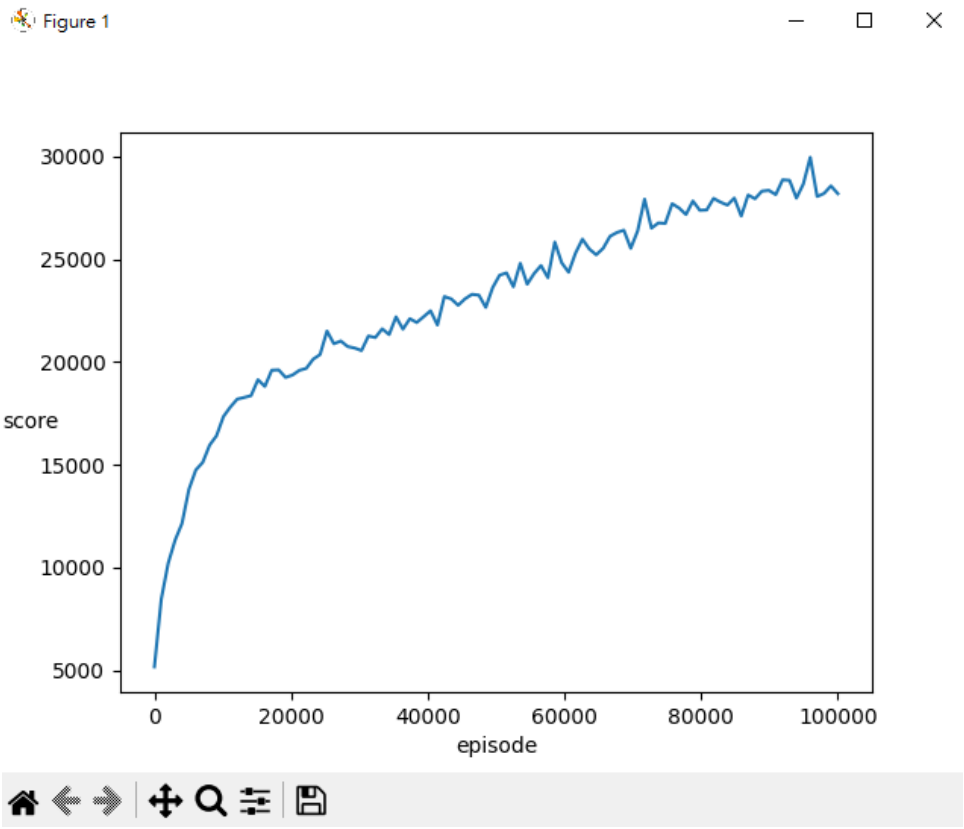
312551163_陳允觀 RL Lab1 Report

Diagram:



Bonus:

1.Describe the implementation and the usage of n-tuple network

While the state of 2048 is large and if we don't use the n-tuple network, we will face most of the state that we haven't seen yet. Therefore, using this method can reduce the memory usage while we train.

2.Explain the mechanism of TD(0). (5%)

TD(0) is a method that can evaluate the value of current state. The value implies the score we can get from the current state to the terminal state. While we play the game from episode to episode, the terminal state's score and those moves' score is modified by the algorithm.

3.Describe your implementation in detail including action selection

and TDbackup diagram.

First, I select the most reward state from the four possible moves. For the TDbackup diagram, I store the moves from the episode just played and modify the error, reward and target values referring to the terminal state's score.
For the plot, I used a vector to store the mean scores of every 1000 episodes and output it to a .csv file. Afterall, I plot the diagram using a python code via matplotlib and the source code is given in the zip file.