

Reinforcement Learning Final Project CheckPoint 1

Team members -

Mohammed Junaid Shaik – mshaik6

Naman Khurpia – namankhu

Komal Vaishnavi Pulla – komalvai

INTRODUCTION:

Solve a multi agent environment using reinforcement learning. In this project we will be recreating a Mario game on a grid world which would be a multi-agent environment with the agents: Mario and Luigi in a 10 x 10 grid world trying to achieve the same objective. We plan on using DQN to decide the path that the characters choose while travelling to the common goal

The objective is to create a grid world where mario and luigi competes for the same goal, i.e. to reach/kill the princess respectively, this game depicts a multi agent RL solution to real world game, displaying a competitive environment having multiple rewards for the level setup

ENVIRONMENT DESCRIPTION:

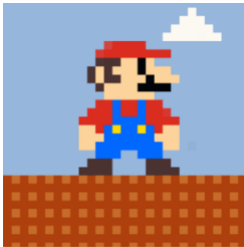
Actions: The agent can move in the direction of up, down, left and right by one step {Right,Left,Up,Down}

State: The states include from position (0,0) initial state to (10,10)

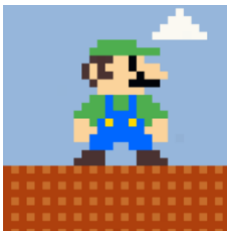
Looks and feels of Grid world -

The agents are -

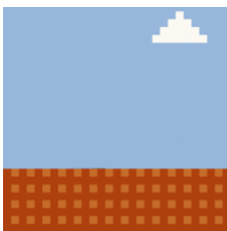
- Mario



- Luigi

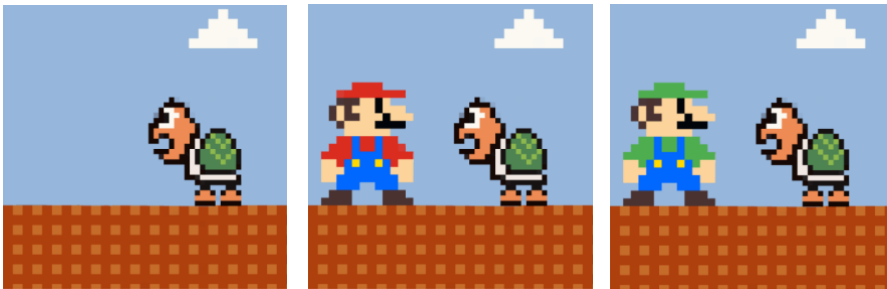


- Sky

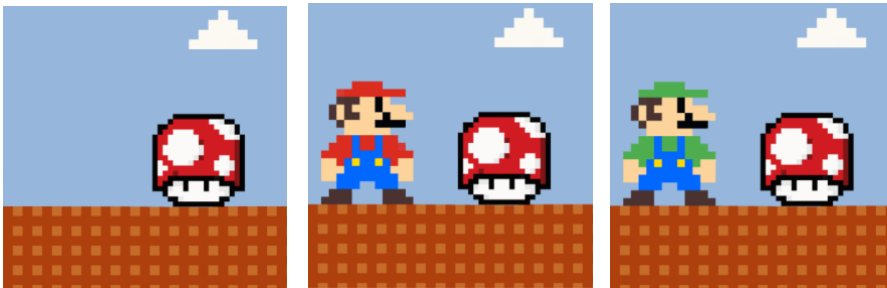


The rewards will be for

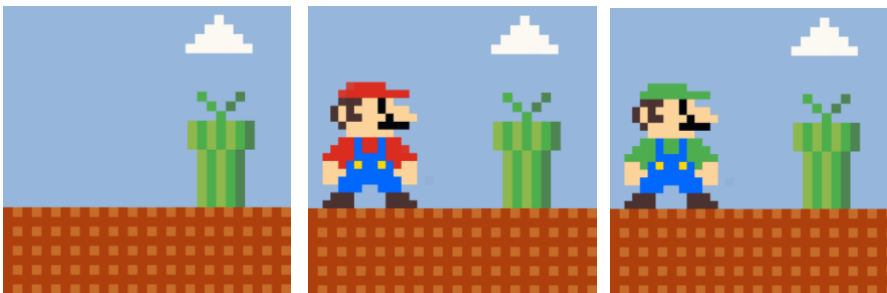
