National Taiwan University
Department of Electrical Engineering
Name: An-Che, Liang
ID: B11705009

112-1 (Fall 2023)
Reinforcement Learning

Assignment 3 Report

Q1. Discuss how you speed up your grid world environment. (10 pts.)

To accelerate the speed of gird world environment. I conduct the following two methods, which in total saves around 0.3 seconds on average for the sample test cases.

1. In the function `_get_next_state()`, some if statements could be omitted to save some computation.
2. In the original code, the class itself maintain a map of state index to state coordinate, however, the reverse relationship is not recorded. We can save time for the environment to map the state coordinate back to the state index by recording the reverse relationship in a hash-map.

Q2. What's your best result in 2048? (5 pts.)

Your answer here.

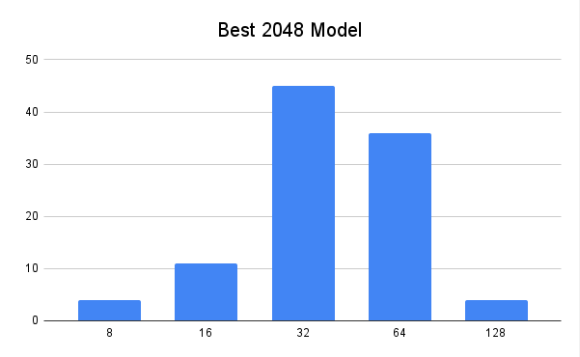

Figure 1: The histogram of the best tile result

Q3. Describe what you have done to train your best model. (15 pts.)

I found out that the PPO algorithm with around 30 epochs will produce a policy with stable performance in a reasonable amount of time, therefore I stick with this set of hyper-parameter. Also, I use the following approach to try to improve the performance, however, the improvement is not significant.

1. Like the instructor suggested, I implemented a mechanism to let the agent try at most 10 illegal moves each time and get the penalty of -100, rather than directly terminate the environment.
2. Also like the instructor suggested, to encourage the agent to put the unpaired tiles at the corner, witch in my experience can improve the score, a weight matrix W is constructed as follow:

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0.1 & 0.3 \\ 0 & 0 & 0.3 & 0.3 \end{bmatrix}$$

   Then the value $\sum W \cdot M$ is added to reward.

3. Finally, to encourage the agent to put tiles with the same value together, I compute the matrix

A, which contains the number of neighboring tiles with the same value as each tiles. Then add $0.4 \times \sum A$ to the reward.

Q4. Choose an environment from the Gymnasium library and train an agent. Show your results or share anything you like. (10 pts.)

Figure 2: Screenshot of environment: LunarLander-v2

The LunarLander-v2 environment is a pre-built environment in the gymnasium/gym package. I use the PPO reinforcement learning algorithm to train the agent to utilize thrusters to land on the designated location.
This environment has the option to render the actual environment like a video game, rather then pure matrices to represent the environment.
It's fun to watch the agent to learn different concepts in action, such as altering left and right thrusters to balance itself, or use the button thruster to decelerate itself from falling too fast.