, 3, 0) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(2, 3, 1) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(2, 3, 2) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(2, 3, 3) {'left': -0.11225000000000002, 'right': -0.11225000000000002, 'forward': -0.11225000000000002, 'backward': -0.11225000000000002, 'up': -0.11225000000000002, 'down': -0.11225000000000002}
(3, 0, 0) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 0, 1) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 0, 2) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 0, 3) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 1, 0) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 1, 1) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 1, 2) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 1, 3) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 2, 0) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 2, 1) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 2, 2) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 2, 3) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 3, 0) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 3, 1) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 3, 2) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(3, 3, 3) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 0, 0) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 0, 1) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 0, 2) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 0, 3) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 1, 0) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 1, 1) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 1, 2) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 1, 3) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 2, 0) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 2, 1) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 2, 2) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 2, 3) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 3, 0) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3351092124249683, 'backward': -0.3351092124249683, 'up': -0.26884866335786743}
(4, 3, 1) {'left': -0.3351092124249683, 'right': -0.3351092124249683, 'forward': -0.3
```

-----Q-Table-Values-----				
-0.5097	-0.5393	-0.3497	-0.0959	
-0.3985	-0.2360	-0.1331	-0.3136	
-0.2982	-0.1575	0.0427	-0.1176	
-0.2134	0.0754	0.2358	-0.1555	
<hr/>				
-0.2478	-0.2298	-0.2037	0.1109	
-0.3120	0.0022	0.0738	0.3253	
-0.0942	-0.1638	0.2818	0.4879	
0.0457	0.4011	0.6539	0.2655	
<hr/>				
-0.0514	0.0862	-0.1379	0.2028	
0.0347	0.2440	0.0941	0.4931	
0.2210	0.3334	0.4973	0.6789	
0.2799	0.5951	0.5226	0.4036	
<hr/>				
0.1552	0.2409	0.2521	-0.3311	
0.3095	0.4910	0.5408	0.8121	
0.3333	0.3846	0.7313	0.9927	
0.6097	0.7289	0.9382	0.0000	

-----Policy-Table-----				
right	forward	up	forward	
right	right	right	right	
right	right	right	forward	
up	forward	forward	left	
<hr/>				
forward	forward	up	right	
right	forward	right	forward	
forward	left	forward	forward	
forward	forward	forward	down	
<hr/>				
forward	forward	up	right	
right	forward	forward	right	
forward	up	right	forward	
forward	forward	forward	down	
<hr/>				
right	right	backward	right	
right	up	right	right	
up	up	up	right	
up	up	up	GOAL	

-----Proportion of time spent updating each part of Q-----				
0.0718	0.0271	0.0161	0.0114	
0.0371	0.0280	0.0180	0.0126	
0.0142	0.0181	0.0125	0.0087	
0.0126	0.0182	0.0106	0.0070	
<hr/>				
0.0228	0.0134	0.0077	0.0089	
0.0158	0.0140	0.0085	0.0141	
0.0111	0.0182	0.0168	0.0174	
0.0078	0.0170	0.0125	0.0087	
<hr/>				
0.0158	0.0105	0.0065	0.0062	
0.0107	0.0077	0.0057	0.0092	
0.0139	0.0166	0.0181	0.0260	
0.0090	0.0244	0.0321	0.0087	
<hr/>				
0.0082	0.0082	0.0056	0.0042	
0.0077	0.0096	0.0103	0.0037	
0.0130	0.0212	0.0287	0.0434	
0.0069	0.0267	0.0430	0.0000	
<hr/>				

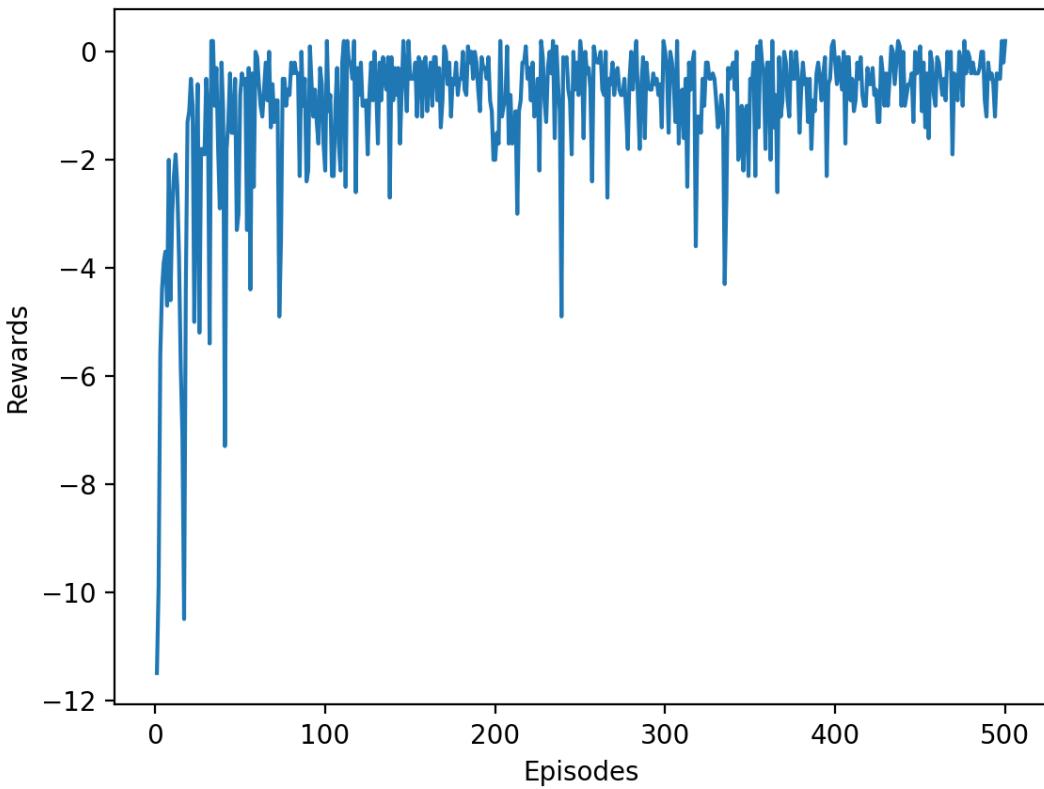


Figure 5 Step_9 episode_rewards vs episodes

Case 2:

Total Steps: 10 version1

step: 0, action: up, reward: 1
episode: 500, total_steps: 10 episode_reward: 0.10000000000000009

Q-Table-FULL	
(0, 0, 0) ('left': -0.264559933318865, 'right': -0.47678566938582644, 'forward': -0.416105214959284, 'backward': -0.4460165134959284, 'up': -0.4303382409836288, 'down': -0.666403095874834)	
(0, 0, 1) ('left': -0.264559933318865, 'right': -0.47678566938582644, 'forward': -0.416105214959284, 'backward': -0.4460165134959284, 'up': -0.4303382409836288, 'down': -0.580923452452079)	
(0, 0, 2) ('left': -0.6717004454627745, 'right': -0.312958564844548, 'forward': -0.5257556058631459, 'backward': -0.425628281749813, 'up': -0.4567939425992877, 'down': -0.615934831321469)	
(0, 0, 3) ('left': -0.3659239456829279, 'right': -0.1015167589653207, 'forward': -0.3422087356238262, 'backward': -0.347935190515765, 'up': -0.35091946628464323, 'down': -0.3456421360999841623)	
(0, 1, 0) ('left': -0.871831387739769, 'right': -0.315842497783922, 'forward': -0.58918559794637378, 'backward': -0.58776346946331752, 'up': -0.5872484445625181, 'down': -0.587339862521407)	
(0, 1, 1) ('left': -0.32823233843313576, 'right': -0.32823233843313576, 'forward': -0.331656484269125, 'backward': -0.30739339324561519, 'up': -0.008778398130133881, 'down': -0.3363342497224467)	
(0, 1, 2) ('left': -0.324598887146437, 'right': 0.267478848049132, 'forward': -0.287568108104623, 'backward': -0.31879108383248423, 'up': -0.28380877343831568, 'down': -0.2947383089678127)	
(0, 1, 3) ('left': -0.4151664437764825, 'right': -0.4047953198515786, 'forward': -0.3282074925726447, 'backward': -0.432304974054574, 'up': -0.26708293295474, 'down': -0.398778398130133881)	
(0, 1, 4) ('left': -0.25562823955637463, 'right': -0.177353317264, 'forward': -0.177353317264, 'backward': -0.177353317264, 'up': -0.18105644451854, 'down': -0.398778398130133881)	
(0, 2, 0) ('left': -0.25562823955637463, 'right': -0.2582083140288841, 'forward': -0.22986120943750883, 'backward': -0.2378674298585582, 'up': 0.31316164647608317, 'down': -0.23669624245898511)	
(0, 2, 1) ('left': -0.2262089484254792, 'right': -0.2142108089988482, 'forward': -0.2262089484254792, 'backward': -0.24383513388492, 'up': -0.23172531385654371)	
(0, 2, 2) ('left': -0.2262089484254792, 'right': -0.2262089484254792, 'forward': -0.2262089484254792, 'backward': -0.24383513388492, 'up': -0.23172531385654371)	
(0, 2, 3) ('left': -0.322533843367338, 'right': -0.3446545785558816, 'forward': 0.1534294897823777, 'backward': -0.328421319771582, 'up': -0.3124227704718787, 'down': -0.3449458151598947)	
(0, 3, 0) ('left': -0.28162993750800001, 'right': -0.19858499275, 'forward': -0.28468784809798764, 'backward': -0.1951182562821285)	
(0, 3, 1) ('left': -0.19858499275, 'right': -0.19858499275, 'forward': 0.5664773589751409, 'backward': -0.14925125, 'up': -0.14925125, 'down': -0.189997678125)	
(0, 3, 2) ('left': -0.4151664437764825, 'right': -0.3578897045582676, 'forward': -0.449129383885454, 'backward': -0.3922747768335798, 'up': -0.4089213088226895)	
(0, 3, 3) ('left': -0.3398308317428544, 'right': -0.3578897045582676, 'forward': -0.3578897045582676, 'backward': -0.334777257264476, 'up': -0.2998181097177883)	
(0, 3, 4) ('left': -0.4047953198515786, 'right': -0.3578897045582676, 'forward': -0.3578897045582676, 'backward': -0.334777257264476, 'up': -0.2998181097177883)	
(0, 3, 5) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 1, 1) ('left': -0.335542169931118, 'right': -0.3446545785558816, 'forward': -0.3194374461675641, 'backward': -0.3165980978114394, 'up': -0.340532972944029)	
