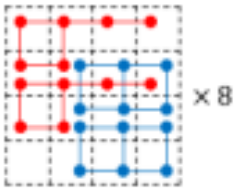
TCG Project2 Report
0716120 林怡安

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。

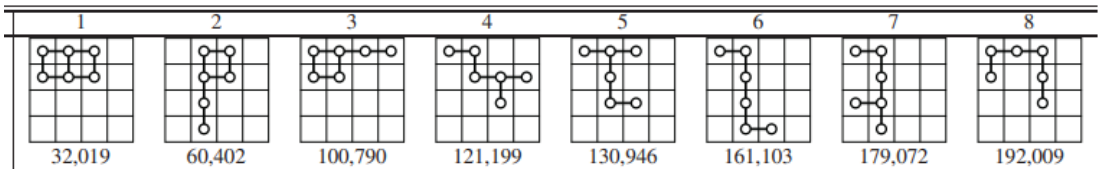
1. Network Design

   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。

2. 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。

```
int network_index[64][6]={
    {0,1,2,4,5,6},
    {2,3,6,7,10,11},
    {9,10,11,13,14,15},
    {4,5,8,9,12,13},
    {8,9,10,12,13,14},
    {0,1,4,5,8,9},
    {1,2,3,5,6,7},
    {6,7,10,11,14,15},

    {1,2,5,6,9,13},
    {4,5,6,7,10,11},
    {2,6,10,14,13,9},
    {4,5,8,9,10,11},
    {1,2,5,6,10,14},
    {6,7,8,9,10,11},
    {1,5,9,10,13,14},
    {4,5,6,7,8,9},

    {0,1,2,3,4,5},
    {2,6,3,7,11,15},
    {12,13,14,15,10,11},
    {0,4,8,12,9,13},
    {8,9,12,13,14,15},
    {0,1,4,5,8,12},
    {0,1,2,3,6,7},
    {3,7,10,11,14,15},
```

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。

```
int evaluate_feature(board& after, int net_index[6]){
    return after(net_index[0])*16*16*16*16*16+after(net_index[1])*16*16*16*16+after(net_index[2])*16*16*16+
    after(net_index[3])*16*16+after(net_index[4])*16+after(net_index[5]);
}
```

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。

```
float evaluate_score(board& after){
    float score=0;
    for(int i=0;i<64;i++){
        int j=i/8;
        score+=net[j][evaluate_feature(after,network_index[i])];
    }

    return score;
}
```

之所以要算 score 是因為要用來調整 network 的 weight。

$$\textbf{TD error: } \Theta[\phi(s'_t)] \leftarrow \Theta[\phi(s'_t)] + \alpha\left(r_{t+1} + V(s'_{t+1}) - V(s'_t)\right).$$

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。

```
void train_weights(board& after, float target){
    float temp=evaluate_score(after);
    float err=target-temp;
    float adjust_value=err*alpha;
    for(int i=0;i<64;i++){
        int j=i/8;
        net[j][evaluate_feature(after,network_index[i])]+=adjust_value;
    }
```

```
virtual void open_episode(const std::string& flag = "") {
    episode.clear();
}
virtual void close_episode(const std::string& flag = "") {
    if(episode.empty()){
        return;
    }
    train_weights(episode[episode.size()-1].after,0);
    for(int i=episode.size()-2;i>=0;i--){
        train_weights(episode[i].after,episode[i+1].reward+evaluate_score(episode[i+1].after));
    }
}
```

3. Training Process

以下為使用上述 network 跑出來的分數。

100k training score=71.7

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 36147, max = 92739, ops = 782920 (409956|14042627)
        12      100%    (0.1%)
        24      99.9%   (0.2%)
        48      99.7%   (1.8%)
        96      97.9%   (5.7%)
        192     92.2%   (7.5%)
        384     84.7%   (17.6%)
        768     67.1%   (30.5%)
        1536    36.6%   (36.6%)

Judging the actions... Passed
Judging the speed... Passed, expected 219590 ops
Assessment: 71.7 points
```

200k training score=82.6

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 51307, max = 89316, ops = 736125 (383621|11840410)
        24      100%    (0.3%)
        48      99.7%   (1%)
        96      98.7%   (3.3%)
        192     95.4%   (4.5%)
        384     90.9%   (9.5%)
        768     81.4%   (23.5%)
        1536    57.9%   (57.9%)

Judging the actions... Passed
Judging the speed... Passed, expected 219590 ops
Assessment: 82.6 points
```