- Tortoise : -100



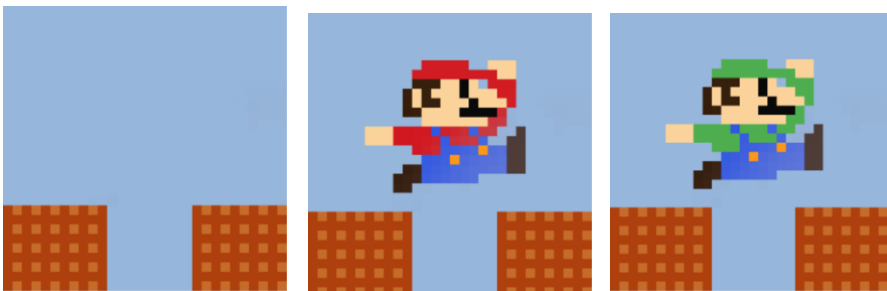
- Oneup : -50



- Plant : -20



- Jumping point: -10



- Princess : +100



Positions:

- Mario : [0,0]
- Luigi:[9,0]
- Princess: [8,9]
- Tortoise Position : [7,4],[2,5],[4,9],[8,9]
- Plant Position: ([2, 0], [4, 0], [6, 0],[3, 1], [5, 1], [8, 1], [0, 2], [7, 2], [9, 2], [1, 3], [5, 3], [8, 3],[0, 4], [4, 4], [6, 4],[5, 5], [9, 5], [2, 6], [8, 6], [1, 7], [3, 7], [9, 7], [0, 8], [2, 8], [1, 9])
- Jump position : [7, 3], [2, 4], [6, 4], [8, 4], [1, 5], [3, 5],[7, 5], [2, 6], [4, 8], [8, 8], [3, 9], [5, 9],[7, 9], [9, 9]
- One up Position: [3, 0], [5, 0], [8, 2], [0, 3], [5, 4], [9, 6], [2, 7], [0, 9]

Policy :

- Either the ghost will surround the jumping point or the Turtle will surround the jumping point
- The aim of Mario would be to reach the princess, avoid turtles, Ghosts, and Jumping points.
- The aim of Luigi would be to kill the princess, avoid turtles, Ghosts, and Jumping points.

Aim:

The game will end when Mario reaches the princess with a maximum reward.

Though the primary goal of the game is for Mario to reach the princess, we also keep track of the rewards that Mario has accumulated along the way. We will try to finish the game and tune it in such a way that Mario gets the maximum reward. We will be using 4 Q tables to choose actions, 2 for Mario and 2 for Luigi.