(0, 1, 2) ('left': -0.25562823955637463, 'right': -0.22884384389646879, 'forward': -0.22884384389646879, 'backward': -0.24383513388492, 'up': -0.2353629218666466983)	
(0, 1, 3) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 2, 0) ('left': -0.4437721744658684, 'right': -0.424664516353245, 'forward': -0.424664516353245, 'backward': -0.373764542815771, 'up': -0.4399521247475182, 'down': -0.4399521247475182)	
(0, 2, 1) ('left': -0.288687721744658684, 'right': -0.288687721744658684, 'forward': -0.288687721744658684, 'backward': -0.278042120111645, 'up': -0.2662237112989896)	
(0, 2, 2) ('left': -0.2367510984384833, 'right': -0.2367510984384833, 'forward': -0.19858499275, 'backward': -0.19858499275, 'up': -0.19858499275, 'down': -0.18547803125)	
(0, 2, 3) ('left': -0.3374361684438957, 'right': -0.3374361684438957, 'forward': -0.3374361684438957, 'backward': -0.3374361684438957, 'up': -0.3374361684438957, 'down': -0.3374361684438957)	
(0, 3, 0) ('left': -0.3374361684438957, 'right': -0.3374361684438957, 'forward': -0.3374361684438957, 'backward': -0.3374361684438957, 'up': -0.3374361684438957, 'down': -0.3374361684438957)	
(0, 3, 1) ('left': -0.3374361684438957, 'right': -0.3374361684438957, 'forward': -0.3374361684438957, 'backward': -0.3374361684438957, 'up': -0.3374361684438957, 'down': -0.3374361684438957)	
(0, 3, 2) ('left': -0.3374361684438957, 'right': -0.3374361684438957, 'forward': -0.3374361684438957, 'backward': -0.3374361684438957, 'up': -0.3374361684438957, 'down': -0.3374361684438957)	
(0, 3, 3) ('left': -0.3374361684438957, 'right': -0.3374361684438957, 'forward': -0.3374361684438957, 'backward': -0.3374361684438957, 'up': -0.3374361684438957, 'down': -0.3374361684438957)	
(0, 3, 4) ('left': -0.3374361684438957, 'right': -0.3374361684438957, 'forward': -0.3374361684438957, 'backward': -0.3374361684438957, 'up': -0.3374361684438957, 'down': -0.3374361684438957)	
(0, 3, 5) ('left': -0.3374361684438957, 'right': -0.3374361684438957, 'forward': -0.3374361684438957, 'backward': -0.3374361684438957, 'up': -0.3374361684438957, 'down': -0.3374361684438957)	
(0, 1, 0) ('left': -0.335542169931118, 'right': -0.335542169931118, 'forward': -0.3194374461675641, 'backward': -0.3165980978114394, 'up': -0.340532972944029)	
(0, 1, 1) ('left': -0.25562823955637463, 'right': -0.22884384389646879, 'forward': -0.22884384389646879, 'backward': -0.24383513388492, 'up': -0.2353629218666466983)	
(0, 1, 2) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 1, 3) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 2, 0) ('left': -0.335542169931118, 'right': -0.335542169931118, 'forward': -0.3194374461675641, 'backward': -0.3165980978114394, 'up': -0.340532972944029)	
(0, 2, 1) ('left': -0.25562823955637463, 'right': -0.22884384389646879, 'forward': -0.22884384389646879, 'backward': -0.24383513388492, 'up': -0.2353629218666466983)	
(0, 2, 2) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 2, 3) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 3, 0) ('left': -0.335542169931118, 'right': -0.335542169931118, 'forward': -0.3194374461675641, 'backward': -0.3165980978114394, 'up': -0.340532972944029)	
(0, 3, 1) ('left': -0.25562823955637463, 'right': -0.22884384389646879, 'forward': -0.22884384389646879, 'backward': -0.24383513388492, 'up': -0.2353629218666466983)	
(0, 3, 2) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 3, 3) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 3, 4) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 1, 0) ('left': -0.335542169931118, 'right': -0.335542169931118, 'forward': -0.3194374461675641, 'backward': -0.3165980978114394, 'up': -0.340532972944029)	
(0, 1, 1) ('left': -0.25562823955637463, 'right': -0.22884384389646879, 'forward': -0.22884384389646879, 'backward': -0.24383513388492, 'up': -0.2353629218666466983)	
(0, 1, 2) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 1, 3) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 2, 0) ('left': -0.335542169931118, 'right': -0.335542169931118, 'forward': -0.3194374461675641, 'backward': -0.3165980978114394, 'up': -0.340532972944029)	
(0, 2, 1) ('left': -0.25562823955637463, 'right': -0.22884384389646879, 'forward': -0.22884384389646879, 'backward': -0.24383513388492, 'up': -0.2353629218666466983)	
(0, 2, 2) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 2, 3) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 3, 0) ('left': -0.335542169931118, 'right': -0.335542169931118, 'forward': -0.3194374461675641, 'backward': -0.3165980978114394, 'up': -0.340532972944029)	
(0, 3, 1) ('left': -0.25562823955637463, 'right': -0.22884384389646879, 'forward': -0.22884384389646879, 'backward': -0.24383513388492, 'up': -0.2353629218666466983)	
(0, 3, 2) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 3, 3) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 3, 4) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 1, 0) ('left': -0.335542169931118, 'right': -0.335542169931118, 'forward': -0.3194374461675641, 'backward': -0.3165980978114394, 'up': -0.340532972944029)	
(0, 1, 1) ('left': -0.25562823955637463, 'right': -0.22884384389646879, 'forward': -0.22884384389646879, 'backward': -0.24383513388492, 'up': -0.2353629218666466983)	
(0, 1, 2) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 1, 3) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 2, 0) ('left': -0.335542169931118, 'right': -0.335542169931118, 'forward': -0.3194374461675641, 'backward': -0.3165980978114394, 'up': -0.340532972944029)	
(0, 2, 1) ('left': -0.25562823955637463, 'right': -0.22884384389646879, 'forward': -0.22884384389646879, 'backward': -0.24383513388492, 'up': -0.2353629218666466983)	
(0, 2, 2) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 2, 3) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 3, 0) ('left': -0.335542169931118, 'right': -0.335542169931118, 'forward': -0.3194374461675641, 'backward': -0.3165980978114394, 'up': -0.340532972944029)	
(0, 3, 1) ('left': -0.25562823955637463, 'right': -0.22884384389646879, 'forward': -0.22884384389646879, 'backward': -0.24383513388492, 'up': -0.2353629218666466983)	
(0, 3, 2) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 3, 3) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 3, 4) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 1, 0) ('left': -0.335542169931118, 'right': -0.335542169931118, 'forward': -0.3194374461675641, 'backward': -0.3165980978114394, 'up': -0.340532972944029)	
(0, 1, 1) ('left': -0.25562823955637463, 'right': -0.22884384389646879, 'forward': -0.22884384389646879, 'backward': -0.24383513388492, 'up': -0.2353629218666466983)	
(0, 1, 2) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.25562823955637463, 'backward': -0.25562823955637463, 'up': -0.25562823955637463)	
(0, 1, 3) ('left': -0.25562823955637463, 'right': -0.25562823955637463, 'forward': -0.2556	

```
*****
-----Q-Table-Values-----
-----|-----|-----|-----|-----|
-0.4745|-0.5422|-0.3129|-0.1015|
-0.3150|-0.2359| 0.0088| 0.2675|
-0.2623|-0.1810| 0.3132| 0.3801|
-0.1676| 0.1534| 0.4358| 0.5065|
-----|-----|-----|-----|
-0.3019|-0.2025|-0.1328|-0.2836|
-0.3408|-0.0765|-0.1080| 0.1704|
-0.1374| 0.2207| 0.4568| 0.5501|
 0.0494| 0.1364| 0.4259| 0.7206|
-----|-----|-----|-----|
-0.2176|-0.0192| 0.0121| 0.4080|
-0.1349| 0.0831| 0.2391| 0.5914|
 0.0611| 0.3230| 0.4051| 0.7614|
 0.3337| 0.5826| 0.5500| 0.9748|
-----|-----|-----|-----|
 0.0316| 0.0901| 0.1608| 0.5036|
-0.0813| 0.2249| 0.3834| 0.7501|
 0.1192| 0.1907| 0.5535| 0.9695|
 0.4399| 0.6272| 0.7636| 0.0000|
*****-----Policy-Table-----|
-----|-----|-----|-----|-----|
