Q-Learning Assignment Report

K BHARGAV SASHANK
23634_Assignment3_report

April 9, 2025

1 Maze Generation and Verification

Maze was generated using the provided script in `QL_Assignment.ipynb` with the SR No :- 23634 as seed. The script in path.ipynb file was used to verify that a valid path exists from the start to the goal for the generated maze.

2 Training Under Trap and Boost Scenarios

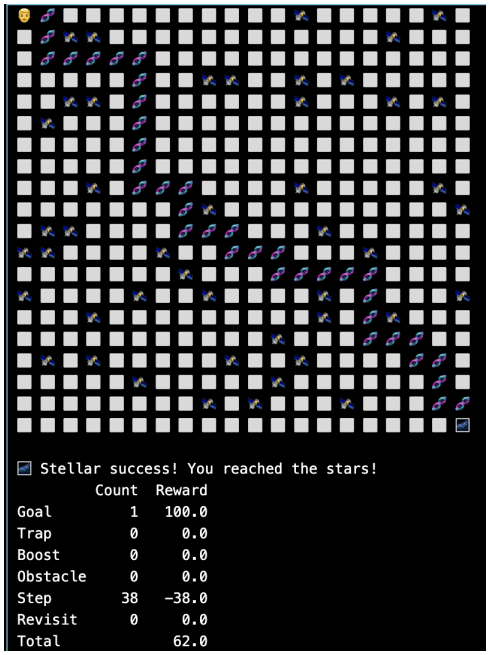
Two scenarios were tested:

2.1 Scenario 1: Traps and Boosts Disabled

The agent learned the optimal path (given by BFS). The number of steps matched exactly with the BFS path, confirming correctness. In both the scenarios, the lengths of the paths taken are the same (39) which is same as the path given by **BFS** algorithm.

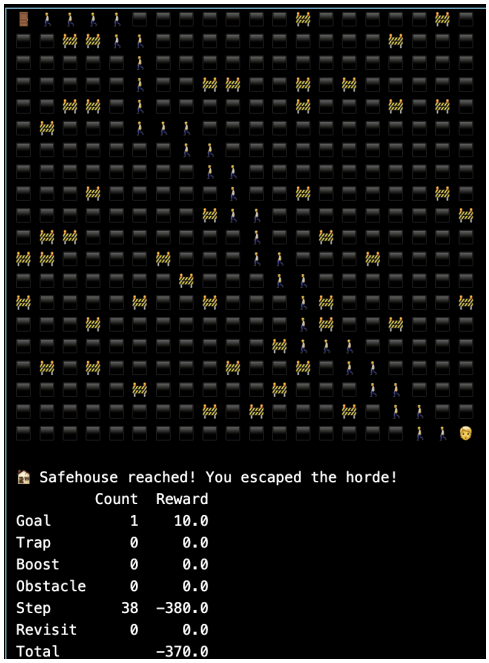
2.1.1 Configuration 1

REWARD_GOAL = 100
REWARD_TRAP = 0
REWARD_OBSTACLE = -100
REWARD_REVISIT = -10
REWARD_ENEMY = -100
REWARD_STEP = -1
REWARD_BOOST = 0
Pickle File Name :- 23634_*disabled*_1.*pkl*



2.1.2 Configuration 2

REWARD_GOAL = 10
REWARD_TRAP = 0
REWARD_OBSTACLE = -10
REWARD_REVISIT = -100
REWARD_ENEMY = -10
REWARD_STEP = -10
REWARD_BOOST = 0
Pickle File Name :- 23634_*disabled*_2.*pkl*

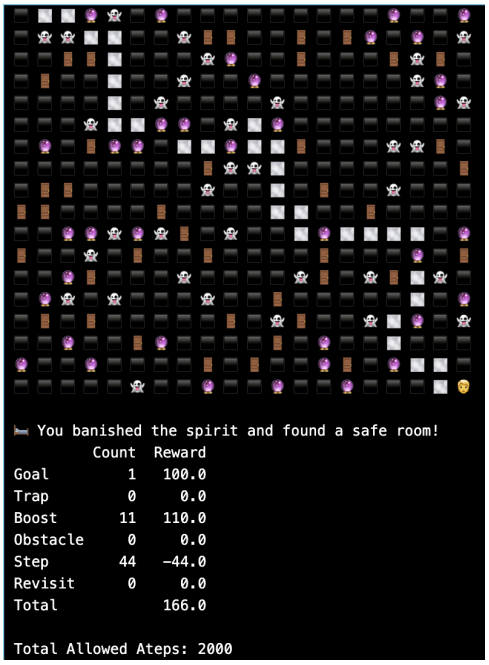


2.2 Scenario 2: Traps and Boosts Enabled

The agent learned to avoid traps and take advantage of boosts. The learned path was different, more reward-efficient, but slightly longer due to new path. Both the paths here are different, the first one with a path length of 45 and the second one with a path length of 39. This difference is observed because of change in REWARD_TRAP and REWARD_BOOST. When the configuration has high REWARD_BOOST, the agent chose paths that are passing through boosts, thereby increasing path lengths. When the configuration has high REWARD_TRAP, the agent aggressively avoided traps,

