元智大學 工業工程與管理學系

畢業專題期末報告

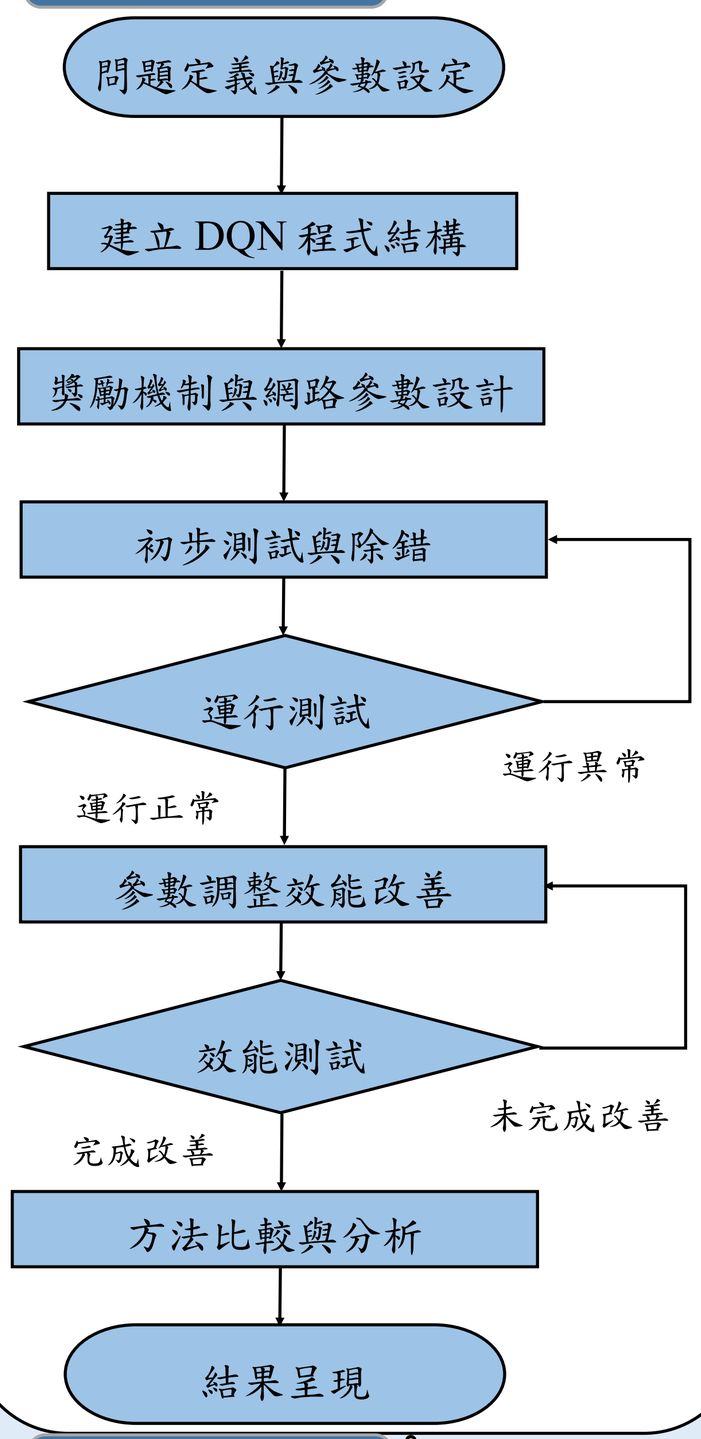
指導教授:吳政翰教授 B4_1
組長：王昊仁
組員：劉書妘、詹采瑄

專題名稱：使用 DQN 求解考量序列相依整備時間之單機排程最小化總延遲時間問題

動機與目的

因傳統、啟發式方法在大規模或動態環境下效能受限，所以本研究想利 DQN 的經驗回放及目標網路提升穩定性，並調整獎勵機制以改善排程品質。最終將其與其他方法進行比較，以驗證其在不同條件下的效能優勢。