 right | right | right | right |
 right | forward | up | right |
 up | up | up | forward |
 forward | forward | up | forward |
-----|-----|-----|-----|-----|
 forward | right | forward | down |
 up | right | forward | forward |
 forward | right | up | right |
 up | up | forward | forward |
-----|-----|-----|-----|-----|
 right | forward | right | forward |
 forward | up | up | forward |
 up | right | up | forward |
 up | up | up | forward |
-----|-----|-----|-----|-----|
 up | right | up | right |
 right | up | up | right |
 up | backward | up | right |
 up | up | up | GOAL |
*****-----Proportion of time spent updating each part of Q-----
-----|-----|-----|-----|
 0.0775| 0.0322| 0.0177| 0.0070|
 0.0446| 0.0231| 0.0127| 0.0094|
 0.0191| 0.0191| 0.0142| 0.0163|
 0.0086| 0.0081| 0.0075| 0.0062|
-----|-----|-----|-----|
 0.0190| 0.0120| 0.0151| 0.0068|
 0.0242| 0.0231| 0.0197| 0.0069|
 0.0144| 0.0144| 0.0100| 0.0174|
 0.0107| 0.0162| 0.0149| 0.0161|
-----|-----|-----|-----|
 0.0153| 0.0101| 0.0154| 0.0077|
 0.0123| 0.0121| 0.0296| 0.0277|
 0.0096| 0.0137| 0.0161| 0.0263|
 0.0045| 0.0082| 0.0215| 0.0319|
-----|-----|-----|-----|
 0.0064| 0.0049| 0.0077| 0.0099|
 0.0076| 0.0059| 0.0112| 0.0296|
 0.0055| 0.0053| 0.0128| 0.0538|
 0.0038| 0.0030| 0.0063| 0.0000|
*****
```

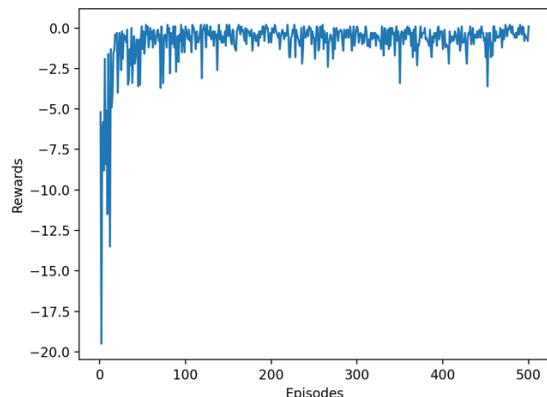


Figure 6 Step_10 episode_rewards vs episodes

Case 3:

Total Steps 10 version2

episode: 500, total_steps: 10 episode reward: 0.1000000000000009

```

Q-Table-FULL:
(0, 0, 0) ('left': -0.819382508514882, 'right': -0.66629223169992, 'forward': -0.8310238134453, 'backward': -0.843451243653129, 'up': -0.7989439177584514, 'down': -0.83250557440833276)
(0, 1, 0) ('left': -0.727343783827214, 'right': -0.7317734389138176, 'forward': -0.6214562731756502, 'backward': -0.6943577952402146, 'up': -0.5330784683664371, 'down': -0.7019879876854654)
(0, 0, 2) ('left': -0.624137088391281, 'right': -0.35542343686889794, 'forward': -0.5966259586786, 'backward': -0.6066994226384487, 'up': -0.5866723661435686, 'down': -0.5858665279374299)
(0, 0, 3) ('left': -0.599629598979574592, 'right': -0.555948488453702426, 'forward': -0.5910595742157535, 'backward': -0.5684341726498729, 'up': -0.599697213706074, 'down': -0.570202833549431)
(0, 1, 0) ('left': -0.6908054345620803, 'right': -0.7104730138286839, 'forward': -0.4351815771154893, 'backward': -0.69108439666239087, 'up': -0.7104302379614763, 'down': -0.6729567984638596)
(0, 1, 1) ('left': -0.5819446025674883, 'right': -0.5819446025674883, 'forward': -0.6806336825196255, 'backward': -0.586133884925787, 'up': -0.586133884925787, 'down': -0.56018158291727631)
(0, 1, 2) ('left': -0.484569795538825735, 'right': -0.4586772556709584, 'forward': -0.4465953087149343, 'backward': -0.4465953087149343, 'up': -0.47411166644007665, 'down': -0.432524720214188)
(0, 1, 3) ('left': -0.54587161568191971, 'right': -0.473161568191971, 'forward': -0.473161568191971, 'backward': -0.498587046847607, 'up': -0.498587046847607, 'down': -0.498587046847607)
(0, 2, 0) ('left': -0.60956458808089874, 'right': -0.5285423365813428, 'forward': -0.60956458808089874, 'backward': -0.60956458808089874, 'up': -0.60956458808089874, 'down': -0.60956458808089874)
(0, 2, 1) ('left': -0.6392395622556481, 'right': -0.555643107807774, 'forward': -0.6392395622556481, 'backward': -0.6392395622556481, 'up': -0.5177984695856285, 'down': -0.46579248524637544, 'up': -0.5177984695856285, 'down': -0.46579248524637544)
(0, 2, 2) ('left': -0.4261596795484666, 'right': -0.3744273636566343, 'forward': -0.4261596795484666, 'backward': -0.4261596795484666, 'up': -0.46579248524637544, 'down': -0.46579248524637544)
(0, 2, 3) ('left': -0.464633956532246176, 'right': -0.4513565017759697, 'forward': -0.464633956532246176, 'backward': -0.464633956532246176, 'up': -0.411842836536176, 'down': -0.411842836536176, 'up': -0.411842836536176, 'down': -0.411842836536176)
(0, 3, 0) ('left': -0.4150765837445514, 'right': -0.441042164244831, 'forward': -0.4150765837445514, 'backward': -0.4150765837445514, 'up': -0.38476475393758566, 'down': -0.38476475393758566, 'up': -0.38476475393758566, 'down': -0.38476475393758566)
(0, 3, 1) ('left': -0.425496880277151, 'right': -0.2876475393758566, 'forward': -0.425496880277151, 'backward': -0.425496880277151, 'up': -0.387561473273994, 'down': -0.387561473273994, 'up': -0.387561473273994, 'down': -0.387561473273994)
(0, 3, 2) ('left': -0.3936475393758566, 'right': -0.3936475393758566, 'forward': -0.3936475393758566, 'backward': -0.3936475393758566, 'up': -0.3936475393758566, 'down': -0.3936475393758566, 'up': -0.3936475393758566, 'down': -0.3936475393758566)
(0, 3, 3) ('left': -0.4374284406556987, 'right': -0.42999588662111926, 'forward': -0.4374284406556987, 'backward': -0.4374284406556987, 'up': -0.4374284406556987, 'down': -0.4374284406556987, 'up': -0.4374284406556987, 'down': -0.4374284406556987)
(0, 1, 0) ('left': -0.6549654359881831, 'right': -0.6311882829674676, 'forward': -0.6549654359881831, 'backward': -0.6484840704748522, 'up': -0.7020138400813288, 'down': -0.6548295070887323, 'up': -0.6548295070887323, 'down': -0.6548295070887323)
(0, 1, 1) ('left': -0.6295733803558222, 'right': -0.681858546715080914, 'forward': -0.6295733803558222, 'backward': -0.6999462510343191, 'up': -0.399932165686336, 'down': -0.60962248213372, 'up': -0.60962248213372, 'down': -0.60962248213372)
(0, 1, 2) ('left': -0.48765367098675776, 'right': -0.492218618869807, 'forward': -0.48765367098675776, 'backward': -0.492218618869807, 'up': -0.492218618869807, 'down': -0.5383438088886314, 'up': -0.492218618869807, 'down': -0.5383438088886314)
(0, 1, 3) ('left': -0.41893963104945177, 'right': -0.41893963104945177, 'forward': -0.41893963104945177, 'backward': -0.41893963104945177, 'up': -0.313181670919668015, 'down': -0.4188842744679484, 'up': -0.4188842744679484, 'down': -0.4188842744679484)
(0, 1, 4) ('left': -0.5086991519595461, 'right': -0.490724334995461, 'forward': -0.5086991519595461, 'backward': -0.5086991519595461, 'up': -0.4572631442998194, 'down': -0.4572631442998194, 'up': -0.4572631442998194, 'down': -0.4572631442998194)
(0, 1, 5) ('left': -0.3936475393758566, 'right': -0.378306475393758566, 'forward': -0.3936475393758566, 'backward': -0.3936475393758566, 'up': -0.3936475393758566, 'down': -0.3936475393758566, 'up': -0.3936475393758566, 'down': -0.3936475393758566)
(0, 1, 6) ('left': -0.3523759635656515, 'right': -0.3523759635656515, 'forward': -0.3523759635656515, 'backward': -0.3523759635656515, 'up': -0.3514087283244475, 'down': -0.3514087283244475, 'up': -0.3514087283244475, 'down': -0.3514087283244475)
(0, 1, 7) ('left': -0.1744787555583439, 'right': -0.316502023815324, 'forward': -0.1744787555583439, 'backward': -0.1744787555583439, 'up': -0.402256252712373, 'down': -0.5002532327847593, 'up': -0.402256252712373, 'down': -0.402256252712373)
(0, 1, 8) ('left': -0.370718772180778724, 'right': -0.370718772180778724, 'forward': -0.370718772180778724, 'backward': -0.370718772180778724, 'up': -0.3284659275831678, 'down': -0.34320821651520765, 'up': -0.3284659275831678, 'down': -0.34320821651520765)
(0, 1, 9) ('left': -0.2732568334685938, 'right': -0.27798723114189633, 'forward': -0.2732568334685938, 'backward': -0.27798723114189633, 'up': -0.3053797234370871644, 'down': -0.3053797234370871644, 'up': -0.3053797234370871644, 'down': -0.3053797234370871644)
(0, 1, 10) ('left': -0.3215871210262329, 'right': -0.2858911332832574, 'forward': -0.3215871210262329, 'backward': -0.2858911332832574, 'up': -0.3027512092911338885, 'down': -0.3027512092911338885, 'up': -0.3027512092911338885, 'down': -0.3027512092911338885)
