

HW4-2 比較報告：Double DQN 與 Dueling DQN

一、演算法簡介

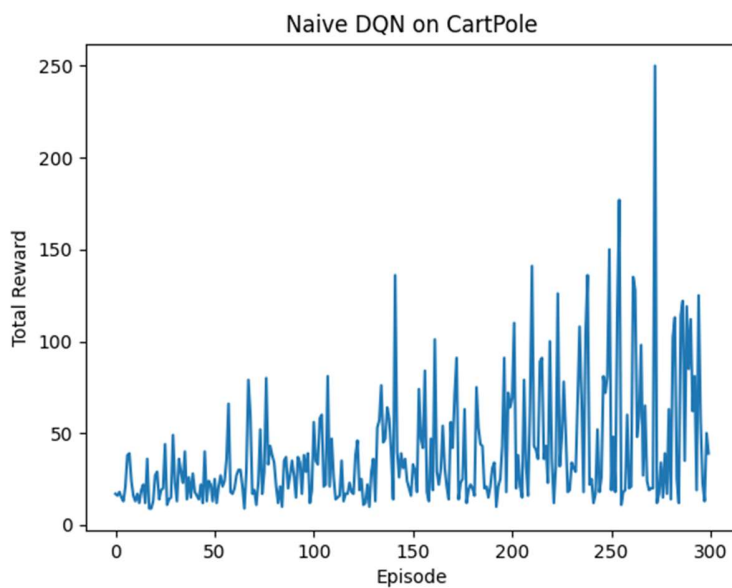
本次實驗探討三種 DQN 變體在 CartPole 環境下的訓練效果：

- Naive DQN：標準深度 Q-learning，直接以單一網路估算 Q 值。
- Double DQN：將動作選擇與價值估算分開，有效減少 Q 值過估。
- Dueling DQN：拆分出狀態價值 (V) 與優勢函數 (A)，提升策略學習效率。

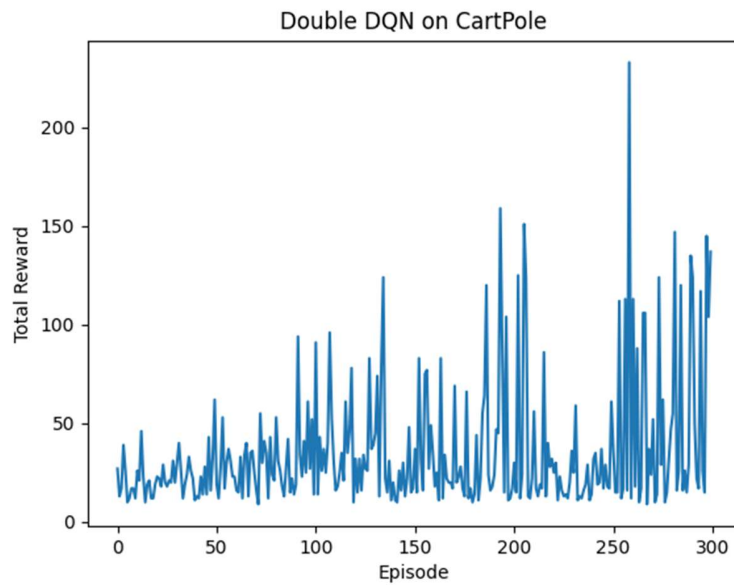
二、訓練結果圖表

下列為三種演算法在 300 集訓練中的 reward 變化趨勢：

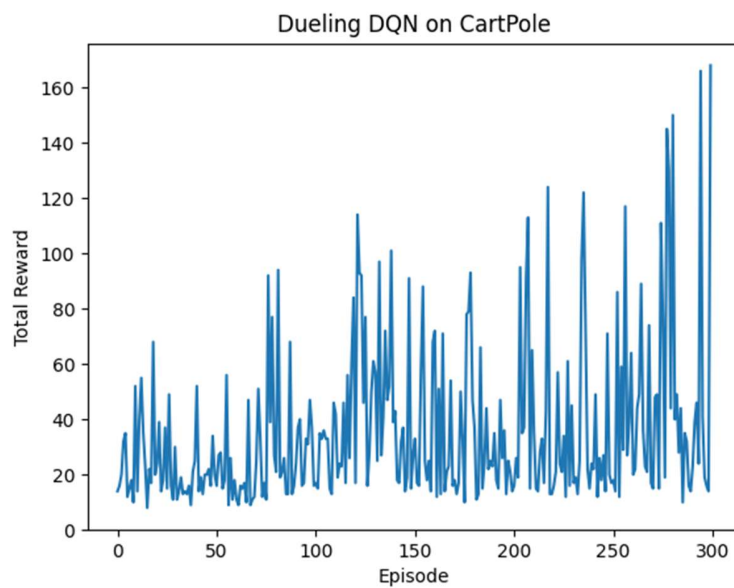
1. Naive DQN



2. Double DQN



3. Dueling DQN



三、訓練結果分析與比較

從 reward 曲線觀察：

- Naive DQN：雖有少數 episode 達到高 reward（約 250），但整體收斂不穩定。
- Double DQN：收斂稍快，約第 200 集開始較頻繁達成高 reward（約 220）。
- Dueling DQN：early stage 表現不明顯，但最末段收斂曲線明顯上升，顯示其在穩定性方面較佳。

四、結論與建議

本次比較發現，Double DQN 在 early stage 有更好的 Q 值穩定性，而 Dueling DQN 則展現了良好的後期收斂能力。

若需應用於高動作空間或複雜策略環境，建議可進一步實作 Double + Dueling DQN 結合架構。

此外，可考慮延長訓練集數與加入學習率調整、gradient clipping 等技巧以進一步穩定學習效果。