

基於多模型深度強化學習之多容器裝填系統

Jin-Ling Lin and Chih-Jung Chang

世新大學資訊管理學系

E-mail: jlilin@mail.shu.edu.tw, a110223057@gmail.com

摘要

本研究提出了一套結合注意力機制和長短期記憶的深度強化學習架構，用於解決多容器裝填問題；運用所提出學習架構訓練的模型，能夠有效擷取容器和物品裝填現況的特徵，並進行裝填動作的決策；實驗結果顯示，所提出的模型在動態裝填場景中，仍保有高度的適應性，對於不同維度的容器和物品，都能有不錯的空間利用率和決策能力，尤其在物品維度差異較小的情況下，裝填結果有相當高的空間使用率。

關鍵字：深度強化學習、長短期記憶模型、注意力機制、多容器裝填

Abstract

This study proposes a DRL framework that combines an attention mechanism and LSTM to address the multi-container packing problems. The model trained using the proposed framework can effectively capture the current packing features of containers and items, and make decisions regarding packing actions. Experimental results show that the proposed model maintains a high level of adaptability in dynamic packing scenarios, achieving good space utilization and decision-making performance across containers and items of different dimensions.

Keywords : DRL, LSTM, AM, Container Packing.

1. 前言

裝箱問題(Bin Packing Problem, BPP)為經典 NP-Hard 組合優化問題；BPP 問題相當多樣化，如單一或多容器裝填、相同或不同維度的物品或容器、即時裝填與否等等；此外尚需考慮有效堆疊的多重限制，如物品需平穩堆疊、物品不可傾斜或受重壓等等。由於問題過於複雜，除了將問題限制在

特定的前提下之外，僅能尋求近似最佳解的方式，來產生合適的物品置放方法。過往的研究多採用線性規劃、分支界限法、整數規劃、演化演算法、啟發式演算法等方法，求解最大空間使用率的有效裝填方式；然而，這些解題方法，高度依賴特定或明確的問題定義，及高效能的決策規則規劃，導致自動化系統擴充及應變的能力受到限制，無法因應實務應用中的多變需求。

深度學習能有效地進行特徵的擷取及決策規則的學習，特別是結合無需事先收集訓練資料數據的深度強化學習(DRL)，已成為解決組合優化問題的熱門研究領域；DRL 於 3D-BPP 的研究中，將問題拆解成順序、方向、和位置三個子問題，以編碼器-解碼器結合深度強化學習[2]，來降低問題的複雜度，以解決可擴展性的問題；然而這些研究多侷限於理論上或個案的探討[3]，設定一些假設的前提，如容器沒有容量的限制(長寬固定，但高度可以無限延伸)[2]、所有容器維度相同[1]，或待裝填物品維度一致等，使得研究成果與真實世界的應用，存在較大的差距。因此，本論文提出一套多模型深度強化學習方法，以克服容器裝填實務應用的靈活性。

2. 研究方法

容器中的物品裝填屬於序列式的連續決策問題，物品的裝填順序會讓容器裝填狀態產生不同的變化，換句話說，物品裝填的軌跡，在選取要裝填的物品、及決定物品的置放位置時，具有關鍵性的影響；此外，裝填過程中，系統需要掌握物品本身的特徵及容器裝填現況的特徵，這些物品狀態的特徵資訊有些用於物品裝入容器前，有些用於裝入容器後，不同的狀態僅部分物品特徵對於決策有影響；

因此，採用焦點區域概念的注意力機制，結合具記憶能力的 LSTM，在於強化學習中推導不同維度多容器裝填有效可行的決策模型，容器裝填的多模型深度強化學習架構，如圖 1。