(0, 1, 11) ('left': -0.268464671778724, 'right': -0.30237080572264063, 'forward': -0.268464671778724, 'backward': -0.268464671778724, 'up': -0.2729326364308305, 'down': -0.29711285261321191, 'up': -0.2729326364308305, 'down': -0.29711285261321191)
(0, 1, 12) ('left': -0.25858088694954366, 'right': -0.25858088694954366, 'forward': -0.25858088694954366, 'backward': -0.25858088694954366, 'up': -0.2745334623346819554, 'down': -0.26229131146381574, 'up': -0.2745334623346819554, 'down': -0.26229131146381574)
(0, 1, 13) ('left': -0.226983114949631007, 'right': -0.2223118162912542, 'forward': -0.226983114949631007, 'backward': -0.226983114949631007, 'up': -0.30237080572264063, 'down': -0.2729326364308305, 'up': -0.30237080572264063, 'down': -0.2729326364308305)
(0, 2, 0) ('left': -0.520470424840981, 'right': -0.52081799111187093, 'forward': -0.520470424840981, 'backward': -0.520470424840981, 'up': -0.35086228707470607, 'down': -0.543421998680871667, 'up': -0.543421998680871667, 'down': -0.543421998680871667)
(0, 2, 1) ('left': -0.5195972191808361, 'right': -0.50284785234218948, 'forward': -0.5195972191808361, 'backward': -0.5195972191808361, 'up': -0.1847399864163304, 'down': -0.502486868088510636, 'up': -0.502486868088510636, 'down': -0.502486868088510636)
(0, 2, 2) ('left': -0.4840853233593274, 'right': -0.40374463181908953, 'forward': -0.4840853233593274, 'backward': -0.40374463181908953, 'up': -0.41409293807982797, 'down': -0.4963373646403731, 'up': -0.41409293807982797, 'down': -0.4963373646403731)
(0, 2, 3) ('left': -0.392218618869807, 'right': -0.37798723114189633, 'forward': -0.392218618869807, 'backward': -0.37798723114189633, 'up': -0.3499621308369896, 'down': -0.39188517318864903, 'up': -0.3499621308369896, 'down': -0.39188517318864903)
(0, 2, 4) ('left': -0.3523759635656515, 'right': -0.3523759635656515, 'forward': -0.3523759635656515, 'backward': -0.3523759635656515, 'up': -0.3514087283244475, 'down': -0.3514087283244475, 'up': -0.3514087283244475, 'down': -0.3514087283244475)
(0, 2, 5) ('left': -0.350885725344411874, 'right': -0.350885725344411874, 'forward': -0.350885725344411874, 'backward': -0.350885725344411874, 'up': -0.350885725344411874, 'down': -0.350885725344411874, 'up': -0.350885725344411874, 'down': -0.350885725344411874)
(0, 2, 6) ('left': -0.308867254838823988, 'right': -0.308867254838823988, 'forward': -0.308867254838823988, 'backward': -0.308867254838823988, 'up': -0.308867254838823988, 'down': -0.308867254838823988, 'up': -0.308867254838823988, 'down': -0.308867254838823988)
(0, 2, 7) ('left': -0.29887474841559781, 'right': -0.29887474841559781, 'forward': -0.29887474841559781, 'backward': -0.29887474841559781, 'up': -0.29887474841559781, 'down': -0.305619272999999, 'up': -0.29887474841559781, 'down': -0.305619272999999, 'up': -0.29887474841559781, 'down': -0.305619272999999)
(0, 2, 8) ('left': -0.28987215180120285, 'right': -0.28987215180120285, 'forward': -0.28987215180120285, 'backward': -0.28987215180120285, 'up': -0.28987215180120285, 'down': -0.3272880678239529, 'up': -0.28987215180120285, 'down': -0.3272880678239529, 'up': -0.28987215180120285, 'down': -0.3272880678239529)
(0, 2, 9) ('left': -0.280874779588585, 'right': -0.280874779588585, 'forward': -0.280874779588585, 'backward': -0.280874779588585, 'up': -0.3051887794137814, 'down': -0.29664284537814, 'up': -0.3051887794137814, 'down': -0.29664284537814, 'up': -0.3051887794137814, 'down': -0.29664284537814)
(0, 2, 10) ('left': -0.2729326364308305, 'right': -0.2729326364308305, 'forward': -0.2729326364308305, 'backward': -0.2729326364308305, 'up': -0.2729326364308305, 'down': -0.2729326364308305, 'up': -0.2729326364308305, 'down': -0.2729326364308305, 'up': -0.2729326364308305, 'down': -0.2729326364308305)
(0, 2, 11) ('left': -0.2649988964839923, 'right': -0.31304837457555, 'forward': -0.2649988964839923, 'backward': -0.2649988964839923, 'up': -0.297181595999741, 'down': -0.306394711681317023, 'up': -0.297181595999741, 'down': -0.306394711681317023, 'up': -0.297181595999741, 'down': -0.306394711681317023)
(0, 2, 12) ('left': -0.24878362374135882, 'right': -0.2277420112297467, 'forward': -0.24878362374135882, 'backward': -0.2277420112297467, 'up': -0.19114970780321105, 'down': -0.41978527853178953, 'up': -0.19114970780321105, 'down': -0.41978527853178953, 'up': -0.19114970780321105, 'down': -0.41978527853178953)
(0, 2, 13) ('left': -0.15565125, 'right': -0.15565125, 'forward': -0.15565125, 'backward': -0.15565125, 'up': -0.15910517151829997, 'down': -0.15910517151829997, 'up': -0.15910517151829997, 'down': -0.15910517151829997, 'up': -0.15910517151829997, 'down': -0.15910517151829997)
(0, 2, 14) ('left': -0.30101518729933, 'right': -0.294537934545759, 'forward': -0.30101518729933, 'backward': -0.294537934545759, 'up': -0.1944740731022855, 'down': -0.19568625561033169, 'up': -0.1944740731022855, 'down': -0.19568625561033169, 'up': -0.1944740731022855, 'down': -0.19568625561033169)
(0, 2, 15) ('left': -0.17240731022855, 'right': -0.17240731022855, 'forward': -0.17240731022855, 'backward': -0.17240731022855, 'up': -0.151731275118797, 'down': -0.151731275118797, 'up': -0.151731275118797, 'down': -0.151731275118797, 'up': -0.151731275118797, 'down': -0.151731275118797)
(0, 2, 16) ('left': -0.2047182436210289, 'right': -0.50578131275118797, 'forward': -0.2047182436210289, 'backward': -0.50578131275118797, 'up': -0.15775735618565, 'down': -0.2327271785618565, 'up': -0.15775735618565, 'down': -0.2327271785618565, 'up': -0.15775735618565, 'down': -0.2327271785618565)
(0, 2, 17) ('left': -0.17517522704540923, 'right': -0.2061238753256527, 'forward': -0.17517522704540923, 'backward': -0.2061238753256527, 'up': -0.17738682798636417, 'down': -0.17738682798636417, 'up': -0.17738682798636417, 'down': -0.17738682798636417, 'up': -0.17738682798636417, 'down': -0.17738682798636417)
(0, 2, 18) ('left': -0.17517522704540923, 'right': -0.10279779899112168761, 'forward': -0.17517522704540923, 'backward': -0.10279779899112168761, 'up': -0.10279779899112168761, 'down': -0.10279779899112168761, 'up': -0.10279779899112168761, 'down': -0.10279779899112168761, 'up': -0.10279779899112168761, 'down': -0.10279779899112168761)
(0, 2, 19) ('left': 0, 'right': 0, 'forward': 0, 'backward': 0, 'up': 0, 'down': 0)
(0, 3, 0) ('left': 0, 'right': 0.09975, 'forward': 0.8176467450161555, 'backward': -0.13137614834936612, 'up': -0.05, 'down': -0.07750715538789843)
(0, 3, 1) ('left': 0, 'right': 0.323492332167322, 'forward': -0.325793724149568824, 'backward': -0.07923397231402151, 'up': -0.28601075231463813, 'down': -0.2962749064373438)
(0, 3, 2) ('left': 0, 'right': 0.17517522704540923, 'forward': -0.2061238753256527, 'backward': -0.17738682798636417, 'up': -0.17738682798636417, 'down': -0.17738682798636417, 'up': -0.17738682798636417, 'down': -0.17738682798636417)
(0, 3, 3) ('left': 0, 'right': 0.07398412280199494, 'forward': -0.07398412280199494, 'backward': -0.16762874375, 'up': -0.05167653625246713, 'down': -0.163284242159883295)
(0, 3, 4) ('left': 0, 'right': 0.09975, 'right': 0.8176467450161555, 'forward': -0.13137614834936612, 'backward': -0.13137614834936612, 'up': -0.05, 'down': -0.07750715538789843)
(0, 3, 5) ('left': 0, 'right': 0.323492332167322, 'right': 0.325793724149568824, 'forward': -0.325793724149568824, 'backward': -0.07923397231402151, 'up': -0.28601075231463813, 'down': -0.2962749064373438)
