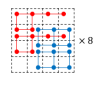
TCG Project2 Report

0716120 林怡安

這次的作業是實作TD(0)-Learning，使用reinforcement learning來讓遊戲能夠透過所創建的網路進而學習更好的走法。

1. Network Design

我原本使用了助教在spec當中提及的network，如下圖。



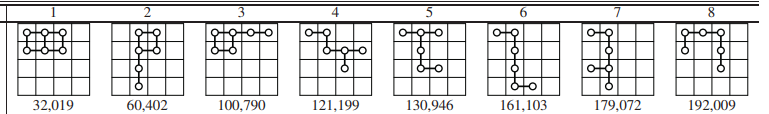
使用這種pattern做8個net，每個net都是這個pattern的上下左右rotation以及對稱後的rotation(正好8種，8個net)。

在training的時候我跑了1M次達到了大約88的成績，後續再跑了1M成績上來到90，但接下來的training卻大概在86-90中來回，所以想說試試其他network。

後來點開了助教在spec當中提及的第二篇reference, K. Matsuzaki, Systematic selection of N-tuple networks with consideration of interinfluence for game 2048, TAAI 2016.。

裡面提及了很多種pattern，也有他自己排名出來覺得最好的pattern network，便想要嘗試看看他的network。

我是選擇他使用6個feature為一個pattern的network，如下圖

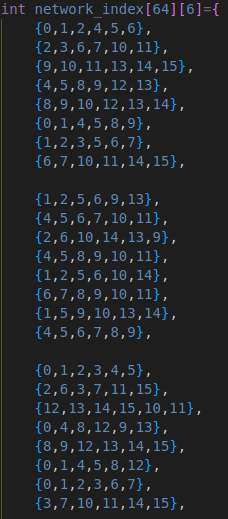


1-8為我所建立的8個network，但每個network都是他們pattern的8種rotation pattern組成(上下左右旋轉以及他們的對稱，總共8種)，所以總共會有8x8=64種features，alpha=1/64。

1. Method Used

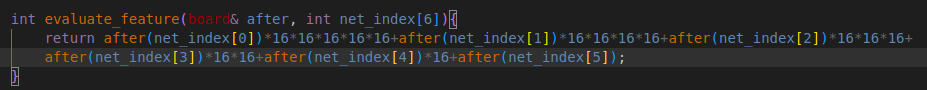
我是先把network的位置都先存到一個2d array裡，方便之後的記算。

如下圖，但由於總共有64列非常長，所以只貼出了一部份。

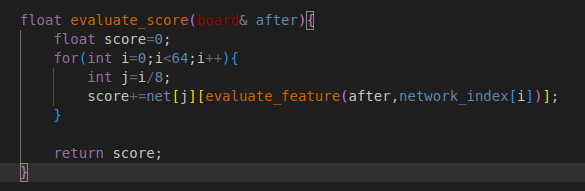


每行皆為一個feature所框到的位置，每8行為一個net總共使用的8個features。

建完network要使用的陣列後，我們接下來需要去算他們features所對應出來的編號。



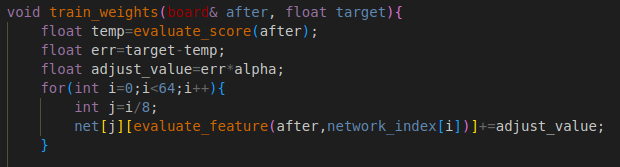
擁有算出features編號的function後，需要一個能計算出此盤面得分為多少的function。



