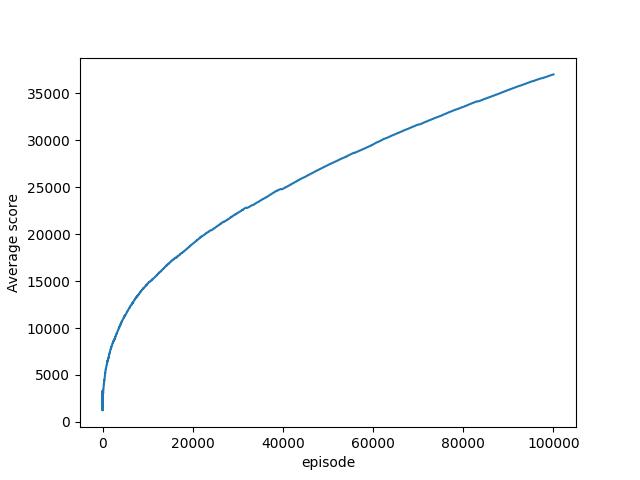
LAB 2 2048TD

謝宇恆 411551022

1. A plot shows scores (mean) of at least 100k training episodes:



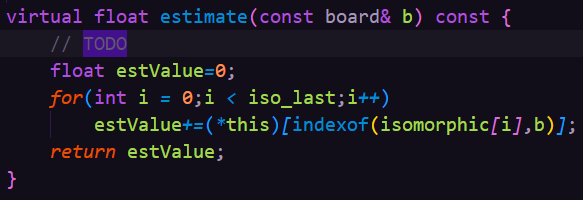
1. Describe the implementation and the usage of n-tuple network

2048的遊戲為4\*4的版面，state的總數為，若是要記錄所有V(s)，需要極大的記憶體。因此需要使用N-tuple Network以降低其所需的記憶體。在實驗中，我們使用到了4\*6-tuple network，這樣的話我們只需要種state，約等於256MB。每一種pattern會有8個isomorphism，並且這8個isomorphism會使用同一個weight table。

1. Explain the mechanism of TD(0)

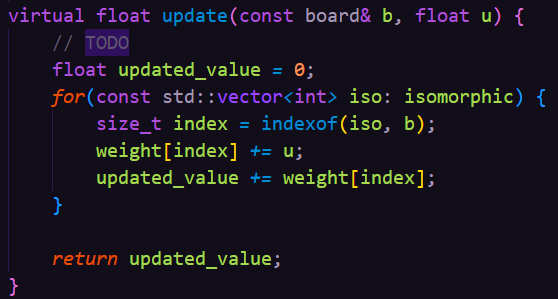
TD(0)的其式子如下: ，其中是對狀態估計能獲得最終結果的獎勵，並且依據實際獎勵值以及下一個狀態估計能獲得最終結果的獎勵值調整。因此TD(0)只需要考慮即可更新。

1. Describe your implementation in detail including action selection and TD-backup diagram.
   1. TODO estimate:



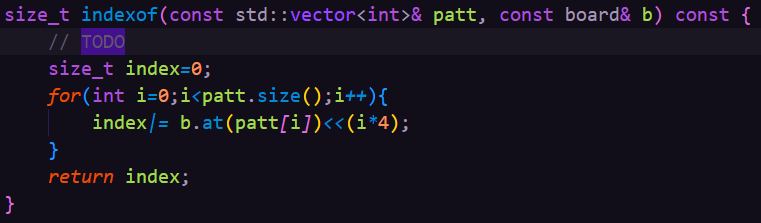
Function estimate是為了計算當前board中Value的期望值。該值等於將每一個feature對應的8個isomorphic加總出來的值。

* 1. TODO update



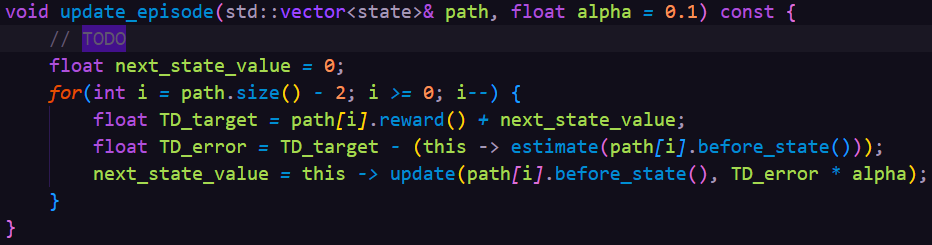
Function update是更新weight table的function，其中u為該原始weight table的TD error乘上learning rate。

* 1. TODO indexof



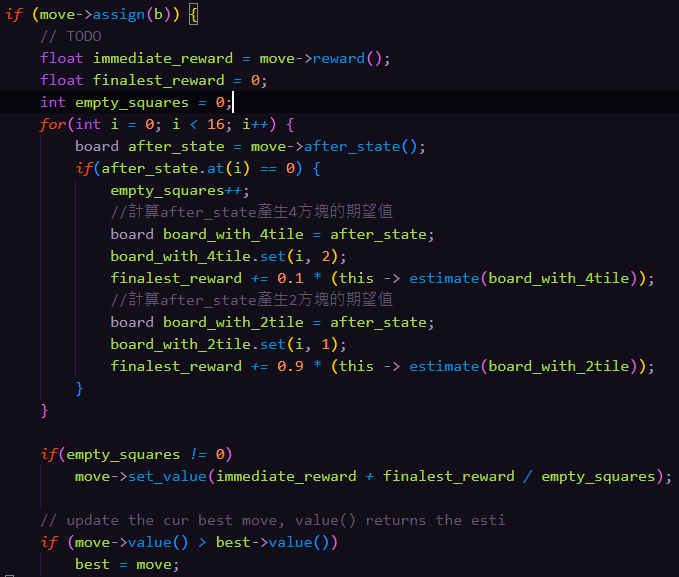
Function indexof是為了將board中每一個pattern儲存成6\*4bit的值，這邊將board中每一個pattern上的值轉換成4bit的數字，並將其合併存入至變數index中。

* 1. TODO update\_episode



Function update\_episode是為了更新V這個函數。因此在每一個episode都會計算TD\_target以及TD\_error，並利用TD\_target以及TD\_error進行V的更新。

* 1. TODO in select\_best\_move



Function select\_best\_move會選擇V()最大的值並進行移動。在這邊我們需要得知V()才能計算V()，所以我們會先進行next state的期望值計算，而next state包含三個可能性: popup一個4的block(有0.1的可能性)、popup一個2的block(有0.9的可能性)以及沒有空間可以動(也就是empty\_square等於0)。在empty\_square不等於0時都需要計算。

* 1. TD-backup diagram