(0, 3, 6) ('left': 0, 'right': 0.17517522704540923, 'right': 0.2061238753256527, 'forward': -0.2061238753256527, 'backward': -0.17738682798636417, 'up': -0.17738682798636417, 'down': -0.17738682798636417, 'up': -0.17738682798636417, 'down': -0.17738682798636417)
(0, 3, 7) ('left': 0, 'right': 0.07398412280199494, 'right': 0.07398412280199494, 'forward': -0.07398412280199494, 'backward': -0.16762874375, 'up': -0.05167653625246713, 'down': -0.163284242159883295)
(0, 3, 8) ('left': 0, 'right': 0.09975, 'right': 0.8176467450161555, 'forward': -0.13137614834936612, 'backward': -0.1313761
```

----Q-Table-Values----				
-0.6063	-0.5331	-0.3554	-0.5550	
-0.4352	-0.3848	-0.0810	-0.4643	
-0.2232	-0.3555	-0.3744	-0.4040	
-0.0142	-0.1876	0.1280	-0.3657	
<hr/>				
-0.6312	-0.3910	-0.2084	-0.1318	
-0.4411	-0.3279	0.0145	-0.0883	
-0.3165	-0.1300	0.0540	0.2075	
0.0621	0.2746	0.5749	0.2770	
<hr/>				
-0.3509	-0.1847	-0.0376	0.2445	
-0.1431	0.0056	0.0654	0.3052	
0.0044	-0.0816	0.4107	0.3150	
0.1641	0.4314	0.7355	0.9376	
<hr/>				
-0.1877	-0.1866	0.1282	0.5165	
-0.0218	0.2831	0.3988	0.5078	
-0.3256	0.2941	0.7398	0.8176	
-0.0792	0.6145	0.9241	0.0000	
<hr/>				
----Policy-Table----				
right	up	right	right	
forward	right	forward	forward	
right	right	right	up	
forward	forward	forward	down	
<hr/>				
right	up	forward	forward	
up	forward	forward	forward	
right	forward	forward	forward	
up	up	forward	down	
<hr/>				
up	forward	right	forward	
up	right	up	forward	
right	up	up	forward	
up	up	up	forward	
<hr/>				
up	right	up	right	
up	up	up	right	
up	up	right	right	
backward	up	up	GOAL	
<hr/>				
----Proportion of time spent updating each part of Q----				
0.0680	0.0209	0.0119	0.0080	
0.0243	0.0104	0.0091	0.0071	
0.0120	0.0074	0.0059	0.0048	
0.0078	0.0073	0.0064	0.0051	
<hr/>				
0.0356	0.0184	0.0113	0.0067	
0.0185	0.0136	0.0108	0.0055	
0.0169	0.0109	0.0078	0.0040	
0.0171	0.0133	0.0113	0.0040	
<hr/>				
0.0258	0.0145	0.0101	0.0076	
0.0115	0.0165	0.0153	0.0116	
0.0115	0.0204	0.0213	0.0100	
0.0193	0.0323	0.0233	0.0097	
<hr/>				
0.0197	0.0095	0.0075	0.0070	
0.0206	0.0147	0.0108	0.0131	
0.0204	0.0218	0.0369	0.0183	
0.0182	0.0388	0.0606	0.0000	

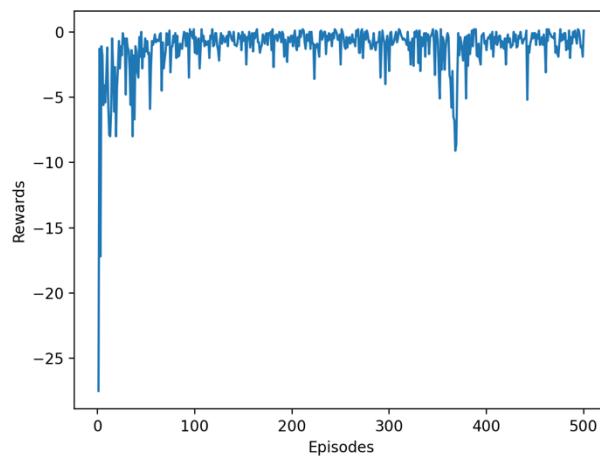


Figure 7 Step_10_version2 episode_rewards vs episodes

Case 4:

Total Steps 11

episode: 500, total_steps: 11 episode reward: 1.110223024625156e-16

```
*****Table-FULL*****
(0, 0, 0) {'left': -0.711443887789397, 'right': -0.6977821419891659, 'forward': -0.35893571788028146, 'backward': -0.6951769551484107, 'up': -0.6702865813838659, 'down': -0.6945545151362145}
(0, 0, 1) {'left': -0.812838482351587, 'right': -0.5881666222035, 'forward': -0.421534615651619, 'backward': -0.615431615651626, 'up': -0.879328935160801, 'down': -0.614477032461476}
(0, 0, 2) {'left': -0.705722145758659, 'right': -0.567522145758659, 'forward': -0.421534615651635, 'backward': -0.615431615651636, 'up': -0.879328935160803, 'down': -0.614477032461478}
(0, 0, 3) {'left': -0.4065393770959457, 'right': -0.281807718419855, 'forward': -0.4497356838366344, 'backward': -0.4157867534457687, 'up': -0.393465642545213, 'down': -0.433468894945484}
(0, 1, 0) {'left': -0.4241848656959577, 'right': -0.6223221453595983, 'forward': -0.6362983810253189, 'backward': -0.6446852525365579, 'up': -0.3521641549723876, 'down': -0.631879847172789}
(0, 1, 1) {'left': -0.5109258156322236, 'right': -0.50513093959526453, 'forward': -0.569668057879566, 'backward': -0.564635344755665, 'up': -0.33595653644274804, 'down': -0.4979531493002479}
(0, 1, 2) {'left': -0.36136664499031599, 'right': -0.15298433825983088, 'forward': -0.3234261084183114, 'backward': -0.3695082443996374, 'up': -0.356934945844153, 'down': -0.3337873465549979}
(0, 1, 3) {'left': -0.36136664499031599, 'right': -0.15298433825983088, 'forward': -0.3234261084183114, 'backward': -0.3695082443996374, 'up': -0.356934945844153, 'down': -0.3337873465549979}
(0, 2, 0) {'left': -0.58179098805663, 'right': -0.5148853425186938, 'forward': -0.5148853425186938, 'backward': -0.51865522422281, 'up': -0.51865522422281, 'down': -0.5138871995084285}
(0, 2, 1) {'left': -0.3513212888133537, 'right': -0.4740854581675246, 'forward': -0.39165854581675246, 'backward': -0.39165854581675246, 'up': -0.20161595503736295, 'down': -0.339534726578364}
(0, 2, 2) {'left': -0.3513212888133537, 'right': -0.4740854581675246, 'forward': -0.39165854581675246, 'backward': -0.39165854581675246, 'up': -0.20161595503736295, 'down': -0.339534726578364}
(0, 2, 3) {'left': -0.2804458795284578, 'right': -0.2871815745399441, 'forward': -0.42126294196427915, 'backward': -0.42126294196427915, 'up': -0.2871815745399441, 'down': -0.3155838804564}
(0, 3, 0) {'left': -0.46744935347359194, 'right': -0.4699521253317533, 'forward': -0.4699521253317533, 'backward': -0.46513272749041934, 'up': -0.395902363811207, 'down': -0.4363172742209127}
(0, 3, 1) {'left': -0.351317924272747715, 'right': -0.332785475335625, 'forward': -0.332785475335625, 'backward': -0.3318346397787873, 'up': -0.3972886514484962, 'down': -0.3338276113606211}
(0, 3, 2) {'left': -0.3013167446886565, 'right': -0.2962749084373438, 'forward': -0.26203364563991144, 'backward': -0.2780835371757813, 'up': -0.26839522911364686, 'down': -0.311128166044539}
(0, 3, 3) {'left': -0.2285876466815868, 'right': -0.24751262724481528, 'forward': -0.056661494955625, 'backward': -0.414347296494955628, 'up': -0.44849173348688302, 'down': -0.4187880713781144}
(1, 0, 0) {'left': -0.636466867436156, 'right': -0.570729934767194, 'forward': -0.4957747721634247, 'backward': -0.6482370525970851, 'up': -0.6122372754108734, 'down': -0.610236753236183}
(1, 0, 1) {'left': -0.4957747721634247, 'right': -0.469485993948867, 'forward': -0.469485993948867, 'backward': -0.526345621565165, 'up': -0.4957747721634247, 'down': -0.46369573525734687}
(1, 0, 2) {'left': -0.469485993948867, 'right': -0.469485993948867, 'forward': -0.469485993948867, 'backward': -0.469485993948867, 'up': -0.4957747721634247, 'down': -0.46369573525734687}
(1, 0, 3) {'left': -0.2657333992578243, 'right': -0.469521253317533, 'forward': -0.469521253317533, 'backward': -0.4621134736774748, 'up': -0.4674441492660924}
(1, 0, 4) {'left': -0.498386455450444, 'right': -0.512624482709282, 'forward': -0.2794185243281565, 'backward': -0.462126294196427915, 'up': -0.42126294196427915, 'down': -0.5164498341695953}
(1, 1, 0) {'left': -0.41859179989489214, 'right': -0.2343598834137431, 'forward': -0.39405631878544645, 'backward': -0.4912828820892456, 'up': -0.4154459498832735, 'down': -0.41244698164687794}
(1, 1, 1) {'left': -0.2185976468748836, 'right': -0.2185976468748836, 'forward': -0.2185976468748836, 'backward': -0.212178861958164, 'up': -0.212178861958164, 'down': -0.29115481971340774}
(1, 1, 2) {'left': -0.2664787426113046, 'right': -0.281229918598916, 'forward': -0.03888854195966637, 'backward': -0.302152342967358, 'up': -0.25064874745119673, 'down': -0.2873043179209622}
(1, 1, 3) {'left': -0.274547972455580783, 'right': -0.25087559965625, 'forward': -0.09685510917133806, 'backward': -0.2724487336494466, 'up': -0.2412332524749263, 'down': -0.29068933264717307}
(1, 2, 0) {'left': -0.24749772455580783, 'right': -0.24749772455580783, 'forward': -0.09685510917133806, 'backward': -0.2724487336494466, 'up': -0.2412332524749263, 'down': -0.29068933264717307}
(1, 2, 1) {'left': -0.24749772455580783, 'right': -0.16756684810875, 'forward': -0.09685510917133806, 'backward': -0.2724487336494466, 'up': -0.2412332524749263, 'down': -0.29068933264717307}
(1, 2, 2) {'left': -0.3339346830778907, 'right': -0.323865451252659865, 'forward': -0.302152342967358, 'backward': -0.323865451252659865, 'up': -0.30641598474271353, 'down': -0.3137302186987133}