之所以要算score是因為要用來調整network的weight。



而調整weight的方式是記算出此盤面的得分與他下個盤面的得分及reward相差多少，最後再乘上alpha(learning rate)來做調整。

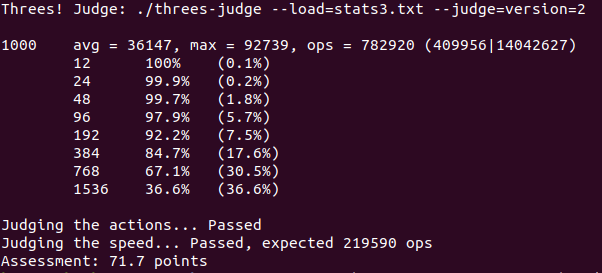




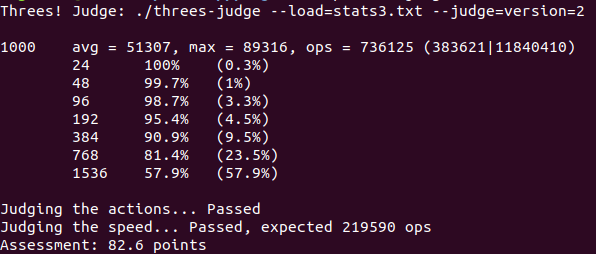
1. Training Process

以下為使用上述network跑出來的分數。

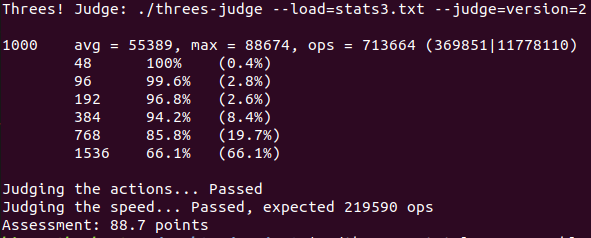
100k training score=71.7



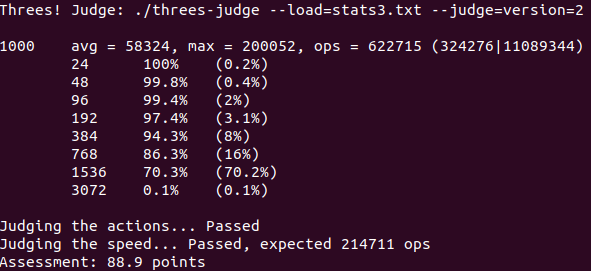
200k training score=82.6



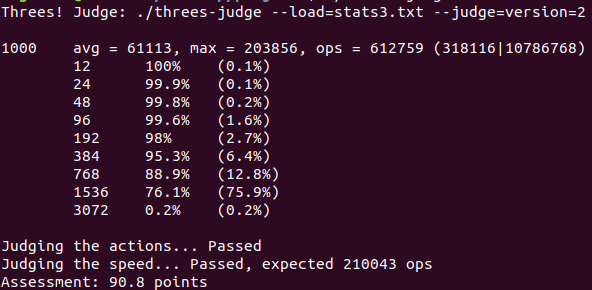
300k training score=88.7



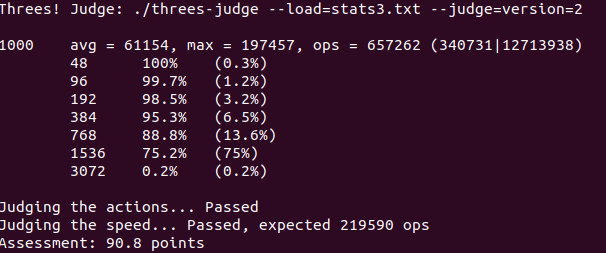
400k training score=88.9



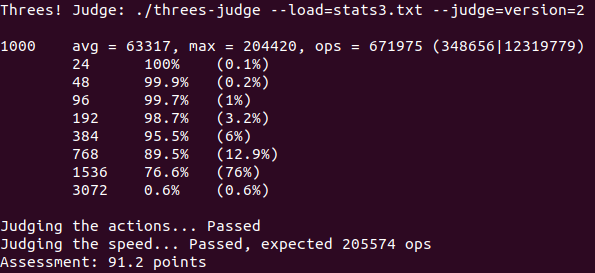
500k training score=90.8



