

作業三 Game Playing

劉彥廷 B03902036

1. 模型敘述

1.1. Policy Gradient

模型參照了 [1] 的介紹。由兩層 fully-connected layer 所構成，第一層由 200 個單元所構成，第二層則對應至輸出的動作數量（強制縮限到只剩下「Up」跟「Down」兩種動作）。

Figure 1: 學習曲線

1.2. DQN

Figure 2: 學習曲線

2. 優化方式

本次作業使用了 teacher force（請參閱前述的文章）的方式來減少訓練的複雜度。

3. 結果

本次作業並沒有實作完畢 attention model，加上花了太久的時間在處理 dataset，沒有足夠的時間訓練出合適的結果。

4. 參考文獻

[1] Deep reinforcement learning: Pong from pixels. <http://karpathy.github.io/2016/05/31/r1/>. (Accessed on 12/16/2017).