

國立臺灣師範大學科技與工程學院電機工程學系

碩士論文

Department of Electrical Engineering
College of Technology and Engineering
National Taiwan Normal University
Master's Thesis

整合自然語言處理與強化式學習之協作機器人開發
Development of Collaborative Robots by Integrating Natural
Language Processing and Reinforcement Learning

駱忠湧

Chung-Yung Lo

指導教授：許陳鑑 博士

Advisor：Chen-Chien Hsu, Ph.D.

中華民國 113 年 7 月

July 2024

誌 謝

首先，我要衷心感謝我的指導老師許陳鑑教授，感謝您的耐心指導和無私分享的知識。您的支持和鼓勵讓我在研究中取得了更深層次的理解，並且成為了我學術生涯中的重要導師。我也要感謝我的實驗室夥伴們，在研究過程中，您們的合作和討論為我提供了寶貴的意見和啟發，讓我能夠不斷提升自己的研究能力。此外，我要感謝我的家人和朋友。在整個研究過程中，你們給予了我無條件的支持和理解，讓我能夠全身心地投入到我的研究中。

駱忠湧 謹誌

中華民國一百一十三年七月二十六日

整合自然語言處理與強化式學習之協作機器人開發

學生：駱忠湧

指導教授：許陳鑑博士

國立臺灣師範大學電機工程學系碩士班

摘 要

近年來，隨著人工智慧領域不斷進步，人們對於機器人能夠以自然方式理解和回應人類指令的需求，以及在不斷變化的環境中自主學習完成任務的需求日益增加。本研究旨在通過整合自然語言處理(NLP)和強化式學習(RL)的技術，實現人類與協作機器人(Cobots)之間更直觀的協作，使它們能夠處理複雜任務。首先，我們使用語音轉文字技術將人類的語音指令轉換為機器可理解的文本數據。隨後，利用 CKIP Transformer 進行詞性分析和任務提取，旨在識別和提取人類指令中包含的特定任務內容，並且使用 Nvidia Isaac Gym 作為強化式學習平台，根據 NLP 的分析，訓練 Cobots 在模擬環境中自主執行相應的動作。通過 NLP 和 RL 技術的結合，我們得以實現更智能、更適應性的 Cobots，不僅僅是被動執行特定指令，而是能夠理解人類意圖，並根據上下文靈活適應和反應。我們相信，具備這種主動學習和改進能力的 Cobots 將使其與人類更有效地協作，能夠應對各種任務和情況，無論是在家庭、工業還是服務相關的應用領域，這些智能協作機器人的實現有望提高生產力，改善生活質量，並促進機器人技術的廣泛應用和發展。

關鍵字：深度強化式學習、自然語言處理、協作機器人

Development of Collaborative Robots by Integrating Natural Language Processing and Reinforcement Learning

Student : Chung-Yung Lo

Advisors : Dr. Chen-Chien Hsu

Department of Electrical Engineering
National Taiwan Normal University

ABSTRACT

With the continuous advancement in the field of artificial intelligence, there is a growing need for robots to be able to understand and respond to human commands in a natural way, and to autonomously learn to complete a task in ever-changing environments. This study aims to achieve more intuitive collaboration between humans and collaborative robots (Cobots), allowing them to handle complex tasks by integrating natural language processing (NLP) and reinforcement learning (RL) techniques. Firstly, speech-to-text technology is employed to convert voice commands from a human into machine-understandable textual data. Subsequently, CKIP Transformers are utilized for part-of-speech analysis and task extraction, aiming to identify and extract specific task content contained within the human commands. Furthermore, Nvidia Isaac Gym serves as the reinforcement learning platform for training the Cobots in simulated environments to autonomously execute corresponding actions based on the NLP analysis. Through this combination of NLP and RL techniques, we aim to realize smarter and more adaptable Cobots that not only passively execute specific commands, but also understand human intentions, adapting and reacting flexibly to the context. We believe that such proactive learning and improvement capabilities in the Cobots will enable more effective collaboration with humans and enable them to address various tasks and situations, whether in the realms of household, industrial, or service-related applications.

The realization of these intelligent collaborative robots is expected to enhance productivity, improve quality of life, and promote the widespread applications and development of robotic technology.

Keywords: Deep Reinforcement Learning, Natural Language Processing, Collaborative Robots

目 錄

誌 謝	I
摘 要	II
ABSTRACT	III
目 錄	V
表 目 錄	VII
圖 目 錄	VIII
第一章 緒論	1
1.1 研究背景與動機	1
1.2 挑戰與貢獻	1
1.3 論文架構	2
第二章 文獻探討	4
2.1 自然語言處理	4
2.2 深度強化式學習	5
2.3 近端策略優化	8
2.4 自然語言處理與協作機器人應用	11
第三章 研究方法	13
3.1 自然語言處理模組	13
3.2 深度強化式學習模組	17
3.3 整合自然語言處理與強化式學習之協作機器人系統架構	22
第四章 實驗場景與結果	25
4.1 硬體設備與軟體	25

4.2	模擬環境設置.....	28
4.3	模擬環境訓練結果.....	30
第五章	結論與未來展望	35
5.1	結論.....	35
5.2	未來展望.....	35
參 考 文 獻	37
自 傳	40

表 目 錄

表 1、獎勵尺度對應表	22
表 2、伺服器規格表	25
表 3、TM5M-700 規格表[29].....	26
表 4、FRANKA HAND 規格表[30]	27
表 5、神經網路參數	30

圖 目 錄

圖 1、強化式學習架構圖[20]	7
圖 2、剪切目標函數示意圖[25]	10
圖 3、近端策略優化虛擬碼[25]	10
圖 4、自然語言處理之任務提取流程圖	14
圖 5、任務提取虛擬碼	16
圖 6、協作機器人系統架構圖	24
圖 7、自然語言處理介面	28
圖 8、NVIDIA ISAAC GYM 訓練環境	29
圖 9、強化式學習訓練結果	31
圖 10、夾爪靠近物件	32
圖 11、機械手臂夾起物件	32
圖 12、物件靠近目標位置	33
圖 13、機械手臂翻轉物件	33
圖 14、物件到達目標位置及姿態(完成任務)	34

