

CSC411: Assignment 4 Bonus

Due on Monday, Apr. 2, 2018

Zhongtian Ouyang/Yihao Ni

Problem 1

X and O

This is a harder problem than project4. The policy now has to consider whether its turn is 1 or 2. The win rates are calculated by playing 400 games. The best win rate when it moves first is around 95% and the best win rate when it moves second is around 70%. The best episode is chosen based on the average of the two win rates. The best episode is 80000.

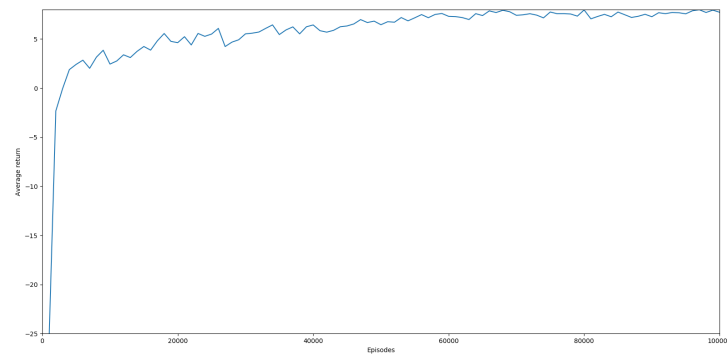


Figure 1: training curve

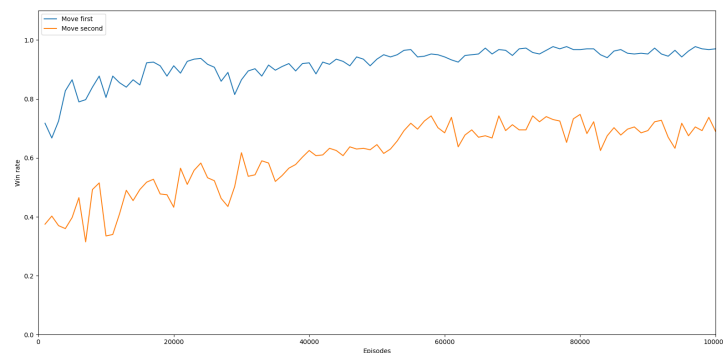


Figure 2: Winrate

Policy plays first

===== Game1 =====

X..

...

...

=====

X..

.O.

...

=====

X..

```
xO.
...
```

```
=====
```

```
xO.
xO.
...
```

```
=====
```

```
xO.
xO.
X..
```

```
=====
```

Learned policy wins against random! (learned policy moves first)

```
===== Game2 =====
```

```
X..
...
...
```

```
=====
```

```
xO.
...
...
```

```
=====
```

```
xO.
X..
...
```

```
=====
```

```
xO.
X..
..O
```

```
=====
```

```
xO.
X..
X.O
```

```
=====
```

Learned policy wins against random! (learned policy moves first)

```
===== Game1 =====
```

```
X..
...
...
```

```
=====
```

```
X..
O..
...
```

```
=====
```

```
X..
O..
X..
```

```
=====
```

```
X..
OO.
X..
```

```
=====
```

```
X..
OOX
X..
```

```
=====
```

```
X..
OOX
```

```
x.o
```

```
=====
```

```
xx.
```

```
oox
```

```
x.o
```

```
=====
```

```
xxo
```

```
oox
```

```
x.o
```

```
=====
```

```
xxo
```

```
oox
```

```
xxo
```

```
=====
```

```
Learned policy ties against random! (random moves first)
```

```
===== Game2 =====
```

```
x..
```

```
...
```

```
...
```

```
=====
```

```
x..
```

```
o..
```

```
...
```

```
=====
```

```
x..
```

```
o..
```

```
..x
```

```
=====
```

```
x..
```

```
oo.
```

```
..x
```

```
=====
```

```
x..
```

```
oox
```

```
..x
```

```
=====
```

```
x.o
```

```
oox
```

```
..x
```

```
=====
```

```
xxo
```

```
oox
```

```
..x
```

```
=====
```

```
xxo
```

```
oox
```

```
.ox
```

```
=====
```

```
xxo
```

```
oox
```

```
xox
```

```
=====
```

```
Learned policy ties against random! (random moves first)
```

```
===== Game3 =====
```

```
x..
```

```
...
```

```
...
```

```
=====
```

```
x . .  
o . .  
. . .  
=====
```

```
x . .  
o . .  
. . x  
=====
```

```
x . .  
oo .  
. . x  
=====
```

```
x . .  
oox  
. . x  
=====
```

```
x . o  
oox  
. . x  
=====
```

```
x . o  
oox  
. xx  
=====
```

```
x . o  
oox  
oxx  
=====
```

Learned policy wins against random! (random moves first)

Problem 2

Self-play

The winrate for both playing first and playing second doesn't really increase. In fact they seems to decrease a bit. So I use the weights at 60000 episode for testing

The graphs also include the win and tie rate because I feel like for tictactoe game, tie is inevitable some time.

