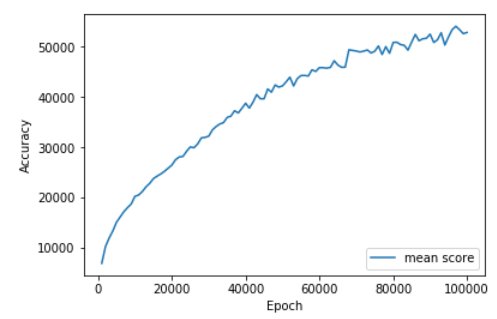
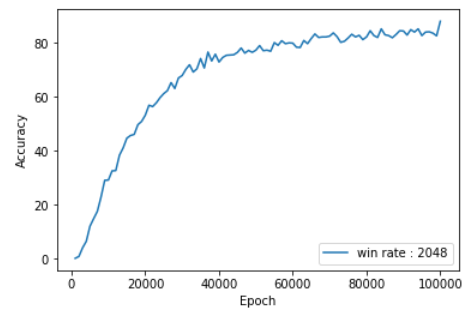
Lab7 : Temporal Difference Learning

0516069 翁英傑

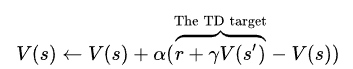
1. Score Plot :

Win rate of 2048 tile mean score

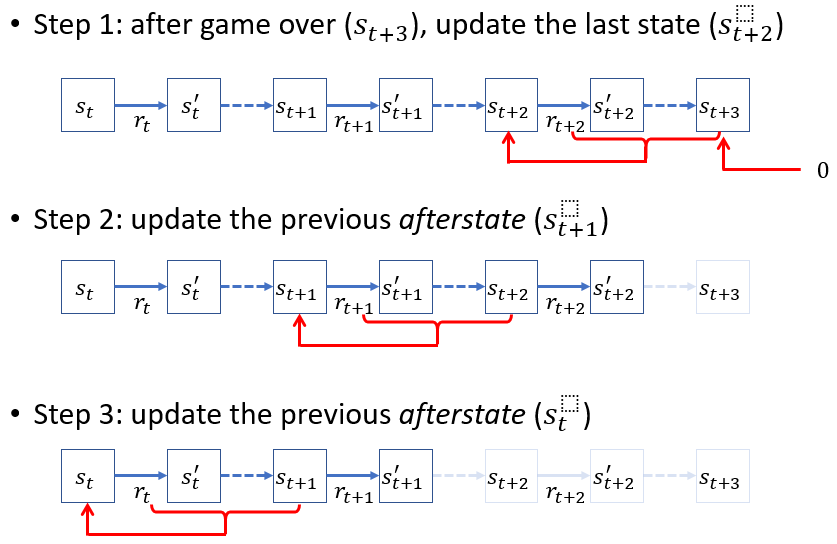


1. Mechanism of TD :

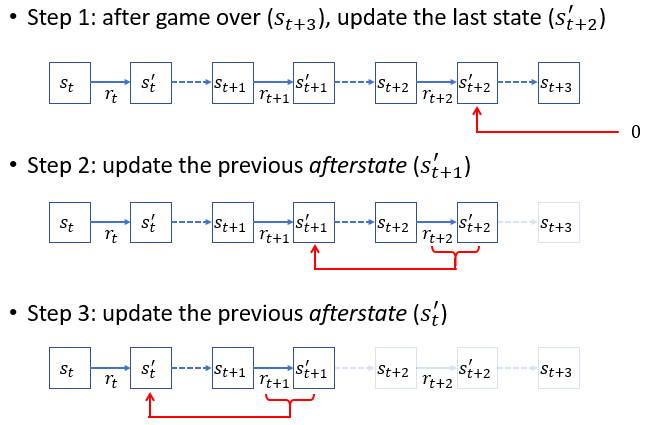
TD learning is to merge the distance of two state to the reward it gets with the action. The formula looks like this:



1. V(state) :



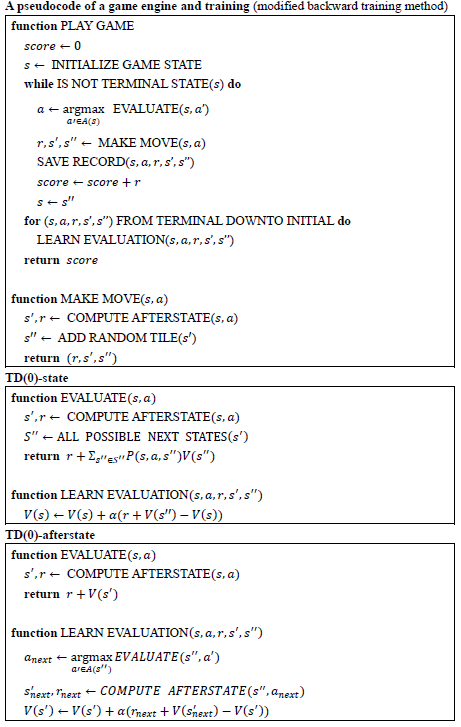
1. V(after-state) :



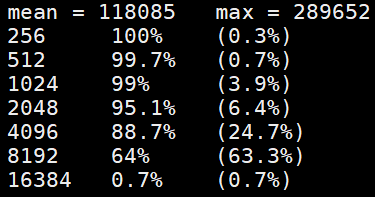
1. Code :

The code is divided into five classes:

* Board: functions of moving the board with up, down, right and left four actions. Also contain the status of the board.
* Pattern: the function for computing the index for storing value of the board with certain pattern.
* Feature: a virtual class for pattern, has the same function as pattern.
* State: contains the information of each state in a play, which contains what action was taken, what reward was gained, what state turned to what state…etc.
* Learning: main methods of td-learning, including determine function for best action of each state, backward update function…etc.



1. Result



1. Discussion

We can see that the value of before-state contains a probability of each possible of after-state, this may cause the score estimation to add an extra variance and leads to a worse performance.