第一章 緒論

1.1 研究背景與動機

隨著人工智慧和機器人技術的快速發展，人們對於協作機器人的需求正不斷提升。協作機器人在各種應用場景中，如製造業、服務業及醫療保健領域，展現了巨大的潛力。然而，現有的大多數協作機器人僅專注於執行數個預先設計的固定任務，缺乏與人類自然互動的能力，這使得它們在與人類共同完成任務時，無法根據實際情況進行即時調整與應變。

為了解決上述問題，本研究旨在透過整合自然語言處理和強化式學習技術，來提升協作機器人的互動能力和任務執行範圍。自然語言處理技術使計算機能夠理解和解析人類語言，這不僅使人機互動變得更加簡單直觀，也大幅提升了協作機器人的操作靈活性。另一方面，強化式學習則通過模擬不同的環境和情境，訓練協作機器人學習並適應多變的任務需求，從而能夠在更多元且複雜的情境中執行任務。

基於以上敘述，本研究希望藉由自然語言處理和強化式學習技術的結合，不僅改善人類與機器人的互動體驗，還能使協作機器人具備更高的智能和適應性，以應對未來更多樣化的應用需求和挑戰。這將不僅提升協作機器人的應用價值，也將為人機協作開創新的可能性。

1.2 挑戰與貢獻

傳統機械手臂的控制通常依賴於預先設計的運動軌跡和控制算法，這些控制方法在執行固定任務時表現出色。然而，當面臨動態變化的環境和需要靈活應對的任務時，這些方法的局限性便顯現出來。傳統機械手臂缺乏自主學習和即時調整能力，因此無法滿足現代智能製造和服務業中對高靈活性和適應性的需求。

為了解決這些問題，本研究結合了自然語言處理和強化式學習技術。自然語言處理技術使機械手臂能夠理解並回應人類的語言指令，使人機互動變得更加直觀和高效。這種交互方式不僅簡化了操作過程，也減少了對專業技術知識的依賴，使得非專業人員也能輕鬆操作機械手臂。

另一方面，強化式學習技術允許機械手臂在不斷變化的環境中學習並適應新的任務需求。通過設計合理的獎勵機制，機械手臂可以在訓練過程中逐步優化其行為策略，從而在實際應用中表現出更高的靈活性和適應性。

本研究的創新點在於成功結合自然語言處理和強化式學習技術，開發出一套具有高度互動性和適應性的協作機器人系統。這一系統不僅突破了傳統機械手臂在固定任務和預設軌跡上的局限性。通過這種技術融合，協作機器人能夠在動態環境中靈活應對多樣化的任務需求，提升了其應用範圍和實用價值。

1.3 論文架構

本論文的研究架構分為五個章節，各章節的內容概要說明如下：

第一章 緒論

說明研究的背景與動機，指出現有技術的局限性和挑戰，並闡述本研究的創新點和貢獻。最後，介紹論文的整體架構，讓讀者了解各章節的安排。

第二章 文獻探討

針對自然語言處理、深度強化式學習及近端策略優化演算法的相關文獻進行回顧。分析這些技術在協作機器人中的應用現狀，並指出現有研究的不足之處，為本研究提供理論基礎。

第三章 研究方法與系統設計

詳細介紹本研究所設計的整合系統，包括自然語言處理和強化式學習的技術框架。說明所使用的設備及軟體，並描述系統的實現步驟和技術細節。

第四章 實驗場景與結果

描述實驗的設計和場景設定，展示實驗結果並進行分析。通過多個實驗案例，評估所提出系統的性能和有效性，並與現有技術進行比較，討論實驗發現。

第五章 結論與未來展望

總結整理以及未來可能的應用發展。

第二章 文獻探討

2.1 自然語言處理

自然語言處理(Natural Language Processing, NLP)是人工智慧和語言學領域的一個分支，旨在使計算機能夠理解、處理和生成自然語言。這一領域涉及多個方面，包括語言的認知、理解和生成。認知和理解使得計算機能夠理解輸入語言的含義、上下文和關係，而生成則使計算機能夠將結構化數據轉化為自然語言的形式。NLP的目標不僅僅是創建能夠處理語言的系統，還包括建立模型來表示語言的能力和特徵，以及開發應用這些模型的各種實用系統。這包括許多應用領域，如機器翻譯、文本分類、情感分析、自動摘要、問答系統等。

早期的自然語言處理方法主要基於統計和機器學習技術，常見的方法包括：

- 隱馬爾可夫模型(Hidden Markov Models, HMMs)[1]：用於序列標注任務，例如詞性標註和命名實體識別。
- 條件隨機場(Conditional Random Fields, CRFs)[2]：用於序列預測任務，在處理上下文依賴性方面比HMM[1]更為有效。
- 支持向量機(Support Vector Machines, SVMs)[3]：用於文本分類和情感分析等任務。

深度學習方法在自然語言處理中的應用大大提升了各項任務的性能，以下是一些重要的深度學習模型和技術：

- 詞嵌入(Word Embeddings)[4]：如 Word2Vec[5]、GloVe[6]和FastText[7]等模型通過將詞彙表示為連續向量來捕捉詞與詞之間的語義關係。
- 卷積神經網路(Convolutional Neural Networks, CNNs)[8]：主要用於文本分類和句子建模，能夠自動學習文本中的關鍵特徵。
- 循環神經網路(Recurrent Neural Networks, RNNs)[9]：如 LSTM[10]和GRU[11]，能夠有效處理序列數據，廣泛應用於機器翻譯和語音識別。

- 注意力機制(Attention Mechanisms)[12]：提高了模型在長距離依賴和句子對齊方面的性能，是Transformer模型[13]的基礎。

Transformer模型[13]由Vaswani等人於2017年提出，徹底改變了自然語言處理的發展。該模型基於自注意力機制[12]，具有高度並行性和更強的語境理解能力。其衍生模型如BERT(Bidirectional Encoder Representations from Transformers)[14]、GPT(Generative Pre-trained Transformer)[15] 和 T5(Text-To-Text Transfer Transformer)[16]在多項自然語言處理基準測試中達到了最先進的性能。