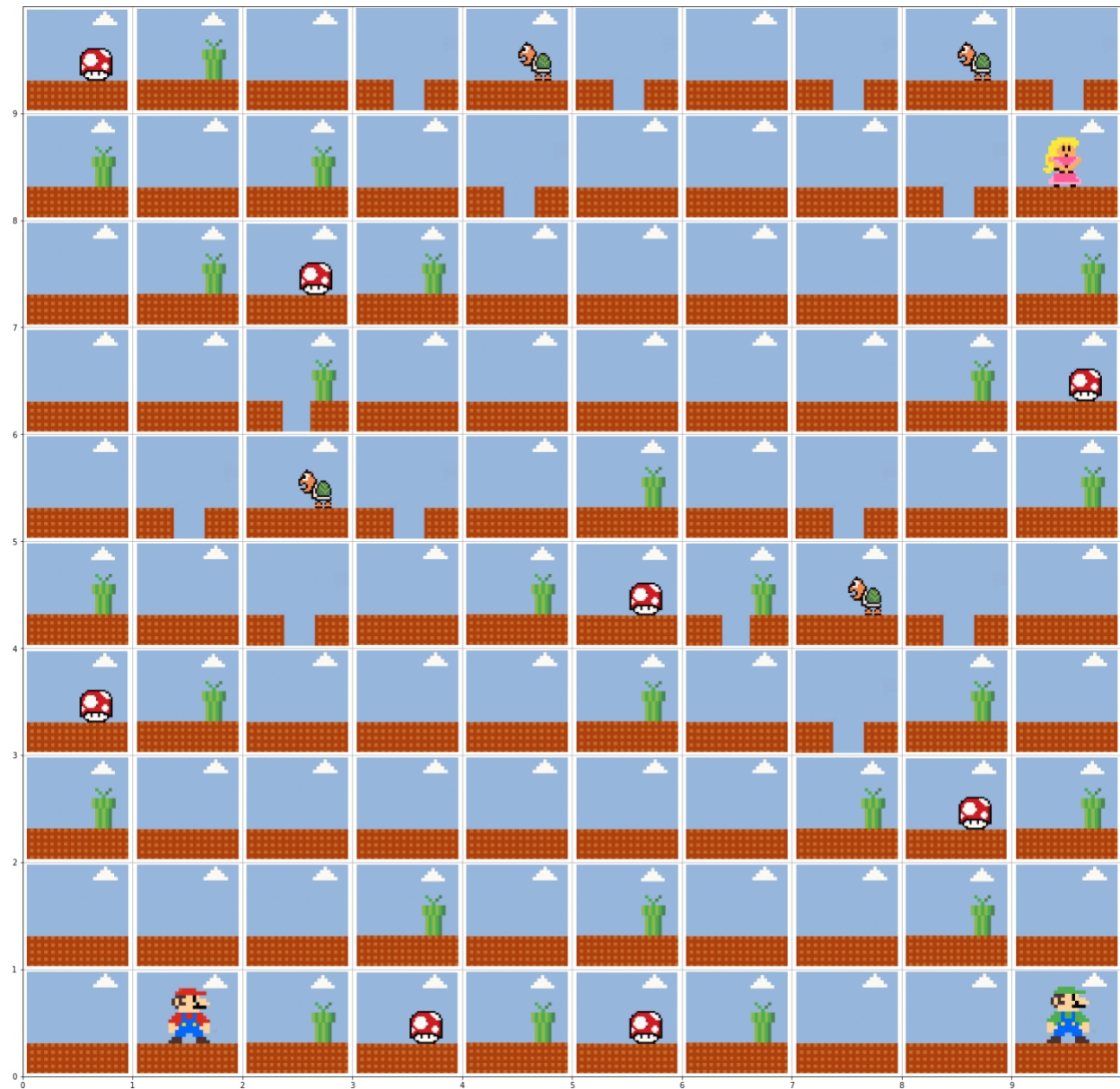
INITIAL ENVIRONMENT:

The movement of luigi and mario is managed with two q tables so far.



The agent will take 10 steps now

(1, 9, 0, 0, False, False, {})



SARSA Results –

Mario's Q Table –

[illegible]

[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[-2.15010063	-0.71272348	-2.16177022	-2.08386012]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[-8.53097992	-10.93702809	-12.13063419	-9.322493]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[0.	0.	0.]
[-8.92881044	-8.02778352	-7.90640263	-10.15858809]

```
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ -4.62739466 -2.90423079 -2.65506371 -3.84700709]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ -5.16186224 -3.23886883 -5.91959608 -6.57668693]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ -3.56847285 -5.77425156 -2.14698092  0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
[ 0.    0.    0.    0.    ]
```

$$\begin{bmatrix} 0. & 0. & 0. & 0. \\ 0. & 0. & 0. & 0. \end{bmatrix}]$$

Luigi's Q Table –

Luigi's Q table

[illegible]

[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[-8.17897817 -12.11457926 -13.06946414 -10.66977897]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[-8.40807277 -7.70667218 -7.59014653 -11.41131584]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[-4.12672592 -2.74819483 -2.54886116 -3.57776687]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[0. 0. 0. 0.]
[-6.52041363 -2.9820189 -7.25389588 -7.65227339]
[0. 0. 0. 0.]
[0. 0. 0. 0.]