研究流程



研究方法

卻保研究在可控環境下進行

研究假設

機台假設

1. 單一機台
2. 機台可持續操作
3. 機台一次僅可進行一個作業處理

作業假設

1. 機台作業不允許中斷
2. 作業處理時間、交期及序列相依整備時間已知且固定

環境假設

1. 靜態排程環境
2. 相關參數與變數皆為非負整數值

研究資料與參數設定

排程資料生成方式

自行生成之測試資料集
共生成 100 個作業資料

兩種獎勵機制

$$\boldsymbol{reward}(\text{最小寬放}) = \boldsymbol{min(slack)} - \boldsymbol{\beta} * \boldsymbol{s_{ij}}$$

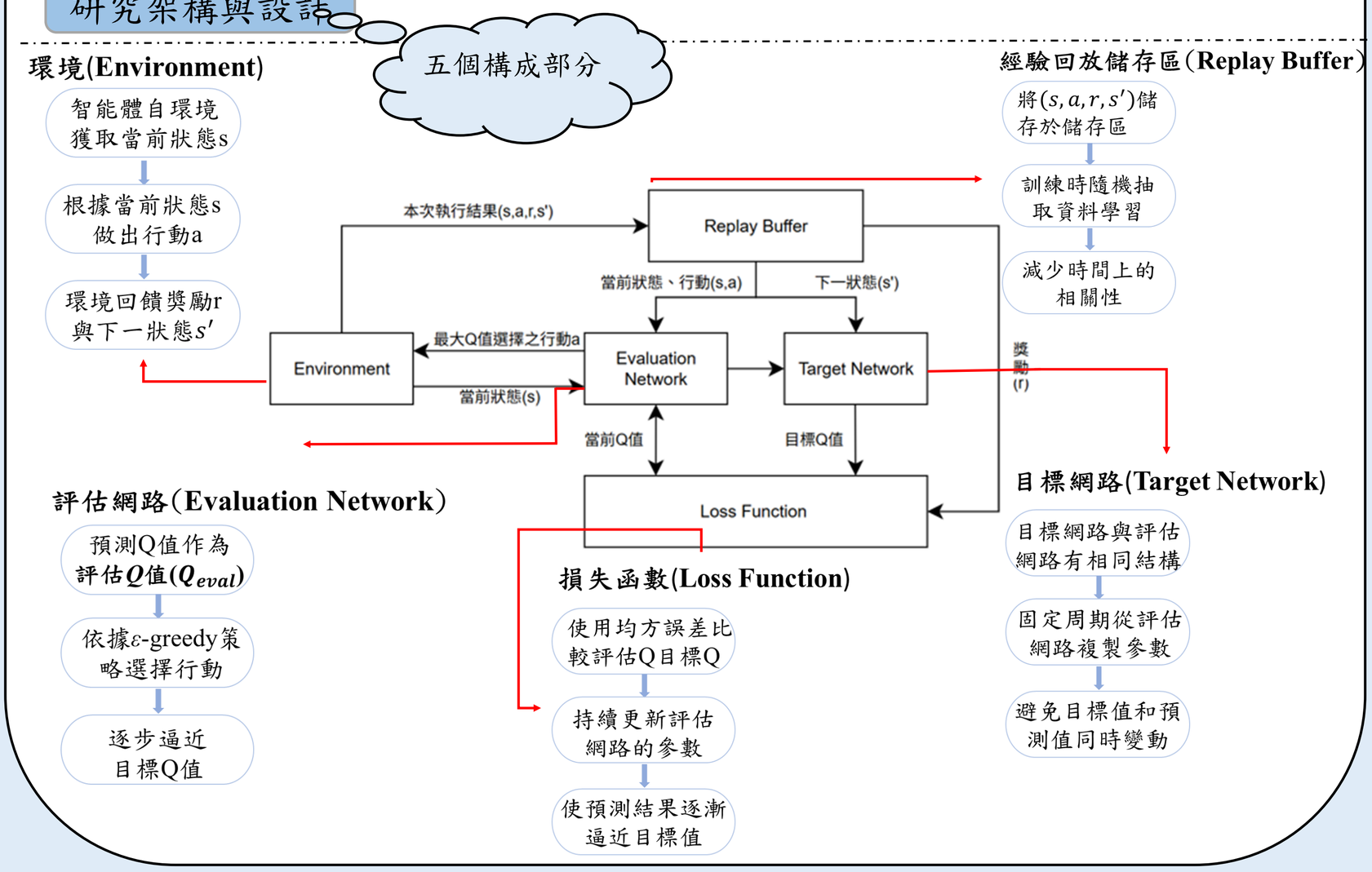
$$\boldsymbol{reward}(\text{平均寬放}) = \boldsymbol{avg(slack)} - \boldsymbol{\beta} * \boldsymbol{s_{ij}}$$