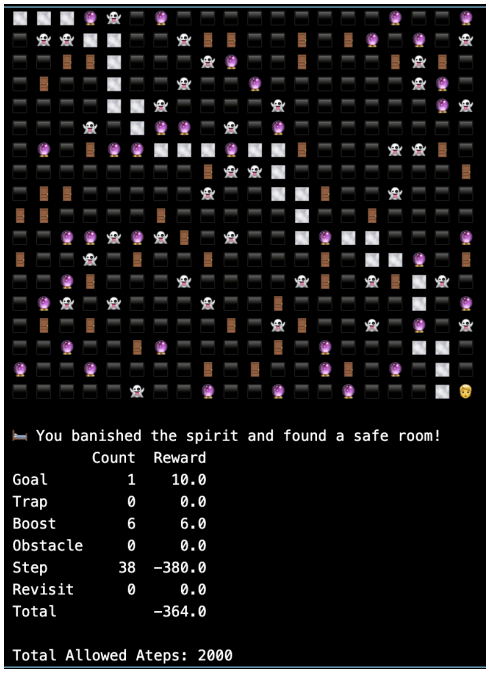
2.2.1 Configuration 1

REWARD_GOAL = 100
REWARD_TRAP = -50
REWARD_OBSTACLE = -100
REWARD_REVISIT = -10
REWARD_ENEMY = -100
REWARD_STEP = -1
REWARD_BOOST = 10
Pickle File Name :- 23634_*enabled*_1.*pkl*



2.2.2 Configuration 2

REWARD_GOAL = 10
REWARD_TRAP = -500
REWARD_OBSTACLE = -10
REWARD_REVISIT = -100
REWARD_ENEMY = -10
REWARD_STEP = -10
REWARD_BOOST = 1
Pickle File Name :- 23634_*enabled*_2.*pkl*



3 Manual Q-value Update for First 5 Steps using Scenario 1 Configuration 2

Using the Q-Table generated after training (with trap and boost disabled), the following manual updates were computed for the first 5 steps of a new episode:

We use the Q-learning update formula:

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r + \gamma \cdot \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right]$$

Given:

- Learning rate $\alpha = 0.7$
- Discount factor $\gamma = 0.9$
- Step reward $r = -10$
- Initial Q-values :-
  State (0,0) :- -847.4282525668123
  State (0,1) :- -836.9633384841969
  State (0,2) :- -801.5731150641324
  State (0,3) :- -780.5693598806217
  State (0,4) :- -773.9093920839804

Step-by-step Calculations

Step 1:

$$Q((0,0), \rightarrow) = Q((0,0), \rightarrow) + 0.7 \left[-10 + 0.9 \cdot \max_a Q((0,1), a) - Q((0,0), \rightarrow) \right] = -846.5447978200666$$

Step 2:

$$Q((0,1),\rightarrow) = Q((0,1),\rightarrow)+0.7\left[-10+0.9\cdot\max_a Q((0,2),a) - Q((0,1),\rightarrow)\right] = -833.6227430271263$$

Step 3:

$$Q((0,2),\rightarrow) = Q((0,2),\rightarrow)+0.7\left[-10+0.9\cdot\max_a Q((0,3),a) - Q((0,2),\rightarrow)\right] = -799.6921701859007$$

Step 4:

$$Q((0,3),\rightarrow) = Q((0,3),\rightarrow)+0.7\left[-10+0.9\cdot\max_a Q((0,4),a) - Q((0,3),\rightarrow)\right] = -780.129453708836$$

Step 5:

$$Q((0,4),\downarrow) = Q((0,4),\downarrow)+0.7\left[-10+0.9\cdot\max_a Q((1,4),a) - Q((0,4),\downarrow)\right] = -772.0625498029146$$

Summary Table

Step	**State**	**Action**	**Reward**	**Next State**	$\max_a Q(s_{t+1}, a)$	**New** $Q(s_t, a_t)$
1	$(0,0)$	$'Right'$	-10	$(0,1)$	-836.9633384841969	-846.5447978200666
2	$(0,1)$	$'Right'$	-10	$(0,2)$	-801.5731150641324	-833.6227430271263
3	$(0,2)$	$'Right'$	-10	$(0,3)$	-780.5693598806217	-799.6921701859007
4	$(0,3)$	$'Right'$	-10	$(0,4)$	-773.9093920839804	-780.1294537088736
5	$(0,4)$	$'Down'$	-10	$(1,4)$	-752.9706760336594	-772.0625498029146