- BERT[14]：通過雙向預訓練技術，能夠更好地理解上下文語義。
- GPT[15]：主要用於生成任務，生成高質量的自然語言文本。
- T5[16]：將所有自然語言處理任務統一建模為文本到文本的轉換，具有高度靈活性和強大的性能。

隨著自然語言處理技術不斷進步，以下是一些發展趨勢：

- 大規模語言模型(LLMs)的進化：如GPT-4[17]和Llama2[18]，這些模型能夠處理更複雜的問題並在多領域應用中表現出色。
- 組合AI(Combinational AI)：技術如LangChain[19]，允許將多個LLM結合使用，以解決更具挑戰性的問題，如客戶行為分析和市場趨勢預測。
- 強化式學習在自然語言處理中的應用：強化式學習技術正在用於改進自然語言處理模型的性能，特別是在連續改進和自適應方面。
- 多語言處理能力的提升：利用大型多語言訓練數據集，使自然語言處理模型能夠更有效地處理多語言文本。

2.2 深度強化式學習

強化式學習是一種機器學習的方法，它是通過觀察環境、從環境中獲取反饋、學習如何選擇動作來使獲得的獎勵最大化的過程。強化式學習的目標是找到一個

策略，使得在特定的環境下，代理能夠獲得最大的累積獎勵。

強化式學習主要以五個部分組成，分別是代理(Agent)、環境(Environment)、狀態(State)、行動(Action)、獎勵(Reward)，這些部分共同作用，形成一個閉環的學習和決策過程，如圖1所示。

- 代理(Agent)是在環境中學習並採取行動的實體，它的目標是最大化累積獎勵。代理根據當前狀態選擇行動，並根據獲得的獎勵更新策略。
- 環境(Environment)是代理所互動的外部系統，它接收代理的行動並返回新的狀態和即時獎勵。環境可以是物理世界，也可以是模擬的數字環境。
- 狀態(State)表示環境在某一時刻的情況或配置，代理根據這些狀態信息來做出決策。狀態可以包括多種感知數據，例如位置、速度、圖像等。
- 行動(Action)是代理在每個狀態下所選擇的操作。行動空間可以是離散的（例如，向左或向右移動）或連續的（例如，調整控制變量的值）。
- 獎勵(Reward)是環境在代理採取某個行動後給予的即時反饋，表示行動的好壞。獎勵信號用於指導代理學習，目標是通過學習策略最大化累積獎勵。

強化式學習主要研究如何讓代理在特定環境中通過試錯(trial and error)學習一個策略(Policy)，以最大化累積獎勵。代理通過探索(Exploration)和利用(Exploitation)平衡試驗新行動和使用已知最佳行動來收集環境訊息。強化式學習問題通常建模為馬可夫決策過程(MDP)，包含狀態、行動、轉移機率(Transition Probability)和獎勵。

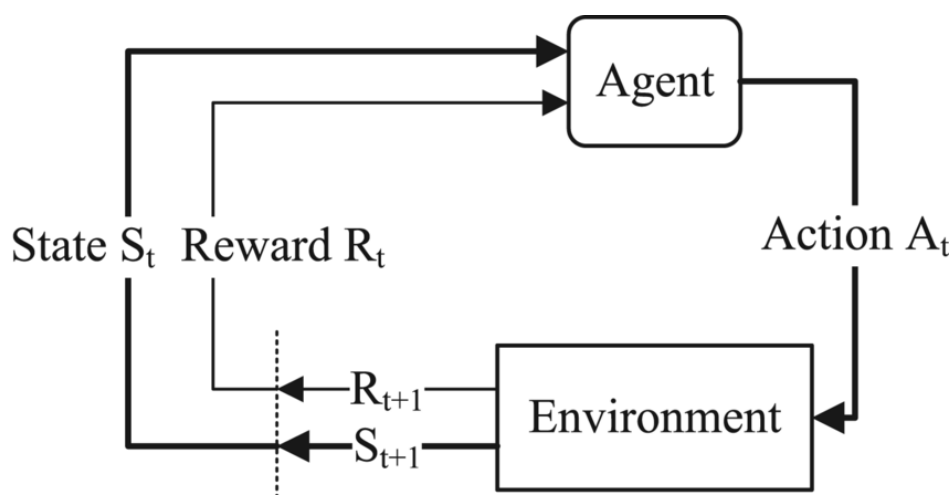


圖 1、強化式學習架構圖[20]

深度強化式學習(Deep Reinforcement Learning, DRL)[21]和傳統強化式學習在目標和方法上有許多相似之處，但在技術實現和應用範圍上存在顯著差異。傳統強化式學習通過代理與環境互動並獲取反饋（獎勵），使代理學習如何在不同狀態下選擇最佳行動以最大化累積獎勵。它依賴於價值函數和策略函數來評估行動的價值，常用的方法包括值迭代、策略迭代、Q-Learning和策略梯度方法。

深度強化式學習結合了深度學習和強化式學習技術，通過使用深度神經網路來處理高維和複雜的狀態空間，使代理能夠在更複雜的環境中進行學習和決策。深度強化式學習使用深度神經網路來近似價值函數、策略函數或Q函數，從而自動從數據中提取特徵，減少了人工設計特徵的需求，提高了模型的泛化能力。深度強化式學習常用的方法包括深度Q網路(Deep Q-Network, DQN)、策略梯度方法以及混合方法。

深度強化式學習的應用範圍廣泛，在遊戲AI、機器人控制和自動駕駛等領域取得了顯著的進展。例如，深度強化式學習被用於開發AlphaGo和Dota 2 AI，在複雜策略遊戲中取得突破性進展；在機器人操作和運動規劃中，深度強化式學習能夠處理高維感知輸入並學習複雜的操作策略；在自動駕駛中，深度強化式學習幫助自動駕駛車輛在複雜環境中做出實時決策。

深度強化式學習可以根據以下幾種方式進行分類：

- 根據策略更新方式，可以分為值函數方法(Value-based Methods)、策略梯度方法(Policy Gradient Methods)和混合方法(Actor-Critic Methods)。
- 根據動作空間，可以分為離散動作空間(Discrete Action Space)和連續動作空間(Continuous Action Space)。
- 根據策略是否確定，可以分為確定性策略(Deterministic Policy)和隨機性策略(Stochastic Policy)。
- 根據學習是否在線進行，可以分為在線策略(On-Policy)和離線策略(Off-Policy)。
- 根據環境模型的依賴性，可以分為無模型方法(Model-free)和基於模型的方法(Model-based)。