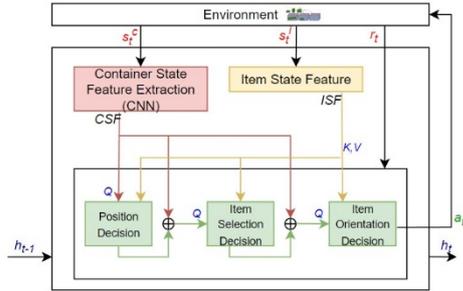


圖 1. 容器裝填多模型深度強化學習架構

動作決策則根據容器裝填狀態、及物品的特徵資訊，選取物品置放的平面位置(\$a^p\$)，要裝填的物品(\$a^i\$)，及物品擺放方向(\$a^o\$)等三種動作；獎勵函數是由物品裝填後的內部及外部閒置空間(公式(1)和(2))所組成。動作決策模型架構如圖 2 所示。

$$IFS_{C_i} = \sum_{x=0}^{CL_i} \sum_{y=0}^{CW_i} C_i(x, y) - \sum_p^{IPI_i} Volume(p) \quad (1)$$

$$EFS_{C_i} = Volume(C_i) - \sum_{x=0}^{CL_i} \sum_{y=0}^{CW_i} C_i(x, y) \quad (2)$$

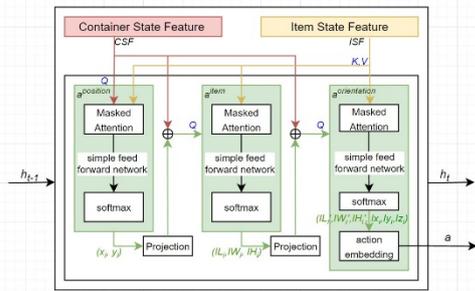


圖 2. 動作決策模型架構

3. 模擬與結果分析

圖 3 及 圖 4 為兩個容器與十個物品的裝填結果，及其相對應的空間使用率；兩個容器的內部閒置空間皆為零，表示能平穩且緊密地裝填，雖然兩個容器的外部閒置空間平均約為 50%，但從開放空間的緊密度來看，閒置空間不到 40%，且這些外部閒置空間主要集中在第二個容器；因此，更換容器大小時，只需針對第二個容器進行調整。系統在單一容器、三個容器中，進行 5-12 個不同維度物品的裝填，結果顯示，當物品維度差異小時，從開放

空間的緊密度來看，外部閒置空間都能保持在 20% 以下，而內部的閒置空間則僅在物品維度存在較大差異時才會出現。

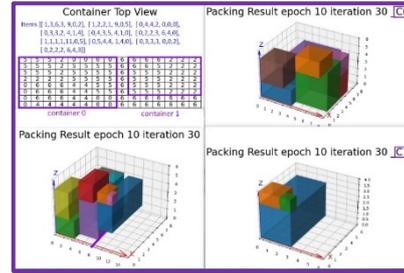


圖 3. 動作決策模型架構

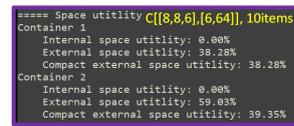


圖 4. 動作決策模型架構

4. 結論

本研究提出了一種多模型深度強化學習的架構，用於解決多容器裝填的問題，並透過模擬實驗驗證所提出方法的有效性及適應性；實驗結果證明，所提出方法不僅能夠有效提升容器的空間利用率，且能保持在處理不同維度和形狀物品時的靈活性與穩定性，為多容器裝填問題提供了一個高效率的解決方法。

5. 參考文獻

- [1] Hang Zhao, H., She, Q., Zhu, C., Yang, Y., & Xu, K. (2022). "Online 3D Bin Packing with Constrained Deep Reinforcement Learning," Machine Learning, Computer Science, Cornell University.
- [2] Jiang, Y., Cao, Z., & Zhang, J. (2023). "Learning to Solve 3-D Bin Packing Problem via Deep Reinforcement and Constraint Program," *IEEE Transactions on Cybernetics*, 53(5), 2864.
- [3] Zhao, H., Zhu, C., Xu, X., Huang, H., & Xu, K. (2022). "Learning practically feasible policies for online 3D bin packing," *Science China Information Sciences*, 65(112105).

6. 誌謝

本研究承蒙行政院國家科學及技術委員會，專題研究計畫 NSTC 112-2221-E-128-002-，NSTC 113-2221-E-128-001- 經費補助，特此申謝。