Lab 5 結報

姓名：賴昱凱　　學號：111511141

1. 請簡述 Resource Allocation 的目的與重要性

目的：

Resource Allocation 的主要目的是在有限的系統資源下，根據不同裝置的需求與通訊條件，合理且有效地分配可用資源（如頻寬、時間、空間、傳輸功率等），以達到系統整體效能最大化、提高服務品質（Quality of Service, QoS），並確保系統公平性與效率。

重要性：

1. **應對不同需求：**

移動通訊系統中，設備之間的服務需求多樣（例如語音、影像串流、即時通訊），資源分配確保各種服務能符合其特定的延遲、頻寬或可靠性要求。

2. **適應通道狀況：**

每台裝置的通道條件（如訊號強度、干擾情形）不同，資源分配機制會根據實際通道狀況進行調整，提升傳輸效率與穩定性。

3. **最大化系統容量：**

合理分配空間、時間、頻率與功率等資源，可在有限頻譜下提升總用戶數量與傳輸速率，增加整體系統容量。

4. **節能與延長裝置壽命：**

適當的資源分配能有效降低不必要的傳輸功率消耗，有助於

裝置節能並延長電池使用時間。

5. **維持公平性：**

資源分配機制也要考慮不同用戶之間的公平性，避免資源過度集中於某些用戶而犧牲其他用戶的服務品質。

2. 請簡述 Q-Learning 的功用

Q-Learning 是一種強化學習 (Reinforcement Learning, RL) 的方法。其原理是讓一個 Agent 透過與環境互動來學習一個策略，以最大化累計的獎勵。它主要依賴一個稱為 Q-table 的表格。

1. State：

Q-Learning 將系統目前的狀況表示為一個狀態 s。在本實驗中（無線通訊系統），狀態可以是對系統總傳輸速率 (Sum Rate) 的量化表示，是一個離散變數。

$$s_t = \left\lceil \frac{R \cdot s_n}{R_{max}} \right\rceil$$

2. Action：

Agent 在觀察到當前 state 後，需要選擇一個行動 a。在本實驗中，action 是指功率分配 (power allocation) 與使用者關聯 (association assignment) 的決策。選擇 Action 的方式採用ϵ-greedy 策略，即以$\boldsymbol{\epsilon}$的機率隨機選擇行動，以$1-\boldsymbol{\epsilon}$的機率選擇 Q-table 中當前狀態下具有最大 Q 值 (期望的未來累計獎勵) 的行動。

$$a_t = \begin{cases} argmax_{a_t} Q(s_t, a_t'), if\ rand() > \epsilon \\ random\ action, otherwise \end{cases}$$

3. Reward：

在執行一個行動後，Agent 會收到來自環境的獎勵 r。在本實驗中，這個 reward 就是系統的總傳輸速率 (throughput)，這個數值越高代表越好，因此很適合作為 RL 的 reward。

4. Q-table update：

Q-Learning 的核心是更新 Q-table 中的值。當 Agent 從狀態 st 採取行動 at 後轉移到狀態 st+1 並獲得獎勵 r，它會根據 **Bellman equation** 來更新 Q-table 中 Q(st, at)的值。更新公式為

$$Q_t(s_t, a_t) \leftarrow Q_t(s_t, a_t) + \eta \left[r_t + \delta \cdot \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q_t(s_t, a_t) \right] + penalty$$

這裡的 η 是學習率 (learning rate)，決定更新數值的幅度，δ 是折扣因子 (discount factor)，決定是過去 Q 的影響大小，兩者都在 0 ~ 1 之間。這個更新過程結合了當前獲得的獎勵 r 與對未來狀態 st+1 的最大期望獎勵 max Q(st+1, at+1)。

而為了處理系統中的限制條件（功率必須小於最大功率），還定義一個**懲罰項 (penalty)** 來懲罰違反限制的行動，因次在實驗中 penalty 必須是負數，才會達到懲罰的目的。

功用：

用一句話來說，上述的步驟就是為了找到在特定狀態下應該採取哪個行動，可以最大化獎勵，而對我們來說，獎勵最大化就代表這個系統表現的最佳化。

而 Q-Learning 在無線通訊系統中的功用，本實驗應用於網路資源分配 (Network Resource Allocation)。其目的是在考慮到不同裝置的通道條件、並在多個基地台 (AP) 和多個使用者 (UE)下，利用 Q-

Learning 來選取較好的通道並分配功率進行下行資料傳輸，以達到最佳的傳輸速率 (throughput)。

3. 心得

本次實驗是最後一個 module 2 實驗，教了一個比較特別的 AI 訓練方式：增強式學習（Reinforcement learning, RL），相較於之前學習的 DNN、CNN 等 NN 模型，利用 supervised 和 Gradient Descent 的方式進行訓練不同，Reinforcement learning 可以達到不需要有 label 過的解答，模型依然可以學會我們希望他做的事情，可以利用加分、扣分的方式讓模型意識到做什麼是對的什麼是錯的，這個結果我覺的十分有趣。

這次不是使用 NN model 來進行 Reinforcement learning，讓我覺得有一點小可惜，因為我一直對如何在 NN 模型上套用 reinforcement learning 十分感興趣，但一直沒有課程有上到，幾乎都在使用 supervised learning，不過多學到一種非模型的 AI 訓練方式也是有所收穫，也期待下一個 module 可以學到更多有趣的 AI 應用。謝謝助教這五個禮拜的教學，助教人都很好而且願意與我們詳細的一對一指導，讓我收穫良多。