深度強化式學習有許多演算法，這些演算法可以根據動作空間分為兩類：離散動作空間和連續動作空間。在離散動作空間中，典型的演算法包括深度Q網路，適用於每個行動都是有限選項的情況。在連續動作空間中，行動可以是無限多種不同選項，典型的演算法包括深度確定性策略梯度(Deep Deterministic Policy Gradient, DDPG)[22]、漸進熵策略梯度(Soft Actor-Critic, SAC)[23]和近端策略優化(Proximal Policy Optimization, PPO)[24]。這些演算法適用於機械臂的精確控制和複雜的機器人控制任務。本論文中的機械手臂的動作屬於連續動作，因此選擇使用近端策略優化來訓練機械手臂的夾取策略。

2.3 近端策略優化

近端策略優化是由OpenAI於2017年提出的一種先進強化式學習演算法。近端策略優化將策略梯度方法與一種新型的目標函數結合，通過限制策略更新的幅度，提高了算法的穩定性和效率，使其在多種強化式學習任務中表現出色。

近端策略優化具有多項優點，包括高穩定性、實現簡單、效率高、泛化能力

強、樣本利用率高、靈活性高、策略穩定性好和無需信賴域計算。這些優點源於近端策略優化使用剪切目標函數或KL散度懲罰來限制策略更新幅度，防止過度更新，提高訓練穩定性；簡化了策略優化過程，降低了實現和計算的複雜度；能高效利用樣本數據，快速收斂到優良策略；具備強大的泛化能力和適應性，在多種應用場景中均表現出色；並且靈活調整超參數以適應不同環境需求，避免了信賴域計算的複雜性，使其成為現代強化式學習中的流行算法。

近端策略優化的核心創新之一是其剪切目標函數，這個目標函數限制了策略更新的幅度，以防止策略變化過大，從而提高訓練的穩定性。剪切目標函數的設計如下：

$$L^{\text{CLIP}}(\theta) = E_t[\min(r_t(\theta)\widehat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\widehat{A}_t)] \quad (2-1)$$

其中， $r_t(\theta)$ 是新舊策略的比率，計算方式為：

$$r_t(\theta) = \frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)} \quad (2-2)$$

$\pi_{\theta}(a_t|s_t)$ 和 $\pi_{\theta_{\text{old}}}(a_t|s_t)$ 分別是當前策略和舊策略在狀態 s_t 下選擇行動 a_t 的概率。 \widehat{A}_t 是優勢估計，用於衡量行動 a_t 在狀態 s_t 下相對於平均水平的好壞。 ϵ 是剪切超參數，通常設置為0.1或0.2，用於控制策略更新的幅度。這種目標函數設計通過限制策略更新的範圍，防止策略變化過大，提高了訓練的穩定性和效率。圖2為剪切目標函數示意圖，當 $r_t(\theta)$ 小於 $1 - \epsilon$ 時輸出 $1 - \epsilon$ ，當 $r_t(\theta)$ 大於 $1 + \epsilon$ 時輸出 $1 + \epsilon$ 。

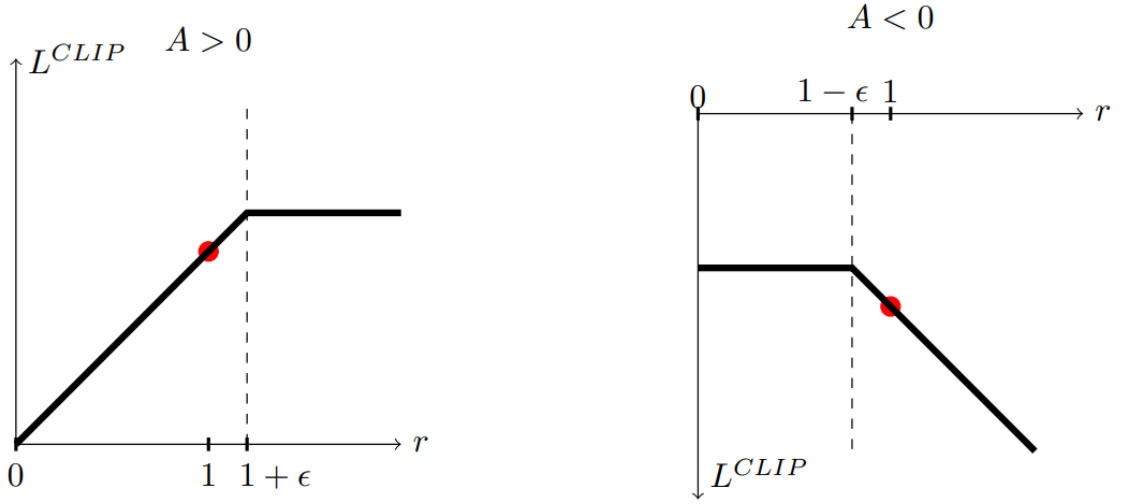


圖2、剪切目標函數示意圖[25]

圖3為近端策略優化的虛擬碼，近端策略優化的流程包括初始化策略和價值函數參數，使用當前策略在環境中收集軌跡數據，計算回報和優勢估計，通過最大化剪切目標函數更新策略，最小化均方誤差損失更新價值函數，並重複這些步驟，直到達到預定的訓練次數或性能標準。這個流程使得近端策略最佳化能夠穩定地更新策略，逐步提高在各種強化式學習任務中的性能。

Algorithm 1 PPO-Clip

- 1: Input: initial policy parameters θ_0 , initial value function parameters ϕ_0
- 2: **for** $k = 0, 1, 2, \dots$ **do**
- 3: Collect set of trajectories $\mathcal{D}_k = \{\tau_i\}$ by running policy $\pi_k = \pi(\theta_k)$ in the environment.
- 4: Compute rewards-to-go \hat{R}_t .
- 5: Compute advantage estimates, \hat{A}_t (using any method of advantage estimation) based on the current value function V_{ϕ_k} .
- 6: Update the policy by maximizing the PPO-Clip objective:

$$\theta_{k+1} = \arg \max_{\theta} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^T \min \left(\frac{\pi_{\theta}(a_t|s_t)}{\pi_{\theta_k}(a_t|s_t)} A^{\pi_{\theta_k}}(s_t, a_t), g(\epsilon, A^{\pi_{\theta_k}}(s_t, a_t)) \right),$$

typically via stochastic gradient ascent with Adam.