300k training score=88.7

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 55389, max = 88674, ops = 713664 (369851|11778110)
        48      100%    (0.4%)
        96      99.6%   (2.8%)
        192     96.8%   (2.6%)
        384     94.2%   (8.4%)
        768     85.8%   (19.7%)
        1536    66.1%   (66.1%)

Judging the actions... Passed
Judging the speed... Passed, expected 219590 ops
Assessment: 88.7 points
```

400k training score=88.9

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 58324, max = 200052, ops = 622715 (324276|11089344)
        24      100%    (0.2%)
        48      99.8%   (0.4%)
        96      99.4%   (2%)
        192     97.4%   (3.1%)
        384     94.3%   (8%)
        768     86.3%   (16%)
        1536    70.3%   (70.2%)
        3072    0.1%    (0.1%)

Judging the actions... Passed
Judging the speed... Passed, expected 214711 ops
Assessment: 88.9 points
```

500k training score=90.8

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 61113, max = 203856, ops = 612759 (318116|10786768)
        12      100%    (0.1%)
        24      99.9%   (0.1%)
        48      99.8%   (0.2%)
        96      99.6%   (1.6%)
        192     98%     (2.7%)
        384     95.3%   (6.4%)
        768     88.9%   (12.8%)
        1536    76.1%   (75.9%)
        3072    0.2%    (0.2%)

Judging the actions... Passed
Judging the speed... Passed, expected 210043 ops
Assessment: 90.8 points
```

600k training score=90.8

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 61154, max = 197457, ops = 657262 (340731|12713938)
        48      100%    (0.3%)
        96      99.7%   (1.2%)
        192     98.5%   (3.2%)
        384     95.3%   (6.5%)
        768     88.8%   (13.6%)
        1536    75.2%   (75%)
        3072    0.2%    (0.2%)

Judging the actions... Passed
Judging the speed... Passed, expected 219590 ops
Assessment: 90.8 points
```

700k training score=91.2

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 63317, max = 204420, ops = 671975 (348656|12319779)
        24      100%    (0.1%)
        48      99.9%   (0.2%)
        96      99.7%   (1%)
        192     98.7%   (3.2%)
        384     95.5%   (6%)
        768     89.5%   (12.9%)
        1536    76.6%   (76%)
        3072    0.6%    (0.6%)

Judging the actions... Passed
Judging the speed... Passed, expected 205574 ops
Assessment: 91.2 points
```

800k training score=89.9

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 60945, max = 206628, ops = 668920 (348286|12495159)
        24      100%    (0.1%)
        48      99.9%   (0.2%)
        96      99.7%   (1.6%)
        192     98.1%   (3.3%)
        384     94.8%   (7.2%)
        768     87.6%   (12.9%)
        1536    74.7%   (74.4%)
        3072    0.3%    (0.3%)

Judging the actions... Passed
Judging the speed... Passed, expected 214711 ops
Assessment: 89.9 points
```

900k training score=94.1

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 63258, max = 217785, ops = 684570 (354837|14224014)
        24      100%    (0.2%)
        48      99.8%   (0.4%)
        96      99.4%   (0.6%)
        192     98.8%   (1.8%)
        384     97%     (7.1%)
        768     89.9%   (12.6%)
        1536    77.3%   (76.7%)
        3072    0.6%    (0.6%)

Judging the actions... Passed
Judging the speed... Passed, expected 205574 ops
Assessment: 94.1 points
```

1000k training score=92.4

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 62645, max = 204072, ops = 655217 (341057|10847719)
        24      100%    (0.1%)
        48      99.9%   (0.1%)
        96      99.8%   (0.6%)
        192     99.2%   (3.1%)
        384     96.1%   (6%)
        768     90.1%   (12.7%)
        1536    77.4%   (76.9%)
        3072    0.5%    (0.5%)

Judging the actions... Passed
Judging the speed... Passed, expected 219590 ops
Assessment: 92.4 points
```

1100k training score=91.6

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 63961, max = 206700, ops = 650786 (339803|10538930)
        48      100%    (0.2%)
        96      99.8%   (1.7%)
        192     98.1%   (2.4%)
        384     95.7%   (4.6%)
        768     91.1%   (12.3%)
        1536    78.8%   (78%)
        3072    0.8%    (0.8%)

