

# Project 1 Report

## Theory of Computer Game

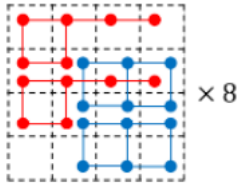
Name: Kai-Jie Lin  
Student ID: 110652019  
October 30, 2022

### 1 Method

The method I use for this game is TD-learning. The feature table is  $8 \times 4 \times 6$ -tuple network with max tile is 14. First, I initialize the value of the table to zero. Then run the game, greedily choose the best policy  $\pi$  which is  $\pi(S_t) = \operatorname{argmax}_a (V(S_t^a))$ .  $V(S_t^a)$  is the value of taking action  $a$  under state  $S_t$ . I stored every state of each game in the episode which is a linear data structure. When the game is end, update the value table:  $V(s'_t) \leftarrow V(s'_t) + \alpha(r_{t+1} + V(s'_{t+1}) - V(s'_t))$  follow the reversed episode.

### 2 Network Design

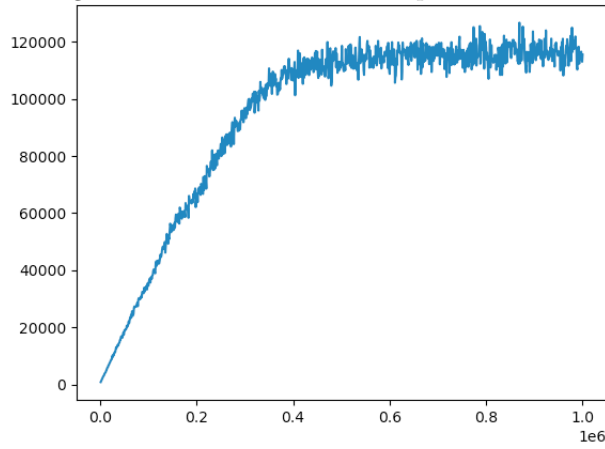
$4 \times 6$ -tuple:



This is the structure of  $4 \times 6$ -tuple. There are eight different direction, so the total feature number is  $8 \times 4$ . I use isomophic pattern to reduce the weight size. The max tile stored in the feature is 16 (3072).

### 3 Experiment

Average score for first one million epoch:



I plot the train log for the first one million epoch on the graph using python. I start training with learning rate equals to 0.003125. And I got near average score 120000. Then I tuned the learning rate to 0.001 and train for another four million epoch and gradually decrease the learning to 0.0005 and finally 0.00005. I have trained for ten million epoch and the average score have reached to nearly 2000000.

Sorry for that I did not present the plot of last nine million epoch. Because I have forgotten to record the train log during the training process, so the training average score is discontinuous. I think this will cause some difficulty to analyze the performance, so I did not present it.

Here is the result:

```
1000    avg = 208386, max = 602247, ops = 415246 (227720|3208619)
48      100%    (0.3%)
96      99.7%  (0.3%)
192     99.4%  (1.1%)
384     98.3%  (0.8%)
768     97.5%  (3.4%)
1536    94.1%  (13.1%)
3072    81%    (78.3%)
6144    2.7%   (2.7%)

Judging the actions... Passed
Judging the speed... Passed, expected 51393 ops
Assessment: 96.6 points
```