- 7: Fit value function by regression on mean-squared error:

$$\phi_{k+1} = \arg \min_{\phi} \frac{1}{|\mathcal{D}_k|T} \sum_{\tau \in \mathcal{D}_k} \sum_{t=0}^T \left(V_{\phi}(s_t) - \hat{R}_t \right)^2,$$

typically via some gradient descent algorithm.

- 8: **end for**
-

圖 3、近端策略優化虛擬碼[25]

2.4 自然語言處理與協作機器人應用

自然語言處理技術的發展，使機器能夠理解和生成自然語言，大大提升了人機交互的自然性和靈活性。同時，協作機器人因其能在無保護柵欄的情況下與人類安全協作，逐漸成為工業和服務業中的重要工具。這些機器人能夠在提升生產效率的同時減少人員的工作負擔，為各種應用場景提供了新的解決方案。

在文獻[26]中，作者介紹了一種在共享工作空間中通過命令協調人機協作的方法。該系統結合了自然語言處理和自動語音識別技術，通過知識庫進行實時推理和動作驗證，使機器人能夠根據操作人員的圖形界面或語音指令來完成任務。結果表明，這種方法在工業應用場景中，如裝配、配套和交接任務中，能夠有效地協調人機合作，提升工作效率。然而，該研究也指出了一些現有技術的缺點和限制。首先，語音識別的準確性在噪音環境下會受到影響，導致機器人無法準確理解指令。其次，語義理解的複雜性仍然是一大挑戰，自然語言的多義性和上下文依賴性使得機器人很難準確解釋指令的意圖。在控制方面，該系統依賴於預定義的動作和編程來實現機器人行為，這使得系統的靈活性受到限制。尤其是當面對動態和非結構化環境時，傳統控制方法難以應對突發情況和快速變化的需求。

在文獻[27]中，作者提出了一種模組化的移動協作機器人系統，該系統包括語音理解、物體定位和多相機定位等三個主要模組。語音理解模組使用自然語言處理技術來解析人類的語音指令，建立動作基礎以描述人類指令。物體定位模組結合了YOLOv4和點雲技術，用於精確定位物體在三維空間中的位置。多相機定位系統則利用ArUco標記和多台相機，實現低成本且高效的移動機器人定位，特別適用於小型工作區域的應用。該研究也在實驗中展示了機器人在組裝木製椅子過程中的表現，顯示出系統在處理未曾見過的場景中能有效協作並執行任務。系統的模組化設計使其易於維護和升級，未來可以進一步改進，以應對不同的工業和服務應用場景。最後作者也提到雖然系統可以通過人工命令準確地執行任務，

但系統仍然存在一些局限性，第一個是如果有任何新物件，需要重新訓練學習模型，第二個是放置物體的位置和移動機器人的導航路徑必須在配備的相機的視野範圍內。

第三章 研究方法

透過將自然語言處理(Natural Language Processing, NLP)與強化式學習(Reinforcement Learning, RL)這兩項先進技術進行整合，本研究利用自然語言處理技術將語音指令轉換成指定的任務格式，再將這些指定的任務格式送入經強化式學習訓練的模型中。這樣的設計實現了電腦能夠更準確地理解並解釋人類所說的話語，進而訓練協作機械手臂完成特定且複雜的任務。

這種整合不僅使得機械手臂能夠在多種情境下與人類無縫協作，共同完成各種任務，同時也顯著提高了整個操作過程中的效率和精度。這樣的進步在多種應用領域中都帶來了更高的生產力和效益，使得技術的應用範圍更加廣泛，從而推動了相關行業的發展和創新。

3.1 自然語言處理模組

中央研究院所開發的CKIP Transformers語言模型[28]，其針對繁體中文訓練了一系列的Transformer[13]模型。除了訓練語言模型之外，亦於各個自然語言任務上訓練了對應的模型（包含斷詞、詞性標記、實體辨識）。

本研究使用CKIP Transformers語言模型[28]將下達任務的命令文本進行斷詞及詞性標記，接著經過本研究所設計的任務提取模組，將任務整理成三個部分，分別是互動的物件、機械手臂的動作以及物件擺放的目的地，最終結果即是訓練機械手臂要完成的任務內容。