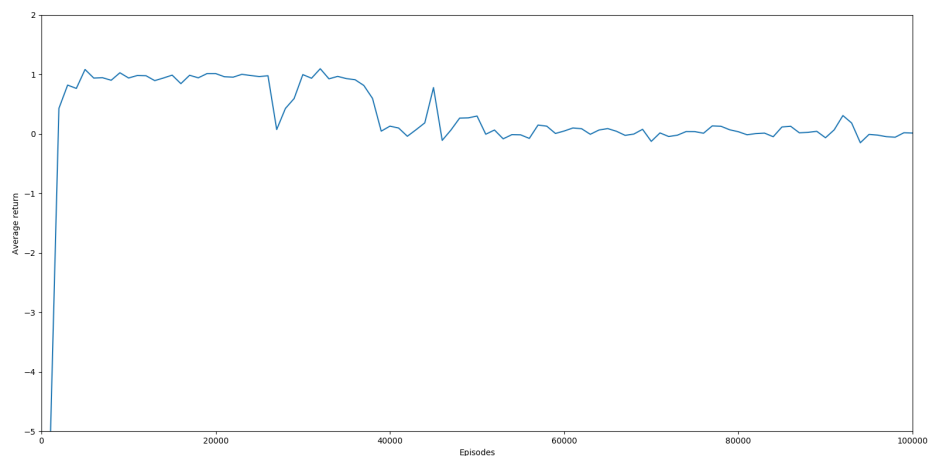


Figure 3: Learning curve

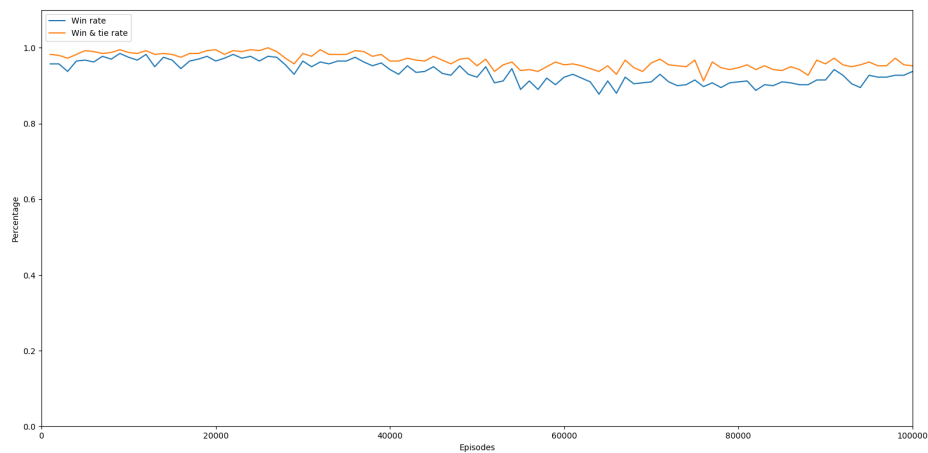


Figure 4: Win and tie rate for playing first for 400 games

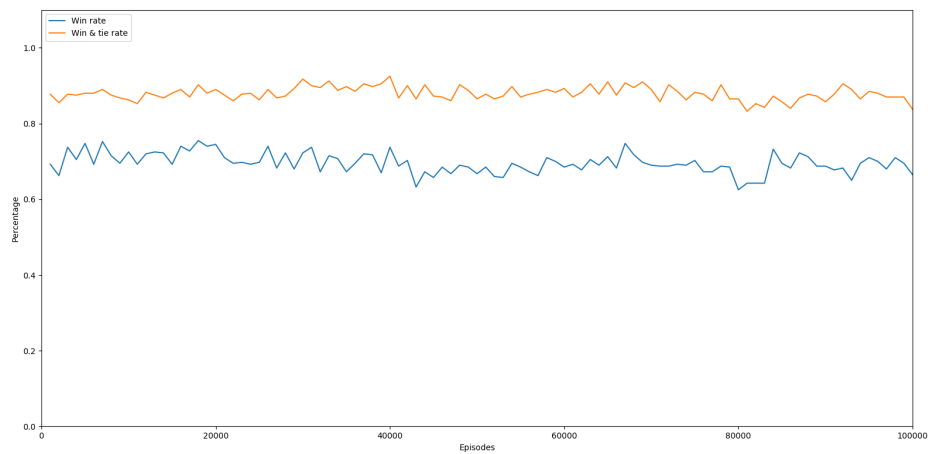


Figure 5: Win and tie rate for playing second for 400 games

Against random player

==== Game1 =====

...
 .X.
 ...

...
 .XO
 ...

X..
 .XO
 ...

X..
 OXO
 ...

X..
 OXO
 ..X

Learned policy wins against random! (learned policy moves first)

==== Game2 =====

...
 .X.
 ...

.O.
 .X.
 ...

XO.

```
. x .
. . .
```

```
=====
x o .
. x .
. o .
```

```
=====
x o .
. x .
x o .
```

```
=====
x o .
o x .
x o .
```

```
=====
x o x
o x .
x o .
```

```
=====
Learned policy wins against random! (learned policy moves first)
```

```
===== Game1 =====
```

```
. . .
. x .
. . .
```

```
=====
o . .
. x .
. . .
```

```
=====
o . .
. x x
. . .
```

```
=====
o . .
o x x
. . .
```

```
=====
o . .
o x x
. . x
```

```
=====
o . .
o x x
o . x
```

```
=====
Learned policy wins against random! (random moves first)
```

```
===== Game2 =====
```

```
. . .
. x .
. . .
```

```
=====
o . .
. x .
. . .
```

```
=====
o . .
. x .
. . x
```



```
=====
O..
OX.
..X
=====
```

```
OX.
OX.
..X
=====
```

```
OX.
OX.
O.X
=====
```

Learned policy wins against random! (random moves first)

===== Game3 =====

```
...
.X.
...
=====
```

```
O..
.X.
...
=====
```

```
O..
.X.
X..
=====
```

```
O.O
.X.
X..
=====
```

