

HW4-1 理解報告：Naive DQN（靜態模式）

一、DQN 基本概念

深度 Q-learning (DQN) 是將傳統 Q-learning 中的 Q-table 替換為神經網路 (Q-network)，用來預測每個狀態對應各動作的 Q 值。透過這種方式，DQN 能夠處理連續狀態空間並學習更複雜的策略。DQN 引入了 Target Network 和 Experience Replay 兩大穩定訓練的技術，前者減少訓練目標震盪，後者打亂資料相關性以提升樣本效率。

二、Experience Replay 的功能與重要性

Experience Replay 是指將智能體與環境互動的經驗（狀態、動作、獎勵、下一狀態、終止）儲存至緩衝區中，並在訓練時隨機抽樣小批次資料進行學習。這樣做能打亂資料順序、避免過度擬合近期經驗，並提高資料利用效率，是強化學習中提升穩定性的關鍵技術。

三、實驗設計

- 使用環境：CartPole-v1
- 訓練次數：300 集 (Episode)
- 最大每集步數：500
- 初始探索率 $\epsilon = 1.0$ ，最終衰減至 0.05
- 損失函數：均方誤差 (MSE)
- Optimizer：Adam，學習率 0.001
- Network 結構：2 層全連接層，每層 128 units

四、訓練結果觀察

訓練曲線顯示在 200 集以後出現明顯提升，reward 在某些集數甚至突破 250，顯示模型已學會控制策略。但整體 reward 波動仍大，表示訓練尚未完全穩定。這可能與訓練次數不足、epsilon 衰減尚未完全、或者模型表達能力有限有關。

五、個人心得

透過實作 Naive DQN，我理解到深度強化學習中神經網路結構如何與強化學習邏輯結合，並透過訓練來逼近最佳策略。在過程中也學習到如經驗回放、target network 等關鍵穩定訓練技巧的實用性。未來可以透過延長訓練時間、使用改良演算法（如 Double DQN）來進一步提升效能。