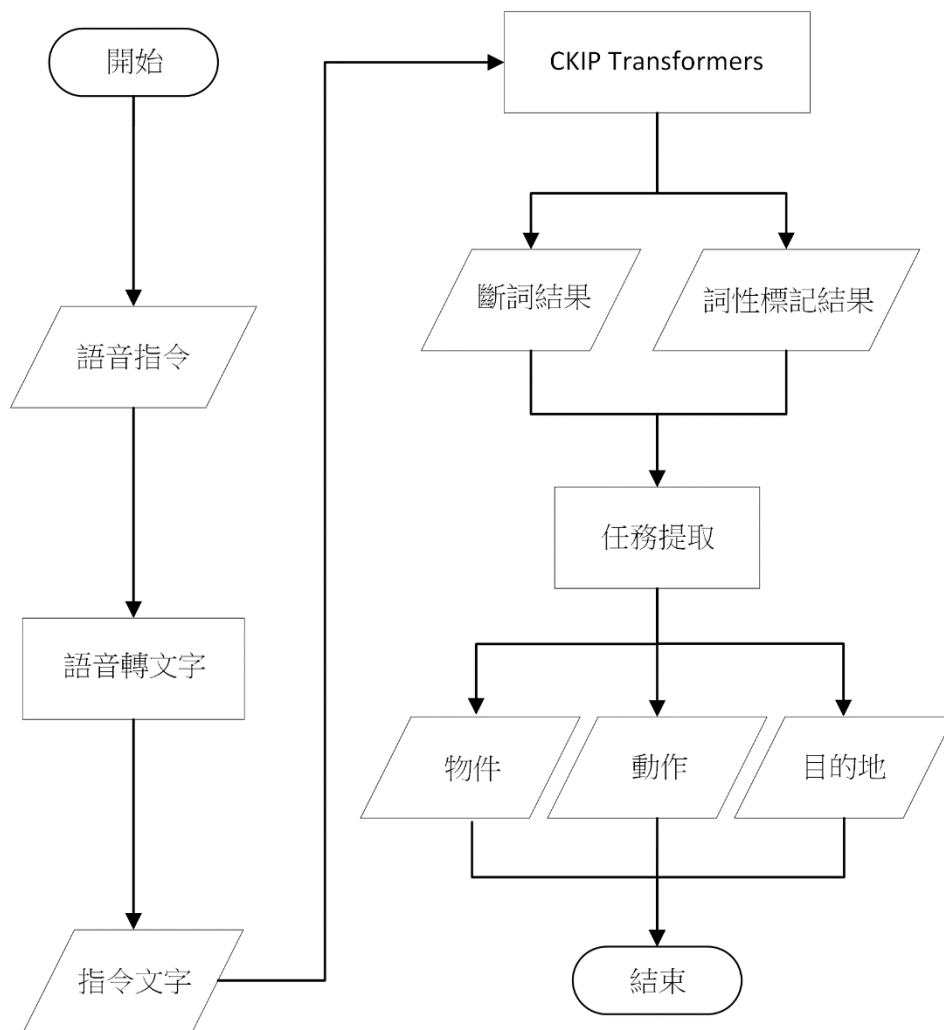


圖 4、自然語言處理之任務提取流程圖

本研究的任務提取流程如圖4所示，首先將使用者說的話透過語音轉文字服務轉成文字後，經過CKIP Transformers語言模型[28]的處理後分別得到斷詞以及詞性標記的結果，最後再經過任務提取模組的處理後，如圖5所示，便可得到最終任務的內容。詳細步驟如下：

步驟一：語音轉文字

- 接收語音指令內容：使用SpeechRecognition這個Python套件，並通過Google Speech Recognition服務處理後，將語音轉成文字輸出。

步驟二：語言模型處理

- 文字處理：將轉成的文字作為輸入送進CKIP Transformers語言模型[28]，

分別進行斷詞和詞性標記的任務處理。再將處理後的結果作為輸入送進任務提取模組。

- 斷詞與詞性標記處理：
 - 透過斷詞快速切割出每個詞。
 - 詞性標記能夠標記每個詞的詞性。

步驟三： 任務提取

- 將斷詞語詞性標記的結果輸入到任務提取模組內，流程如圖5所示
- 任務提取模組
 - 文法規則判斷：利用文法規則判斷出物件、動作以及目的地分別代表的詞。
 - 錯誤處理：在處理過程中，可能會遇到多個名詞而導致提取出的任務有錯誤的問題。針對這些情況，必須進行額外處理，通過找出特定句子組成的規則，成功解決大部分類似問題。

● 範例：

指令：把杯子放到桌子上

↓

斷詞結果：把 杯子 放到 桌子 上

詞性標記結果：P Na VC Na Ncd

↓

送進任務提取模組

↓

最終結果：

動作：夾取，物件：杯子，目的地：桌子

Algorithm 2 Extract Mission Process

```
1: Input: sentence_ws, sentence_pos
2: Initialize res with keys 'VC', 'Na', 'Ncd' and empty lists as values
3: for each  $i$ , ( $word\_ws, word\_pos$ ) in enumerate(zip(sentence_ws,
sentence_pos)) do
4:   if ( $word\_pos \in ['Na', 'Ncd']$ ) or ( $word\_pos.startswith('V')$  and
( $word\_ws$  not in VC_EXPECT)) then
5:     if not  $word\_pos.startswith('V')$  then
6:       if  $word\_pos == 'Ncd'$  and  $sentence\_pos[i - 1] == 'Na'$  then
7:         Append  $sentence\_ws[i - 1]$  to res[word_pos]
8:       else
9:         Append  $word\_ws$  to res[word_pos]
10:      end if
11:    else
12:      if  $i == \text{len}(sentence\_ws) - 1$  then
13:        Append  $word\_ws[0]$  to res['VC']
14:      else
15:        if  $sentence\_ws[i + 1] == 'after'$  then
16:          Append  $word\_ws$  to res['VC']
17:        else
18:          Append  $word\_ws[0]$  to res['VC']
19:        end if
20:      end if
21:    end if
22:  end if
23: end for
24: if  $\text{len}(\text{res}['Ncd']) == 0$  then
25:   Append 'hand' to res['Ncd']
26: end if
27: Rename the keys in res:
28:  $\text{res}['action'] \leftarrow \text{res.pop('VC')}$ 
29:  $\text{res}['object'] \leftarrow \text{res.pop('Na')}$ 
30:  $\text{res}['destination'] \leftarrow \text{res.pop('Ncd')}$ 
```

圖5、任務提取虛擬碼

3.2 深度強化式學習模組

本研究使用Nvidia所開發的深度強化式學習訓練平台Isaac Gym進行深度強化式學習機械手臂控制訓練，訓練時模型的輸入為3.1章節所提到的任務內容、機械手臂的位置姿態以及所有物件的位置姿態等資訊，以此達到人類與機械手臂更直覺的互動，以及共同完成任務的效果。

訓練場景包含16,384個工作單元，場景內有3個待互動的物件、協作機械手臂以及兩個目的地。任務內容為指定其中一個物件進行協作機械手臂夾取，接著指定協作機械手臂需做的動作，動作為翻轉物件，最後再指定物件擺放的目的地，目的地分別設置在協作機械手臂的左右兩邊。深度強化式學習訓練所採用的演算法是近端策略優化演算法，並採用多層感知器(MLP)神經網路架構，架構包括四個隱藏層，第一和第二層隱藏層均包含1024個神經元，第三和第四層隱藏層均包含512個神經元。模型輸入總共為40個維度，分別代表三個待互動物件的x、y、z座標以及姿態以四元數表示、夾爪夾取點的x、y、z座標以及姿態以四元數表示、夾爪兩個關節的自由度狀態、指定要夾取的物件、指定要執行的任務、指定要放置物件的目的地、目的地的x、y、z座標以及姿態以四元數表示。輸出總共為7個維度，分別代表協作機械手臂夾爪的夾取點要移動的x、y、z座標以及x、y、z三個軸的旋轉以及夾爪的開合狀態。