(1, 2, 3) {'left': -0.1767867625, 'right': -0.17400187500000983, 'forward': -0.165191867500000983, 'backward': -0.165191867500000983, 'up': -0.165376874512528}
(1, 3, 0) {'left': -0.32398446303778907, 'right': -0.32398446303778907, 'forward': -0.302152342967358, 'backward': -0.32398446303778907, 'up': -0.30641598474271353, 'down': -0.3137302186987133}
(1, 3, 1) {'left': -0.17400187500000983, 'right': -0.17400187500000983, 'forward': -0.17400187500000983, 'backward': -0.17400187500000983, 'up': -0.16866749375}
(1, 3, 2) {'left': -0.1678767625, 'right': -0.17400187500000983, 'forward': -0.1678767625, 'backward': -0.1678767625, 'up': -0.165376874512528}
(1, 3, 3) {'left': -0.2118661494955625, 'right': -0.2118661494955625, 'forward': -0.2118661494955625, 'backward': -0.2118661494955625, 'up': -0.2118661494955625}
(2, 0, 0) {'left': -0.4185924410517152, 'right': -0.35460884373873888, 'forward': -0.1251852994139627, 'backward': -0.3476669474948871, 'up': -0.439422168625244516, 'down': -0.3746825113829664}
(2, 0, 1) {'left': -0.4185924410517152, 'right': -0.222133625241611396, 'forward': -0.21957674598659113, 'backward': -0.2846690140646684, 'up': -0.2705314937105131}
(2, 0, 2) {'left': -0.222133625241611396, 'right': -0.222133625241611396, 'forward': -0.222133625241611396, 'backward': -0.222133625241611396, 'up': -0.2705314937105131}
(2, 0, 3) {'left': -0.371245834913847, 'right': -0.485202235414486, 'forward': -0.33399334466444885, 'backward': -0.485202235414486, 'up': -0.373944212619397}
(2, 1, 0) {'left': -0.2854416582898674, 'right': -0.356824416582898674, 'forward': -0.356824416582898674, 'backward': -0.356824416582898674, 'up': -0.373944212619397}
(2, 1, 1) {'left': -0.2854416582898674, 'right': -0.2854416582898674, 'forward': -0.2854416582898674, 'backward': -0.2854416582898674, 'up': -0.373944212619397}
(2, 1, 2) {'left': -0.159971093750000082, 'right': -0.4039962346836242, 'forward': -0.159971093750000082, 'backward': -0.20458180310325456, 'up': -0.158319648797827479, 'down': -0.11852868993763633}
(2, 1, 3) {'left': -0.1377, 'right': -0.14925125, 'forward': -0.14925125, 'backward': -0.14925125, 'up': -0.09975, 'down': -0.12437525800000001}
(2, 2, 0) {'left': -0.26408335642171876, 'right': -0.24919810826932464, 'forward': -0.21799462169425267, 'backward': -0.2722081318800662, 'up': -0.19950601399259637}
(2, 2, 1) {'left': -0.24919810826932464, 'right': -0.16162425, 'forward': -0.16162425, 'backward': -0.1925263012499997, 'up': -0.44087475264756914, 'down': -0.19950601399259637}
(2, 2, 2) {'left': -0.16162425, 'right': -0.16162425, 'forward': -0.16162425, 'backward': -0.16162425, 'up': -0.09975, 'down': -0.12437525800000001}
(2, 2, 3) {'left': -0.159971093750000082, 'right': -0.2139532739625082, 'forward': -0.159971093750000082, 'backward': -0.2087134899332895, 'up': -0.2087134899332895, 'down': -0.124440128973568004}
(2, 3, 0) {'left': -0.09975, 'right': -0.14925125, 'forward': -0.14925125, 'backward': -0.14925125, 'up': -0.09975, 'down': -0.14775800000000001}
(2, 3, 1) {'left': -0.09975, 'right': -0.14925125, 'forward': -0.14925125, 'backward': -0.14925125, 'up': -0.09975, 'down': -0.14775800000000001}
(2, 3, 2) {'left': -0.09975, 'right': -0.14925125, 'forward': -0.14925125, 'backward': -0.14925125, 'up': -0.09975, 'down': -0.14775800000000001}
(2, 3, 3) {'left': -0.09975, 'right': -0.14925125, 'forward': -0.14925125, 'backward': -0.14925125, 'up': -0.09975, 'down': -0.14775800000000001}
(3, 0, 0) {'left': -0.3270327439354418, 'right': -0.08156799884914217, 'forward': -0.3780747424931854, 'backward': -0.32238264586808738, 'up': -0.34479351398515705}
(3, 0, 1) {'left': -0.1647657479747026, 'right': -0.256514445454269, 'forward': -0.256514445454269, 'backward': -0.256514445454269, 'up': -0.261779917401367}
(3, 0, 2) {'left': -0.18619186875, 'right': -0.18619186875, 'forward': -0.18619186875, 'backward': -0.18619186875, 'up': -0.23976493700000002, 'down': -0.23938989375}
(3, 0, 3) {'left': -0.18619186875, 'right': -0.18619186875, 'forward': -0.18619186875, 'backward': -0.18619186875, 'up': -0.23976493700000002, 'down': -0.23938989375}
(3, 1, 0) {'left': -0.2864991851114484, 'right': -0.1841924413627402, 'forward': -0.2788747924599724, 'backward': -0.2788747924599724, 'up': -0.34605188280835, 'down': -0.3376519795144957}
(3, 1, 1) {'left': -0.2266868344294875, 'right': -0.19923409469927653, 'forward': -0.22151783904646921, 'backward': -0.2151783904646921, 'up': -0.34644583449324645, 'down': -0.22280559698625}
(3, 1, 2) {'left': -0.1183125, 'right': -0.123580124740174, 'forward': -0.112274188438653, 'backward': -0.14663025, 'up': -0.09975, 'down': -0.09975}
(3, 1, 3) {'left': -0.0924670808777891456, 'right': -0.07128526247440465, 'forward': -0.07128526247440465, 'backward': -0.087152560000000001, 'up': -0.1977291122122418}
(3, 2, 0) {'left': -0.186315, 'right': -0.174001875, 'forward': -0.174001875, 'backward': -0.16156437, 'up': -0.09975, 'down': -0.12437625000000001}
(3, 2, 1) {'left': -0.081777053927764946, 'right': -0.07500000000000001, 'forward': -0.07500000000000001, 'backward': -0.09975, 'up': -0.17554369125269115, 'down': -0.12437625000000001}
(3, 2, 2) {'left': -0.19536080013592624, 'right': -0.065, 'forward': -0.065, 'backward': -0.065, 'up': 0, 'down': 0}
(3, 2, 3) {'left': -0.253702324373856, 'right': -0.162629208735214, 'forward': -0.1972498135945353, 'backward': -0.1972498135945353, 'up': 0, 'down': 0}
(3, 3, 0) {'left': -0.16000000000000001, 'right': -0.16000000000000001, 'forward': -0.16000000000000001, 'backward': -0.16000000000000001, 'up': 0, 'down': 0}
(3, 3, 1) {'left': -0.09975, 'right': -0.09975, 'forward': -0.09975, 'backward': -0.09975, 'up': 0.2026564621277633, 'down': -0.065}
(3, 3, 2) {'left': -0.095, 'right': -0.095, 'forward': -0.095, 'backward': -0.095, 'up': 0.0560130836466994, 'down': 0}
(3, 3, 3) {'left': 0, 'right': 0, 'forward': 0, 'backward': 0, 'up': 0, 'down': 0}
```

-----Q-Table-Values-----

-0.3589	-0.4282	-0.3115	-0.2281
-0.3522	-0.3360	-0.1497	-0.1530
-0.3978	-0.2016	-0.0779	-0.0213
-0.3395	0.0978	0.2620	0.0567
-0.2774	-0.1013	0.0252	0.0720
-0.2794	-0.2344	0.1019	-0.0381
-0.0626	0.0969	0.3008	0.0667
0.1506	0.0761	0.3460	0.1987
-0.1363	0.1251	0.2197	0.3754
0.0353	0.3567	0.4310	0.1781
0.2180	0.4481	0.7075	0.4045
0.2032	0.5548	0.7693	0.4198
-0.0816	0.2631	0.3664	0.4619
0.0142	0.1799	0.3943	0.7128
0.1977	0.1755	0.8461	0.9963
0.3101	0.8206	0.9568	0.0000

----Policy-Table----					
	forward	forward	forward	right	
up	up	forward	right		
up	up	up	forward		
up	up	forward	forward		
	forward	forward	forward	forward	
forward	right	forward	forward		
up	forward	forward	forward		
forward	forward	forward	forward		
	forward	forward	forward	forward	
up	right	right	forward		
forward	up	forward	right		
up	up	forward	down		
	right	up	up	right	
right	right	up	right		
up	up	up	right		
up	up	up	GOAL		

----Proportion of time spent updating each part of Q----			
0.0751	0.0215	0.0154	0.0094
0.0191	0.0143	0.0082	0.0061
0.0092	0.0102	0.0086	0.0050
0.0068	0.0055	0.0054	0.0038
0.0503	0.0229	0.0090	0.0071
0.0146	0.0114	0.0083	0.0057
0.0096	0.0069	0.0078	0.0059
0.0055	0.0053	0.0047	0.0055
0.0347	0.0244	0.0198	0.0099
0.0151	0.0128	0.0143	0.0119
0.0101	0.0129	0.0257	0.0142
0.0060	0.0086	0.0160	0.0092
0.0246	0.0236	0.0355	0.0301
0.0121	0.0094	0.0192	0.0427
0.0091	0.0118	0.0367	0.0641
0.0044	0.0053	0.0209	0.0000

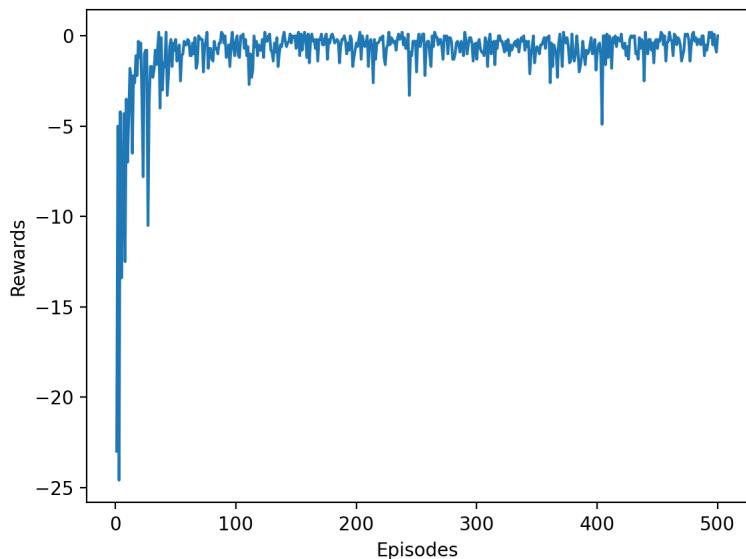


Figure 8 Step_11 episode_rewards vs episodes

Final Observations

Do the values eventually converge?

