深度學習與實務 Lab 8 報告

0556508 翁治平

1. Implementation

AI::get_best_move()

從目前的 state s 往四個方向都走一次，並且計算 reward 和 V(s')的總和，再回傳總和最大者的方向。其中若有任一方向的 reward 為-1，則讓對應的 value 變成一個很小的值，用來忽略該方向。

AI::update_tuple_values()

考慮 experience_buff 的邊界為特例，直接忽略，只取前面的 experience 來進行訓練。按照 TD(0)的 Pseudo Code 計算出$\text{error} = \text{r}_{next} + \text{V}(s'_{next}) - \text{V}(s')$。

2. Discussion

- 無法訓練出預期的模型

  訓練了 2000k 場，maxtile 只能達到 1024，且出現機率低於 0.01%，與助教的實驗結果（400~500k 場可達 95%機率出現 2048）不相符。