● 狀態 (State)

在本研究中，我們從模擬環境中讀取所有訓練強化式學習代理所需要用到的所有資料的狀態。這些狀態總共包含40個維度，具體如下：

1. 待互動物件的x、y、z座標及姿態（21維）：

- 三個待互動物件，每個物件的空間位置由x、y、z三個坐標表示。
- 每個物件的姿態(Orientation)使用四元數(Quaternion)表示，包含四個分量。

- 每個物件的狀態由7個維度表示（三個坐標+四個四元數分量），三個物件共計21個維度。
2. 夾爪夾取點的x、y、z座標及姿態（7維）：
 - 夾爪的空間位置由x、y、z三個坐標表示。
 - 夾爪的姿態同樣使用四元數表示，包含四個分量。
 3. 夾爪兩個關節的自由度狀態（2維）：
 - 夾爪的2個關節的自由度狀態。
 4. 指定要夾取的物件（1維）：
 - 一個整數值，表示當前要夾取的特定物件。
 5. 指定要執行的任務（1維）：
 - 一個整數值，表示當前要執行的特定任務。
 6. 指定要放置物件的目的地（1維）：
 - 一個整數值，表示物件要最後要到達的目的地。
 7. 目的地的x、y、z座標及姿態（7維）：
 - 目標位置的空間坐標，使用x、y、z三個坐標表示。
 - 目標位置的姿態使用四元數表示，包含四個分量。

● 獎勵函數 (Reward Function)

獎勵函數是強化式學習中非常重要的概念。它用於定義系統在特定狀態下採取特定行動後所獲得的即時回饋或報酬。獎勵函數是一個從環境的每一個狀態和行動映射到一個實數值的函數。大部分的情況下，我們希望獎勵函數可以指導代理學習達到某個目標。這意味著對於代理採取的行動，獎勵函數應該給予正面的回饋，而對於不利於達到目標的行動，則給予負面的回饋。設計一個良好的獎勵函數是強化式學習中的一個關鍵挑戰。因為獎勵函數直接影響代理的學習過程和最終的行為策略。一個好的獎勵函數應該能夠在目標明確的情況下，引導代理學習到預期的行為。

因此，針對本研究需要訓練機械手臂完成的任務，設計了以下四個主要的獎

勵函數來引導代理學習，分別為：

1. 夾爪夾取位置及夾爪左右手指與物件平均距離的獎勵(dist reward)：

- 獎勵函數(3-1)旨在鼓勵代理學會讓夾爪夾取物件的位置靠近物件的中心位置。其計算方式包括三個距離指標：夾爪夾取物件位置與物件中心的歐幾里得距離 d ，夾爪左手指與物件中心點的歐幾里得距離 d_{lf} ，以及夾爪右手指與物件中心點的歐幾里得距離 d_{rf} 。當這三個距離的平均值越小時，獎勵值就會越高，從而促使代理學會將夾爪的位置調整得更靠近物件中心。

$$dist_reward = 1 - \tanh\left(10 \times \frac{d + d_{lf} + d_{rf}}{3}\right) \quad (3-1)$$

2. 物件當前距離目標位置的獎勵(aligned reward)：

- 獎勵函數(3-5)旨在鼓勵代理學會將物件夾取並靠近目標位置。首先計算夾爪獎勵(3-2)，計算方式為將夾爪的z坐標(z_{hand})與物件當前位置的z座標($z_{current}$)相減取絕對值後除上0.025，再透過clip函數將值限制在0到1之間，最後再用1減去算出來的值即可得到夾爪獎勵。接著計算物件當前位置與目標位置的距離 $align_target$ ，計算方式為物件當前位置的三維座標(p_1, p_2, p_3)與目標位置的三維座標(q_1, q_2, q_3)距離相減後取歐幾里得距離，將結果取負數後以指數函數計算，並乘上夾爪獎勵。此外，當物件當前位置與目標位置距離大於0.2m時，會乘上 obj_lifted ，即物件當前位置z座標($z_{current}$)大於物件初始位置的z座標($z_{original}$)0.02m時才會計算此獎勵。

$$eef_reward = 1 - clip\left(\frac{|z_{current} - z_{hand}|}{0.025}, 0, 1\right) \quad (3-2)$$

$$obj_lifted = (z_{current} - z_{original}) > 0.02 \quad (3-3)$$

$$align_target = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2} \quad (3-4)$$

$align_reward$

$$= \begin{cases} \exp(-align_target) \times eef_reward, & \text{if } align_target \leq 0.2 \\ \exp(-align_target) \times eef_reward \times obj_lifted, & \text{otherwise} \end{cases} \quad (3-5)$$

3. 物件姿態距離目標姿態的差距獎勵(flip reward)：

- 獎勵函數(3-6)旨在鼓勵代理學會將物件翻轉到指定的目標姿態。其中 $diff_angles$ 為物件姿態與目標姿態之間的四元數夾角差距，其計算方式為通過計算兩個四元數之間的旋轉差異來獲得它們的相對旋轉角度。首先，計算其中一個四元數的共軛，這會將四元數的虛部取反。然後，將這個共軛四元數與另一個四元數相乘，得到一個新的四元數，這個新四元數代表了從一個四元數旋轉到另一個四元數的相對旋轉。接著，通過計算這個結果四元數的虛部的範數，並使用反正弦函數將其轉換為角度值。最後，乘以2得到最終的旋轉角度，這個角度表示了從一個四元數到另一個四元數的最短旋轉角度。此獎勵只會在物件當前位置與目標位置距離小於等於0.2m時產生。

$$flip_reward = \left(1 - \tanh\left(\frac{diff_angles}{180}\right)\right) \times (align_target \leq 0.2) \quad (3-6)$$