Step 9	Step 10 version 1	Step 10 version 2	Step 11																																																																																																																	
<p>----Q-Table-Values----</p> <table border="1"> <tr><td>-0.5897 -0.5393 -0.3497 -0.0959 </td><td>-0.4745 -0.5422 -0.3129 -0.1015 </td><td>-0.6963 -0.5331 -0.3554 -0.5550 </td><td>-0.3589 -0.4282 -0.3115 -0.2281 </td></tr> <tr><td>-0.3985 -0.2360 -0.1331 -0.3136 </td><td>-0.3150 -0.2359 0.0088 0.2675 </td><td>-0.4352 -0.3848 -0.0810 -0.4643 </td><td>-0.3522 -0.3368 -0.1497 -0.1538 </td></tr> <tr><td>-0.2982 -0.1575 0.0427 -0.1176 </td><td>-0.2623 -0.1810 0.3132 0.3801 </td><td>-0.2232 -0.3555 -0.3744 -0.4040 </td><td>-0.3978 -0.2016 -0.0779 -0.0213 </td></tr> <tr><td>-0.2134 0.0754 0.2358 -0.1555 </td><td>-0.1676 0.1534 0.4358 0.5065 </td><td>-0.0142 -0.1876 0.1280 -0.3657 </td><td>-0.3395 0.0978 0.2628 0.0567 </td></tr> <tr><td>-0.2478 -0.2298 -0.2837 0.1109 </td><td>-0.3019 -0.2025 -0.1328 -0.2836 </td><td>-0.6312 -0.3910 -0.2084 -0.1318 </td><td>-0.2774 -0.1013 0.0252 0.0720 </td></tr> <tr><td>-0.3120 0.0022 0.0738 0.3253 </td><td>-0.3408 -0.0765 -0.1880 0.1704 </td><td>-0.4411 -0.3279 0.0145 -0.0883 </td><td>-0.2794 -0.2344 0.1019 -0.0381 </td></tr> <tr><td>-0.0942 -0.1638 0.2818 0.4879 </td><td>-0.1374 0.2207 0.4568 0.5501 </td><td>-0.3165 -0.1300 0.0549 0.2075 </td><td>-0.0626 0.0969 0.3008 0.0667 </td></tr> <tr><td>0.0457 0.4011 0.6539 0.2655 </td><td>0.0494 0.1364 0.4259 0.7206 </td><td>0.0621 0.2746 0.5749 0.2770 </td><td>0.1586 0.0761 0.3460 0.1987 </td></tr> <tr><td>0.0514 0.0862 -0.1379 0.2028 </td><td>-0.2176 -0.0192 0.0121 0.4080 </td><td>-0.3509 -0.1847 -0.0376 0.2445 </td><td>-0.1363 0.1251 0.2197 0.3754 </td></tr> <tr><td>0.0347 0.2440 0.0941 0.4931 </td><td>-0.1349 0.0831 0.2391 0.5914 </td><td>-0.1431 0.0056 0.0654 0.3052 </td><td>0.0353 0.3567 0.4310 0.1781 </td></tr> <tr><td>0.2210 0.3334 0.4973 0.6789 </td><td>0.0611 0.3230 0.4051 0.7614 </td><td>0.0844 -0.0816 0.4107 0.3150 </td><td>0.2180 0.4481 0.7075 0.4045 </td></tr> <tr><td>0.2799 0.5951 0.5226 0.4036 </td><td>0.3337 0.5826 0.5500 0.9748 </td><td>0.1641 0.4314 0.7355 0.9376 </td><td>0.2032 0.5548 0.7693 0.4198 </td></tr> <tr><td>0.1552 0.2409 0.2521 -0.3311 </td><td>0.0316 0.0901 0.1608 0.5036 </td><td>-0.1877 -0.1866 0.1282 0.5165 </td><td>-0.0816 0.2631 0.3664 0.4619 </td></tr> <tr><td>0.3095 0.4910 0.5408 0.8121 </td><td>-0.0813 0.2249 0.3834 0.7501 </td><td>-0.3256 0.2941 0.7398 0.8176 </td><td>0.0142 0.1799 0.3943 0.7128 </td></tr> <tr><td>0.3333 0.3846 0.7313 0.9927 </td><td>0.1192 0.1907 0.5535 0.9695 </td><td>-0.0792 0.6145 0.9241 0.0000 </td><td>0.1977 0.1755 0.8461 0.9963 </td></tr> <tr><td>0.6097 0.7289 0.9382 0.0000 </td><td>0.4399 0.6272 0.7636 0.0000 </td><td>0.4399 0.6272 0.7636 0.0000 </td><td>0.3101 0.8206 0.9568 0.0000 </td></tr> </table> <p>----Policy-Table----</p> <table border="1"> <tr><td>right forward up forward </td><td>right right right right </td><td>right up right right </td><td>forward forward forward right </td></tr> <tr><td>right right right right </td><td>right forward up right </td><td>up up up forward </td><td>up up forward right </td></tr> <tr><td>right right right forward </td><td>up up up forward </td><td>forward forward up forward </td><td>up up up forward </td></tr> <tr><td>up forward forward left </td><td>forward forward up forward </td><td>forward forward forward down </td><td>up up up forward </td></tr> <tr><td>up forward forward down </td><td>forward forward down forward </td><td>right up forward forward </td><td>forward forward forward forward </td></tr> <tr><td>forward forward up right </td><td>forward right forward forward </td><td>up right up forward </td><td>forward right forward forward </td></tr> <tr><td>right forward forward right </td><td>forward right up forward </td><td>right forward forward forward </td><td>up right right forward </td></tr> <tr><td>forward up right forward </td><td>up right up forward </td><td>up up up forward </td><td>forward up forward right </td></tr> <tr><td>forward forward forward down </td><td>forward forward down forward </td><td>up up up forward </td><td>up up up forward down </td></tr> <tr><td>right right backward right </td><td>right forward right forward </td><td>up right up right </td><td>right up up right </td></tr> <tr><td>right right right right </td><td>right right up right </td><td>up up up right </td><td>right right up right </td></tr> <tr><td>up up up right </td><td>up backward up right </td><td>GOAL up up GOAL </td><td>up up up right </td></tr> <tr><td>up up up GOAL </td><td>up up up GOAL </td><td></td><td>up up up GOAL </td></tr> </table>	-0.5897 -0.5393 -0.3497 -0.0959	-0.4745 -0.5422 -0.3129 -0.1015	-0.6963 -0.5331 -0.3554 -0.5550	-0.3589 -0.4282 -0.3115 -0.2281	-0.3985 -0.2360 -0.1331 -0.3136	-0.3150 -0.2359 0.0088 0.2675	-0.4352 -0.3848 -0.0810 -0.4643	-0.3522 -0.3368 -0.1497 -0.1538	-0.2982 -0.1575 0.0427 -0.1176	-0.2623 -0.1810 0.3132 0.3801	-0.2232 -0.3555 -0.3744 -0.4040	-0.3978 -0.2016 -0.0779 -0.0213	-0.2134 0.0754 0.2358 -0.1555	-0.1676 0.1534 0.4358 0.5065	-0.0142 -0.1876 0.1280 -0.3657	-0.3395 0.0978 0.2628 0.0567	-0.2478 -0.2298 -0.2837 0.1109	-0.3019 -0.2025 -0.1328 -0.2836	-0.6312 -0.3910 -0.2084 -0.1318	-0.2774 -0.1013 0.0252 0.0720	-0.3120 0.0022 0.0738 0.3253	-0.3408 -0.0765 -0.1880 0.1704	-0.4411 -0.3279 0.0145 -0.0883	-0.2794 -0.2344 0.1019 -0.0381	-0.0942 -0.1638 0.2818 0.4879	-0.1374 0.2207 0.4568 0.5501	-0.3165 -0.1300 0.0549 0.2075	-0.0626 0.0969 0.3008 0.0667	0.0457 0.4011 0.6539 0.2655	0.0494 0.1364 0.4259 0.7206	0.0621 0.2746 0.5749 0.2770	0.1586 0.0761 0.3460 0.1987	0.0514 0.0862 -0.1379 0.2028	-0.2176 -0.0192 0.0121 0.4080	-0.3509 -0.1847 -0.0376 0.2445	-0.1363 0.1251 0.2197 0.3754	0.0347 0.2440 0.0941 0.4931	-0.1349 0.0831 0.2391 0.5914	-0.1431 0.0056 0.0654 0.3052	0.0353 0.3567 0.4310 0.1781	0.2210 0.3334 0.4973 0.6789	0.0611 0.3230 0.4051 0.7614	0.0844 -0.0816 0.4107 0.3150	0.2180 0.4481 0.7075 0.4045	0.2799 0.5951 0.5226 0.4036	0.3337 0.5826 0.5500 0.9748	0.1641 0.4314 0.7355 0.9376	0.2032 0.5548 0.7693 0.4198	0.1552 0.2409 0.2521 -0.3311	0.0316 0.0901 0.1608 0.5036	-0.1877 -0.1866 0.1282 0.5165	-0.0816 0.2631 0.3664 0.4619	0.3095 0.4910 0.5408 0.8121	-0.0813 0.2249 0.3834 0.7501	-0.3256 0.2941 0.7398 0.8176	0.0142 0.1799 0.3943 0.7128	0.3333 0.3846 0.7313 0.9927	0.1192 0.1907 0.5535 0.9695	-0.0792 0.6145 0.9241 0.0000	0.1977 0.1755 0.8461 0.9963	0.6097 0.7289 0.9382 0.0000	0.4399 0.6272 0.7636 0.0000	0.4399 0.6272 0.7636 0.0000	0.3101 0.8206 0.9568 0.0000	right forward up forward	right right right right	right up right right	forward forward forward right	right right right right	right forward up right	up up up forward	up up forward right	right right right forward	up up up forward	forward forward up forward	up up up forward	up forward forward left	forward forward up forward	forward forward forward down	up up up forward	up forward forward down	forward forward down forward	right up forward forward	forward forward forward forward	forward forward up right	forward right forward forward	up right up forward	forward right forward forward	right forward forward right	forward right up forward	right forward forward forward	up right right forward	forward up right forward	up right up forward	up up up forward	forward up forward right	forward forward forward down	forward forward down forward	up up up forward	up up up forward down	right right backward right	right forward right forward	up right up right	right up up right	right right right right	right right up right	up up up right	right right up right	up up up right	up backward up right	GOAL up up GOAL	up up up right	up up up GOAL	up up up GOAL		up up up GOAL
-0.5897 -0.5393 -0.3497 -0.0959	-0.4745 -0.5422 -0.3129 -0.1015	-0.6963 -0.5331 -0.3554 -0.5550	-0.3589 -0.4282 -0.3115 -0.2281																																																																																																																	
-0.3985 -0.2360 -0.1331 -0.3136	-0.3150 -0.2359 0.0088 0.2675	-0.4352 -0.3848 -0.0810 -0.4643	-0.3522 -0.3368 -0.1497 -0.1538																																																																																																																	
-0.2982 -0.1575 0.0427 -0.1176	-0.2623 -0.1810 0.3132 0.3801	-0.2232 -0.3555 -0.3744 -0.4040	-0.3978 -0.2016 -0.0779 -0.0213																																																																																																																	
-0.2134 0.0754 0.2358 -0.1555	-0.1676 0.1534 0.4358 0.5065	-0.0142 -0.1876 0.1280 -0.3657	-0.3395 0.0978 0.2628 0.0567																																																																																																																	
-0.2478 -0.2298 -0.2837 0.1109	-0.3019 -0.2025 -0.1328 -0.2836	-0.6312 -0.3910 -0.2084 -0.1318	-0.2774 -0.1013 0.0252 0.0720																																																																																																																	
-0.3120 0.0022 0.0738 0.3253	-0.3408 -0.0765 -0.1880 0.1704	-0.4411 -0.3279 0.0145 -0.0883	-0.2794 -0.2344 0.1019 -0.0381																																																																																																																	
-0.0942 -0.1638 0.2818 0.4879	-0.1374 0.2207 0.4568 0.5501	-0.3165 -0.1300 0.0549 0.2075	-0.0626 0.0969 0.3008 0.0667																																																																																																																	
0.0457 0.4011 0.6539 0.2655	0.0494 0.1364 0.4259 0.7206	0.0621 0.2746 0.5749 0.2770	0.1586 0.0761 0.3460 0.1987																																																																																																																	
0.0514 0.0862 -0.1379 0.2028	-0.2176 -0.0192 0.0121 0.4080	-0.3509 -0.1847 -0.0376 0.2445	-0.1363 0.1251 0.2197 0.3754																																																																																																																	
0.0347 0.2440 0.0941 0.4931	-0.1349 0.0831 0.2391 0.5914	-0.1431 0.0056 0.0654 0.3052	0.0353 0.3567 0.4310 0.1781																																																																																																																	
0.2210 0.3334 0.4973 0.6789	0.0611 0.3230 0.4051 0.7614	0.0844 -0.0816 0.4107 0.3150	0.2180 0.4481 0.7075 0.4045																																																																																																																	
0.2799 0.5951 0.5226 0.4036	0.3337 0.5826 0.5500 0.9748	0.1641 0.4314 0.7355 0.9376	0.2032 0.5548 0.7693 0.4198																																																																																																																	
0.1552 0.2409 0.2521 -0.3311	0.0316 0.0901 0.1608 0.5036	-0.1877 -0.1866 0.1282 0.5165	-0.0816 0.2631 0.3664 0.4619																																																																																																																	
0.3095 0.4910 0.5408 0.8121	-0.0813 0.2249 0.3834 0.7501	-0.3256 0.2941 0.7398 0.8176	0.0142 0.1799 0.3943 0.7128																																																																																																																	
0.3333 0.3846 0.7313 0.9927	0.1192 0.1907 0.5535 0.9695	-0.0792 0.6145 0.9241 0.0000	0.1977 0.1755 0.8461 0.9963																																																																																																																	
0.6097 0.7289 0.9382 0.0000	0.4399 0.6272 0.7636 0.0000	0.4399 0.6272 0.7636 0.0000	0.3101 0.8206 0.9568 0.0000																																																																																																																	
