Multi Agent learning on N-player Prisoners Dilemma

Prathyoom Sridharan, Ashwath Srikanthan

December 2019

1 Idea

We trained multiple agents by playing N-player prisoners dilemma game against each other using deep reinforcement learning method.
The agent looks at previous 15 actions of each player and decides the current move. State is defined by the previous 15 moves combined. Each episode consists of T rounds where the agent learns, and changes its weights according the reward provided after each round. M episodes are played, and the state is refreshed after each game.

2 Results

We are displaying the scores of player 1 where N=5.

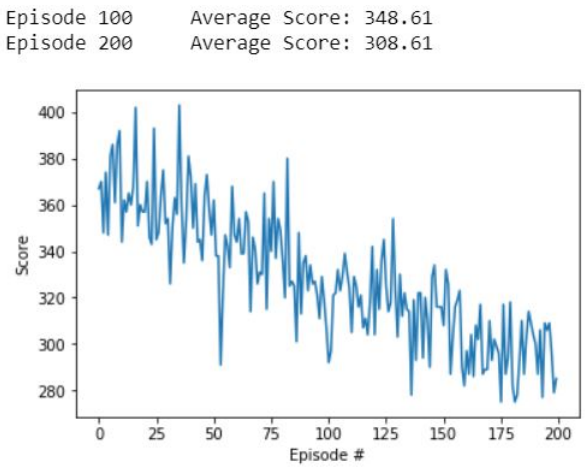

Figure 1: M=200 T=50

```
Episode 100     Average Score: 350.63
Episode 200     Average Score: 308.83
Episode 300     Average Score: 284.89
Episode 400     Average Score: 271.62
Episode 500     Average Score: 262.36
Episode 600     Average Score: 258.23
Episode 700     Average Score: 254.86
Episode 800     Average Score: 252.45
Episode 900     Average Score: 252.43
Episode 1000    Average Score: 256.59
Episode 1100    Average Score: 258.44
Episode 1200    Average Score: 265.57
Episode 1300    Average Score: 271.05
Episode 1400    Average Score: 288.22
Episode 1500    Average Score: 325.70
Episode 1600    Average Score: 338.22
Episode 1700    Average Score: 341.80
Episode 1800    Average Score: 340.12
Episode 1900    Average Score: 319.95
Episode 2000    Average Score: 321.90
```

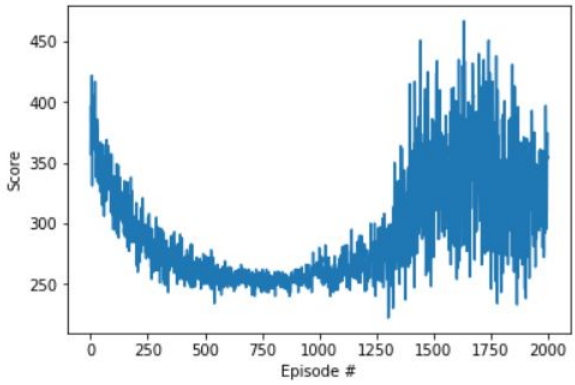

Figure 2: M=2000 T=50

```
Episode 9800    Average Score: 251.28
Episode 9900    Average Score: 251.28
Episode 10000   Average Score: 251.09
```

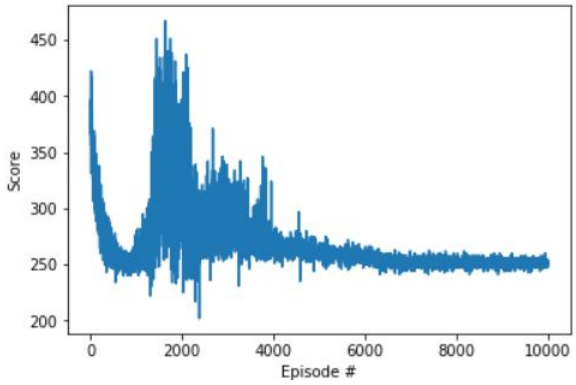

Figure 3: M=10000 T=50

3 Inference

We noticed initially the models were competing with each other, resulting in gradual reduction in score. Later, models experimented cooperating for a possible uptick. When we ran it even further (refer figure 3), we noticed a convergence to a particular score better than all cheat.