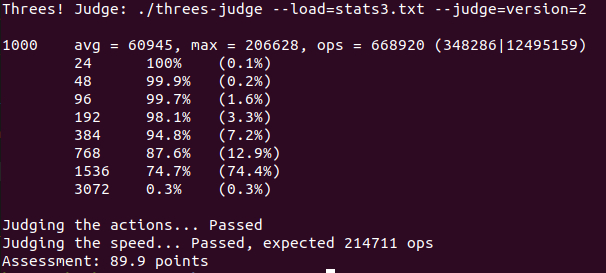
600k training score=90.8



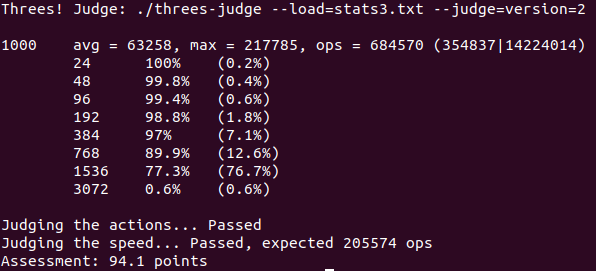
700k training score=91.2



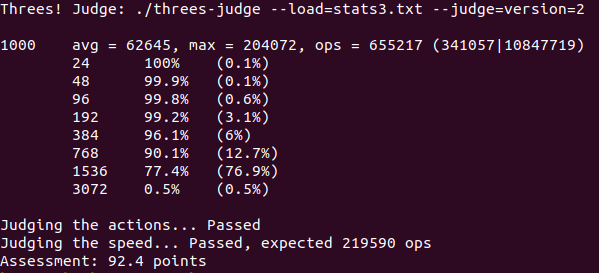
800k training score=89.9



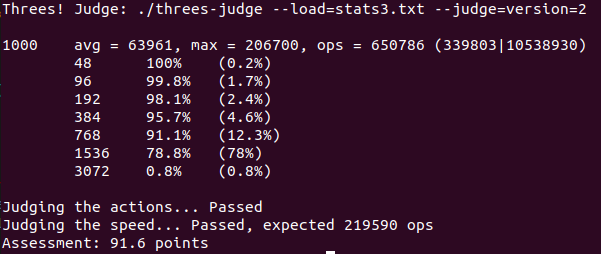
900k training score=94.1



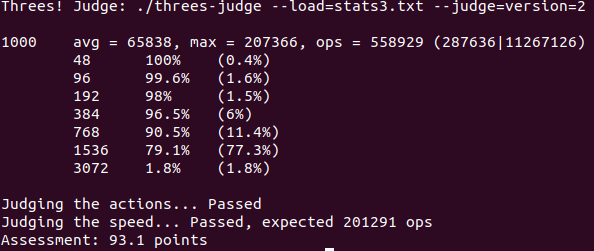
1000k training score=92.4



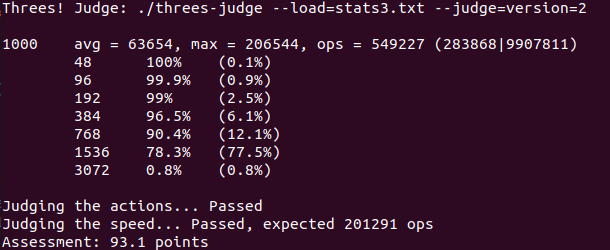
1100k training score=91.6



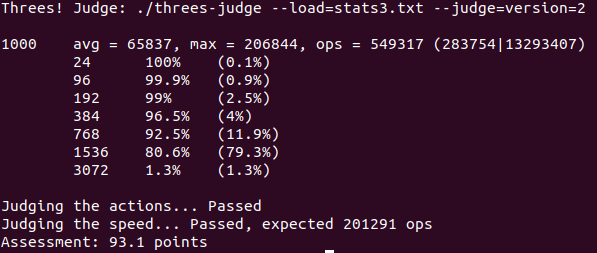
1200k training score=93.1



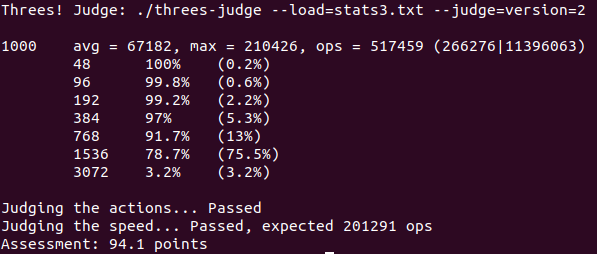
1300k training score=93.1



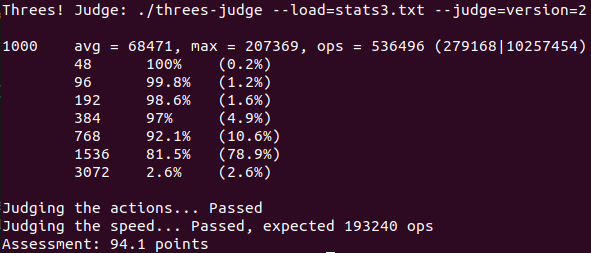
1400k training score=93.1