right forward up forward	right right right right	right up right right	forward forward forward right																																																																																																																	
right right right right	right forward up right	up up up forward	up up forward right																																																																																																																	
right right right forward	up up up forward	forward forward up forward	up up up forward																																																																																																																	
up forward forward left	forward forward up forward	forward forward forward down	up up up forward																																																																																																																	
up forward forward down	forward forward down forward	right up forward forward	forward forward forward forward																																																																																																																	
forward forward up right	forward right forward forward	up right up forward	forward right forward forward																																																																																																																	
right forward forward right	forward right up forward	right forward forward forward	up right right forward																																																																																																																	
forward up right forward	up right up forward	up up up forward	forward up forward right																																																																																																																	
forward forward forward down	forward forward down forward	up up up forward	up up up forward down																																																																																																																	
right right backward right	right forward right forward	up right up right	right up up right																																																																																																																	
right right right right	right right up right	up up up right	right right up right																																																																																																																	
up up up right	up backward up right	GOAL up up GOAL	up up up right																																																																																																																	
up up up GOAL	up up up GOAL		up up up GOAL																																																																																																																	
<p>----Policy-Table----</p> <table border="1"> <tr><td>right forward up right </td><td>right right right right </td><td>right up right right </td><td>forward forward forward right </td></tr> <tr><td>right right right right </td><td>right forward up right </td><td>up up up forward </td><td>up up forward right </td></tr> <tr><td>right right right forward </td><td>up up up forward </td><td>forward forward forward down </td><td>up up up forward </td></tr> <tr><td>up forward forward left </td><td>forward forward up forward </td><td>right up forward forward </td><td>forward forward forward forward </td></tr> <tr><td>up forward forward down </td><td>forward forward down forward </td><td>up right up forward </td><td>forward right forward forward </td></tr> <tr><td>forward forward up right </td><td>forward right forward forward </td><td>right forward forward forward </td><td>up right right forward </td></tr> <tr><td>forward forward forward right </td><td>up right up forward </td><td>up up up forward </td><td>forward up forward right </td></tr> <tr><td>forward forward forward right </td><td>up right up forward </td><td>up up up forward </td><td>up up up forward down </td></tr> <tr><td>right right backward right </td><td>right forward right forward </td><td>up right up right </td><td>right up up right </td></tr> <tr><td>right right right right </td><td>right right up right </td><td>up up up right </td><td>right right up right </td></tr> <tr><td>up up up right </td><td>up backward up right </td><td>GOAL up up GOAL </td><td>up up up right </td></tr> <tr><td>up up up GOAL </td><td>up up up GOAL </td><td></td><td>up up up GOAL </td></tr> </table>	right forward up right	right right right right	right up right right	forward forward forward right	right right right right	right forward up right	up up up forward	up up forward right	right right right forward	up up up forward	forward forward forward down	up up up forward	up forward forward left	forward forward up forward	right up forward forward	forward forward forward forward	up forward forward down	forward forward down forward	up right up forward	forward right forward forward	forward forward up right	forward right forward forward	right forward forward forward	up right right forward	forward forward forward right	up right up forward	up up up forward	forward up forward right	forward forward forward right	up right up forward	up up up forward	up up up forward down	right right backward right	right forward right forward	up right up right	right up up right	right right right right	right right up right	up up up right	right right up right	up up up right	up backward up right	GOAL up up GOAL	up up up right	up up up GOAL	up up up GOAL		up up up GOAL																																																																				
right forward up right	right right right right	right up right right	forward forward forward right																																																																																																																	
right right right right	right forward up right	up up up forward	up up forward right																																																																																																																	
right right right forward	up up up forward	forward forward forward down	up up up forward																																																																																																																	
up forward forward left	forward forward up forward	right up forward forward	forward forward forward forward																																																																																																																	
up forward forward down	forward forward down forward	up right up forward	forward right forward forward																																																																																																																	
forward forward up right	forward right forward forward	right forward forward forward	up right right forward																																																																																																																	
forward forward forward right	up right up forward	up up up forward	forward up forward right																																																																																																																	
forward forward forward right	up right up forward	up up up forward	up up up forward down																																																																																																																	
right right backward right	right forward right forward	up right up right	right up up right																																																																																																																	
right right right right	right right up right	up up up right	right right up right																																																																																																																	
up up up right	up backward up right	GOAL up up GOAL	up up up right																																																																																																																	
up up up GOAL	up up up GOAL		up up up GOAL																																																																																																																	

Theoretically, with increasing iterations, the Q-Table values should eventually converge. This is especially so for majority of the values from Q-Tables with similar Step values. However, there are edge cases for some states with large diverging values (also shown in the policy table with ‘backward’ action), and I would attribute this to the random chance of action occurring in the environment file, which may, by random chance, disrupt the sub-optimal value needed to compute for the subsequent most optimal value.

reward, terminate, next_state = env.step(action)

```

def step(self, action, stochastic=True):
    in_action = action # action from agent
    assert action in self.all_actions
    reward = -0.1
    is_terminate = False
    pre_pos = self.curr_pos
    r = random.random()
    if action == 'right':
        if r < 0.1:
            action = 'up'
        elif r < 0.2:
            action = 'down'
        elif r < 0.3:
            action = 'forward'
        elif r < 0.4:
            action = 'backward'
        else:
            action = 'right'
    else:
        if r < 0.1:
            action = 'up'
        elif r < 0.2:
            action = 'down'
        elif r < 0.3:
            action = 'forward'
        elif r < 0.4:
            action = 'backward'
        else:
            action = 'right'

```

Figure 9 Random Occurring Chance

Further Possible Optimization

There are varying methods to improve the performance of Q-learning, such as using simultaneous Q-values updating [1]. It's possible to increase the number of episodes and further decrease the epsilon (exploration rate) from 0.01 to 0.0001, and toggling stochastic to false. These would produce a more consistent and converging values.

```
def step(self, action, stochastic=False):
    in_action = action # action from agent
    assert action in self.all_actions
    reward = -0.1
    is_terminate = False
    pre_pos = self.curr_pos
    r = random.random()
```

Figure 10 Stochastic = False

Iteration 1 @ eps 0.0001	Iteration 2 @ eps 0.0001
<pre>----Q-Table-Values----- ----- 0.1502 -0.2161 -0.2231 -0.1740 0.2527 0.3563 -0.0503 -0.1493 -0.2132 0.4609 0.0569 -0.0998 -0.1493 0.5666 0.4014 -0.0998 ----- -0.2542 0.0915 -0.1740 -0.1493 -0.1246 0.4377 0.3708 -0.0998 0.2105 0.5651 0.6732 -0.0998 -0.0998 0.6733 0.7811 0.8606 ----- -0.2124 -0.1225 -0.1246 -0.0998 -0.1736 0.1587 0.5261 -0.0998 -0.0367 0.3749 0.7795 0.6550 -0.0998 0.3069 0.8900 1.0000 ----- -0.1493 -0.0689 0.2698 0.0927 0.0379 0.3683 0.7065 0.6390 -0.0500 -0.0500 0.8759 0.9995 -0.0500 -0.0500 0.7500 0.0000 </pre>	<pre>----Q-Table-Values----- ----- 0.1502 -0.2522 -0.1922 -0.1493 -0.2070 -0.1862 -0.1493 -0.1493 0.0468 0.0907 -0.0998 -0.0998 -0.0998 0.3004 -0.0021 -0.0998 ----- 0.2527 -0.1985 -0.1614 -0.1493 -0.1266 0.0710 -0.1246 -0.0998 0.3359 0.5381 0.2836 -0.0750 -0.0998 0.6682 0.7805 -0.0500 ----- 0.3563 0.2345 -0.0998 -0.0998 0.4609 0.5666 -0.0037 -0.0998 -0.0998 0.0637 0.6281 0.7725 -0.0500 -0.0500 0.8900 1.0000 ----- -0.1472 0.5146 0.1307 -0.0998 0.1283 0.6733 0.7811 0.7686 -0.0998 0.6772 0.8900 1.0000 -0.0500 -0.0500 0.7500 0.0000 </pre>
<p>Rewards</p> <p>Episodes</p>	<p>Rewards</p> <p>Episodes</p>
<pre>episode: 1000, total_steps: 9 episode reward: 0.2000000 *****</pre>	<pre>episode: 1000, total_steps: 9 episode reward: 0.2000000 *****</pre>

Reference

- [1] M. Pouyan, "Improving the performance of Q-learning using simultaneous Q-values updating," *International Congress on Technology, Communication and Knowledge (ICTCK)*, vol. 10, pp. 1-6, 2014.
- [2] C. J. Watkins, "Technial Note Q-Learning," *Kluwer Academic*, vol. 8, pp. 279-292, 1992.