```
O.O
.XX
X..
=====
```

```
O.O
OXX
X..
=====
```

```
OXO
OXX
X..
=====
```

```
OXO
OXX
XO.
=====
```

```
OXO
OXX
XOX
=====
```

Learned policy ties against random! (random moves first)

Against self

===== Game1 =====

X . .
 . . .
 . . .
 =====

X . .
 . . .
 O . .
 =====

X . .
 . . .
 O . X
 =====

X . .
 . O .
 O . X
 =====

X . X
 . O .
 O . X
 =====

X O X
 . O .
 O . X
 =====

X O X
 . O X
 O . X
 =====

First player wins!

===== Game2 =====

. . .
 . X .
 . . .
 =====

O . .
 . X .
 . . .
 =====

O . .
 . X .
 . . X
 =====

O . .
 O X .
 . . X
 =====

O . .
 O X .
 X . X
 =====

O . .
 O X .
 X O X
 =====

O . X
 O X .
 X O X

```
=====
First player wins!
```

```
===== Game3 =====
```

```
...
.X.
...
```

```
=====
O..
.X.
...
```

```
=====
O..
.X.
..X
```

```
=====
O..
OX.
..X
```

```
=====
O..
OX.
X.X
```

```
=====
O..
OX.
XOX
```

```
=====
O.X
OX.
XOX
```

```
=====
First player wins!
```

```
===== Game4 =====
```

```
...
.X.
...
```

```
=====
O..
.X.
...
```

```
=====
O..
XX.
...
```

```
=====
O..
XXO
...
```

```
=====
O..
XXO
..X
```

```
=====
O..
XXO
.OX
```

```
=====
O..
```

```

xxo
xox

```

```

o.o
xxo
xox

```

```

oxo
xox
xox

```

```

Tie.

```

```

===== Game5 =====

```

```

...
.x.
...

```

```

o..
.x.
...

```

```

o..
.x.
..x

```

```

o..
ox.
..x

```

```

o..
ox.
x.x

```

```

o..
ox.
xox

```

```

o.x
ox.
xox

```

```

First player wins!

```