Judging the actions... Passed
Judging the speed... Passed, expected 219590 ops
Assessment: 91.6 points
```

1200k training score=93.1

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 65838, max = 207366, ops = 558929 (287636|11267126)
        48      100%    (0.4%)
        96      99.6%   (1.6%)
        192     98%     (1.5%)
        384     96.5%   (6%)
        768     90.5%   (11.4%)
        1536    79.1%   (77.3%)
        3072    1.8%    (1.8%)

Judging the actions... Passed
Judging the speed... Passed, expected 201291 ops
Assessment: 93.1 points
```

1300k training score=93.1

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 63654, max = 206544, ops = 549227 (283868|9907811)
        48      100%    (0.1%)
        96      99.9%   (0.9%)
        192     99%     (2.5%)
        384     96.5%   (6.1%)
        768     90.4%   (12.1%)
        1536    78.3%   (77.5%)
        3072    0.8%    (0.8%)

Judging the actions... Passed
Judging the speed... Passed, expected 201291 ops
Assessment: 93.1 points
```

1400k training score=93.1

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 65837, max = 206844, ops = 549317 (283754|13293407)
        24      100%    (0.1%)
        96      99.9%   (0.9%)
        192     99%     (2.5%)
        384     96.5%   (4%)
        768     92.5%   (11.9%)
        1536    80.6%   (79.3%)
        3072    1.3%    (1.3%)

Judging the actions... Passed
Judging the speed... Passed, expected 201291 ops
Assessment: 93.1 points
```

1500k training score=94.1

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 67182, max = 210426, ops = 517459 (266276|11396063)
        48      100%    (0.2%)
        96      99.8%   (0.6%)
        192     99.2%   (2.2%)
        384     97%     (5.3%)
        768     91.7%   (13%)
        1536    78.7%   (75.5%)
        3072    3.2%    (3.2%)

Judging the actions... Passed
Judging the speed... Passed, expected 201291 ops
Assessment: 94.1 points
```

1600k training score=94.1

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 68471, max = 207369, ops = 536496 (279168|10257454)
        48      100%    (0.2%)
        96      99.8%   (1.2%)
        192     98.6%   (1.6%)
        384     97%     (4.9%)
        768     92.1%   (10.6%)
        1536    81.5%   (78.9%)
        3072    2.6%    (2.6%)

Judging the actions... Passed
Judging the speed... Passed, expected 193240 ops
Assessment: 94.1 points
```

1700k training score=94.9

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 69988, max = 207564, ops = 526476 (272207|10554594)
        24      100%    (0.1%)
        48      99.9%   (0.2%)
        96      99.7%   (0.6%)
        192     99.1%   (1.7%)
        384     97.4%   (6.1%)
        768     91.3%   (10.9%)
        1536    80.4%   (75.9%)
        3072    4.5%    (4.5%)

Judging the actions... Passed
Judging the speed... Passed, expected 205574 ops
Assessment: 94.9 points
```

1800k training score=94.1

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 71848, max = 236901, ops = 533418 (276149|10304364)
        24      100%    (0.2%)
        48      99.8%   (0.1%)
        96      99.7%   (1.2%)
        192     98.5%   (1.5%)
        384     97%     (5.1%)
        768     91.9%   (10%)
        1536    81.9%   (76.9%)
        3072    5%      (5%)

Judging the actions... Passed
Judging the speed... Passed, expected 205574 ops
Assessment: 94.1 points
```

1900k training score=92.4

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 72194, max = 245823, ops = 627764 (325065|11752062)
        24      100%    (0.1%)
        48      99.9%   (0.1%)
        96      99.8%   (1.7%)
        192     98.1%   (2%)
        384     96.1%   (6.4%)
        768     89.7%   (11.1%)
        1536    78.6%   (71.5%)
        3072    7.1%    (7.1%)

Judging the actions... Passed
Judging the speed... Passed, expected 210043 ops
Assessment: 92.4 points
```

2000k training score=92.4

```
Threes! Judge: ./threes-judge --load=stats3.txt --judge=version=2

1000    avg = 73944, max = 236442, ops = 482777 (250217|9581782)
        24      100%    (0.1%)
        48      99.9%   (0.3%)
        96      99.6%   (1.5%)
        192     98.1%   (2%)
        384     96.1%   (7.2%)
        768     88.9%   (12.2%)
        1536    76.7%   (67.5%)
        3072    9.2%    (9.2%)

Judging the actions... Passed
Judging the speed... Passed, expected 205574 ops
Assessment: 92.4 points
```

4. 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。