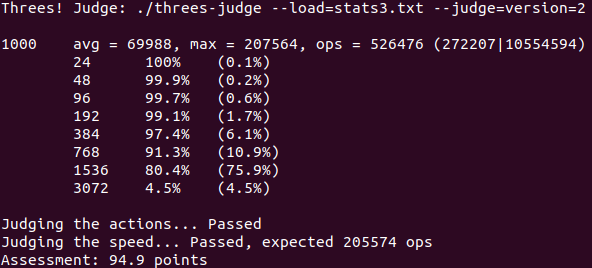
1500k training score=94.1



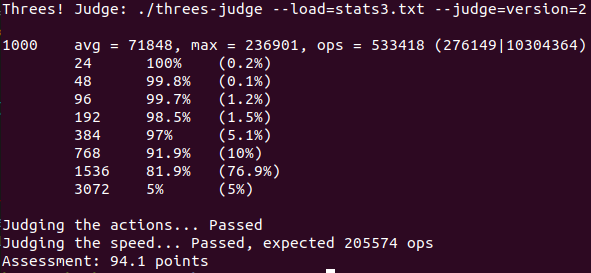
1600k training score=94.1



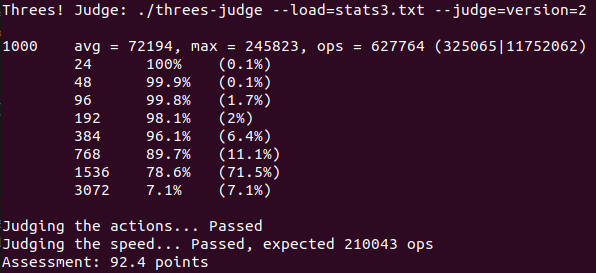
1700k training score=94.9



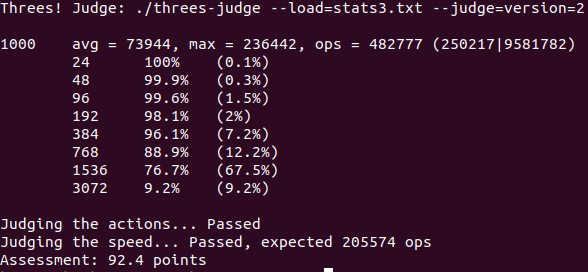
1800k training score=94.1



1900k training score=92.4



2000k training score=92.4



1. Result of Training

透過上面2M次的training process，我們可以發現從一開始的71分經過500k的training就可以很快地來到90分，但從90分開始要有明顯的進步就非常的困難了。不過還是能從得到的數值當中看出network還是有持續再改進的，從3072出現的%數來看，可以發現進入到90分後雖然分數並沒有明顯上升，但3072出現的%數有越來越多。

雖然最後幾次training並沒有持續讓分數變更高，但透過這2M次training我們所得到的最高分數為94.9。