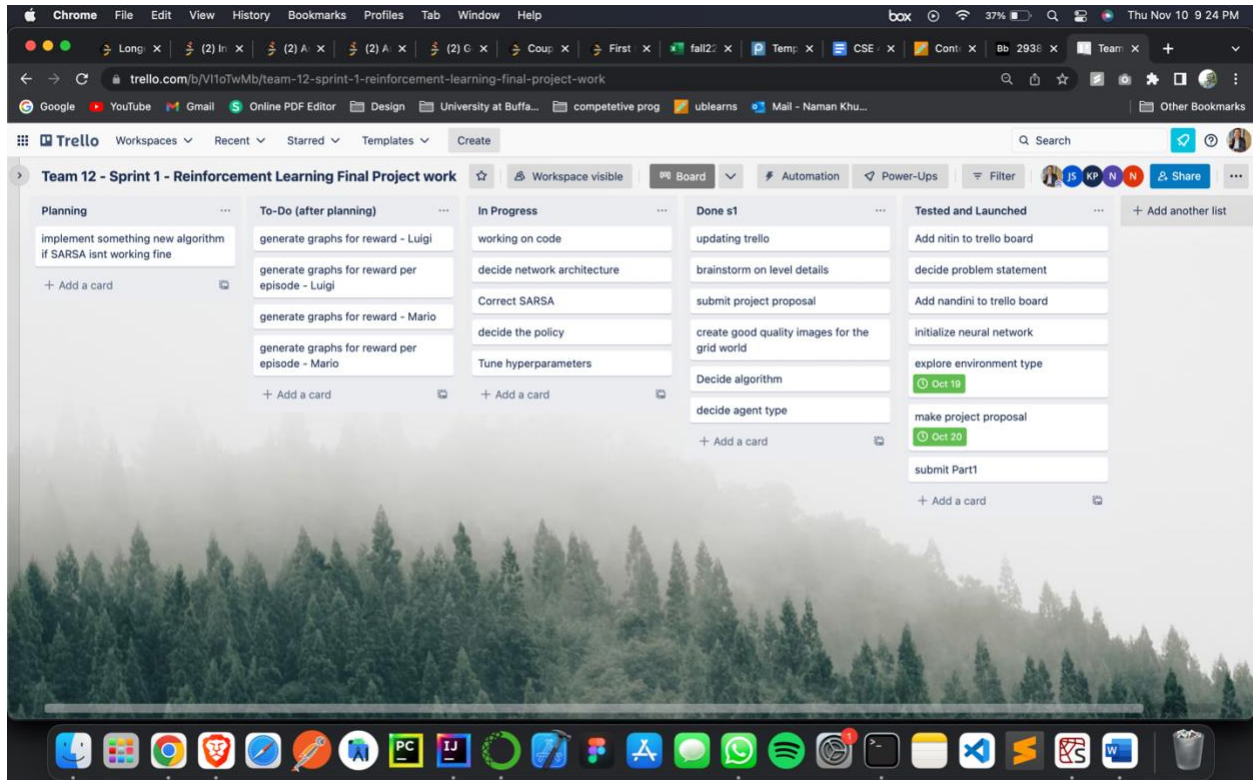
```

[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]
[ -2.85477828 -4.61940125 -1.71758473 0.      ]
[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]
[ 0.      0.      0.      0.      ]

```

NEXT STEPS: SARSA results will be corrected and DQN will be implemented, an epsilon greedy policy will be implemented, proper comparison between algorithm will be provided, and visualized along with proper graphs.

TRACKBOARD-



THANK YOU