Controlling 3D gaming agents in an adversarial setting with Deep Reinforcement Learning

Mehmood Munir
p176075@nu.edu.pk

Bashir Ahmed
p176079@nu.edu.pk

M. Hanzaila
p180453@nu.edu.pk

Supervisor
Dr.Muhammad Nauman

January 6, 2021

Table of contents

Q Learning

Q Learning

$$\underbrace{Q(s,a)}_{\text{New Q-Value}} = Q(s,a) + \underset{\text{Learning Rate}}{\alpha}\,[\underbrace{R(s,a,s')}_{\text{Reward}} + \underset{\text{Discount rate}}{\gamma}\,\overbrace{\max Q'(s',a')}^{\substack{\text{Maximum predicted reward, given}\\ \text{new state and all possible actions}}} - Q(s,a)]$$

Sample = R(s,a,s') + $\gamma(maxQ(s',a'))$

Q(s,a) = (1-α)$Q(s,a) + \alpha(Sample)$

Grid World

Grid World

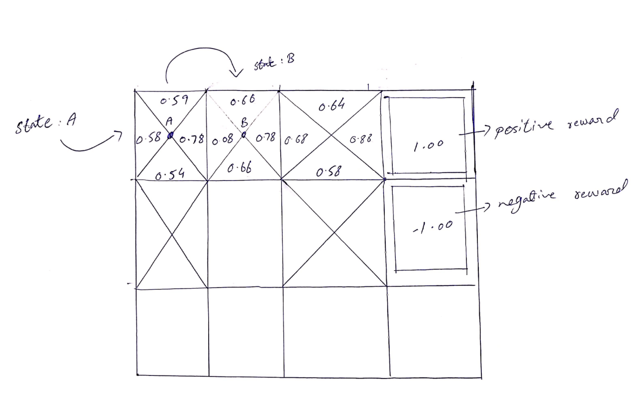

Figure 1: Grid World

Calculation

Calculation

R(s,a,s') = 1

s = state(A),
s' = state(B) next state,
a = move right(Random Action)

Q-Values with different actions 0.66, 0.40, 0.78, 0.66

max Q(s',a') = max (0.66, 0.40, 0.78, 0.66) = 0.78

Sample = R(s,a,s') + $\gamma(maxQ(s', a'))$

Sample = 1 + 1(0.78) = 1.78

Q(s,a) = (1-$\alpha)Q(s, a) + \alpha(Sample)...(A)$

Let $\alpha = 1$
putting values in equation (A)

Q(s,a) = (1 - 1)(0.78) + 1(1.78)

Q(s,a) = 1.78

Bridge Design Pattern

Bridge Design Pattern

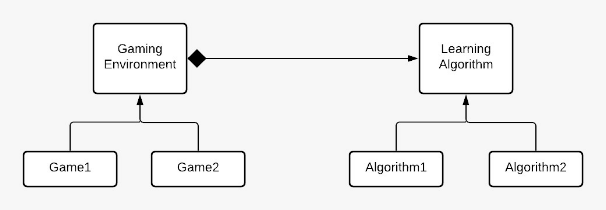

Figure 2: Bridge Design Pattern

Coding Convention

Coding Convention

- Tensorflow v2 vs tensorflow v1

Coding Convention

- Tensorflow v2 vs tensorflow v1
- Snake case naming convention

Coding Convention

- Tensorflow v2 vs tensorflow v1
- Snake case naming convention

Airstriker

Airstriker

- State Space: 215040
- Action Space : 12

CartPole

CartPole

- State Space: 16
- Action Space : 2

Slimevolleygym

Slimevolleygym

- State Space: 12
- Action Space : 3

Questions?