名詞解釋

avg(slack)：所有尚未排程作業各別寬放時間之平均
min(slack)：所有尚未排程作業中最小的寬放時間
β：調整參數，用以調整整備時間對獎勵的影響權重

研究架構與設計

五個構成部分



研究結果

下方 Tab.1、2 中整理了不同β值與兩種獎勵機制排程結果，以總延遲為比較指標，另外用紅色標示出各作業數中總延遲最小結果。

◆以整體的排程能力來看是調整參數β=5 的結果較好

Tab.1 最小寬放機制之排程結果

作業數	β=10	β=5	β=3	β=2
40	8186	7568	8073	7776
50	12631	12814	11750	13127
60	18765	18222	18663	18090
70	28916	26521	27531	25425
80	36101	34197	35065	33491
90	41412	41012	39771	43783
100	54986	53776	54484	54382
加總總延遲	200997	194110	195337	196074

平均寬放與最小寬放獎勵之比較：平均寬放平衡所有作業的緊迫性，避免集中單一最急作業，提升整體排程表現，明顯優於最小寬放。

◆調整參數β=3 時具有很好的結果

Tab.2 平均寬放機制之排程結果

作業數	β=10	β=5	β=3	β=2
40	7245	7268	6080	7736
50	10731	11659	11625	11615
60	17331	17215	17136	16429
70	25609	23670	24467	24255
80	32714	31622	31018	31418
90	39230	39672	39899	38986
100	53500	49830	49707	49804
加總總延遲	186360	180936	179932	180243

β值對獎勵效果之影響：適中β值(如β=3)能兼顧整備時間與整體交期，使 DQN 收斂且達到較低總延遲。

◆ DQN 在不同作業下總延遲時間都低於其他派工法

Tab.3 DQN 與派工方法之比較

作業數	β=3	EDD	LST	ATC
40	6080	8112	8171	7909
50	11625	12771	12856	12821
60	17136	18455	18633	18327
70	24467	25123	25431	25339
80	31018	32732	33169	33042
90	39899	41406	41973	41715
100	49707	50952	51696	51344
加總總延遲	179932	189551	191929	190497

結果比較：從 Tab. 3 中可以看到 DQN 能在每個決策點同時考量未排程的寬放、整備時間，在不同作業數下皆明顯低於其他派工法，隨作業數增加 DQN 在總延遲上的優勢仍持續保持。

★ 結論

本研究證實深度強化學習（DQN）能有效改善單機排程表現。透過結合作業緊迫性與序列相依整備時間（SDST）設計獎勵函數，DQN 在不同規模下皆能降低總延遲，明顯優於傳統派工法。結果表明深度強化學習方法可作為單機排程問題的可行解法，並為未來在更複雜或動態排程環境中的應用提供了參考。

DQN 表現
平均寬放獎勵能提升整體排程效率
β參數可改善收斂與延遲表現
經驗回放與固定目標網路提高學習穩定性
模型在中大型排程問題展現良好的能力