4. 錯誤動作的懲罰(error penalty)：

- 獎勵函數(3-7)旨在讓代理學會在物件尚未抬起時，物件姿態不能與起始姿態差距太大。其中 C_L 表示物件是否被抬起、 C_Q 表示物件是否接近初始姿態、 F_O 表示物件是否翻轉完成。 I 是指示函數，表示條件是否成立。當條件成立時，指示函數 I 返回1，否則返回 0。 \neg 是否定的邏輯

運算，當其運算元為False時，結果為True，反之亦然。 \wedge 是邏輯中表示連接的邏輯運算，只有當兩個運算元都為True時，連接才為True。這個獎勵函數的目的是使代理學會將物件抬起後再翻轉，而不是在桌面上直接翻轉物件。

$$error_penalty = I(\neg C_L) \wedge I(\neg C_Q) \wedge I(\neg F_O) \quad (3-7)$$

接著是完成任務的條件 $success$ ，條件為物件當前位置與目標位置的距離 $align_target$ 小於0.02m，以及物件姿態與目標姿態之間的四元數夾角差距 $diff_angles$ 小於5度。

$$success = (align_target < 0.02) \wedge (diff_angles < 5) \quad (3-8)$$

最後計算總獎勵時，上述獎勵函數將分別乘上對應的獎勵尺度(Reward Scale)後相加(3-9)，並且判斷是否完成任務，給予對應的最終總獎勵 $final_reward$ ，如(3-10)所示。獎勵尺度的具體設置如表1所示，獎勵尺度指的是強化式學習中獎勵值的範圍或比例，影響代理訓練過程中獲取獎勵的大小和方式，調整獎勵尺度可以改善訓練的穩定性和效率。

$$\begin{aligned} total_reward &= dist_reward \times 0.1 \\ &+ align_reward \times 3.5 \\ &+ flip_reward \times 2.5 \\ &+ error_penalty \times (-0.08) \end{aligned} \quad (3-9)$$

$$final_reward = \begin{cases} 100, & \text{if } success \\ total_reward, & \text{otherwise} \end{cases} \quad (3-10)$$

表 1、獎勵尺度對應表

獎勵函數	獎勵尺度(Reward Scale)
夾爪夾取位置獎勵(dist reward)	0.1
物件距離目標位置獎勵(align reward)	3.5
物件姿態差距獎勵(flip reward)	2.5
錯誤動作懲罰(error penalty)	-0.08
成功獎勵(success reward)	100

● 動作(Action)

動作共包含7個維度，具體如下：

1. 夾爪夾取點要移動的x、y、z座標（3維）：
 - 夾爪需要移動到的新位置，由x、y、z三個坐標表示。
2. 夾爪的旋轉（3維）：
 - 夾爪需要沿x、y、z三個軸進行的旋轉角度，分別表示旋轉量。
3. 夾爪的開合狀態（1維）：
 - 一個浮點數，表示夾爪的開合程度。

在模擬環境中，使用Isaac Gym訓練的模型輸出的動作包括夾爪夾取點要移動的x、y、z座標以及夾爪的旋轉角度和開合程度。這些輸出會透過操作空間控制(Operational Space Control, OSC)直接計算出協作機械手臂的每個關節需要施加的扭矩，以實現對機械手臂的精確控制。

3.3 整合自然語言處理與強化式學習之協作機器人系統架構

將3.1自然語言處理模組以及3.2深度強化式學習模組整合，構成協作機器人系統，如圖6所示。該系統能夠接收語音指令，並通過語音轉文字模組將語音指令

轉換成文本，然後利用CKIP Transformers進行斷詞和詞性標記，最終經過任務提取模組提取出具體的任務資訊。這些任務內容、物件資訊及夾爪狀態一併輸入到強化式學習代理模型中，代理基於這些資訊生成相應的行動，從而達到透過語音指令控制協作機械手臂完成指定任務的效果。

首先，從模擬環境中讀取所有訓練強化式學習代理所需的資料狀態。這些狀態總共包含30個維度，具體如下：

1. 物件資訊(Object)：

- 待互動物件的x、y、z座標及姿態共21個維度，包含三個待互動物件。每個物件的空間位置由x、y、z三個坐標表示，並且每個物件的姿態使用四元數表示，共包含四個分量。因此，每個物件的狀態由7個維度（3個坐標+4個四元數分量）表示，三個物件共計21個維度。

2. 夾爪資訊(Gripper)：

- 夾爪夾取點的x、y、z座標及姿態共7個維度，夾爪的空間位置由x、y、z三個坐標表示，夾爪的姿態同樣使用四元數表示，共包含四個分量。此外，夾爪兩個關節的自由度狀態共2個維度，表示夾爪的兩個關節的自由度狀態。

以上合計30個維度的資訊僅包含物件和夾爪的基本狀態資訊。然而，為了完整描述強化式學習代理需要處理的所有資訊，我們還需考慮以下來自任務提取後結果的資訊，其中第4點目的地的x、y、z座標及姿態的資訊是根據第3點指定要放置物件的目的地的資訊轉換而成的：

1. 指定要夾取的物件（1維）：

- 一個整數值，表示當前要夾取的特定物件。

2. 指定要執行的任務（1維）：

- 一個整數值，表示當前要執行的特定任務。

3. 指定要放置物件的目的地（1維）：

- 一個整數值，表示物件要最後要到達的目的地。

4. 目的地的x、y、z座標及姿態（7維）：

- 目標位置的空間坐標，使用x、y、z三個坐標表示。目標位置的姿態使用四元數表示，包含四個分量。

因此，全部的資料狀態維度為物件資訊21個維度、夾爪資訊9個維度、指定要夾取的物件1個維度、指定要執行的任務1個維度、指定要放置物件的目的地1個維度，以及目的地的x、y、z座標及姿態7個維度，合計共40個維度，最後再將這40個維度的資訊送入強化式學習代理模型。

協作機器人系統的目的是實現從語音指令到機械手臂動作的全自動化流程，確保系統能夠準確理解並執行使用者的指令。這種整合不僅提高了機械手臂操作的精度和效率，還擴展了機械手臂在實際應用中的範圍，從而推動相關技術的進一步發展。透過這種綜合方法，系統不僅能處理更複雜的任務，還能在動態環境中靈活應對各種挑戰，實現更高效、更準確的操作。

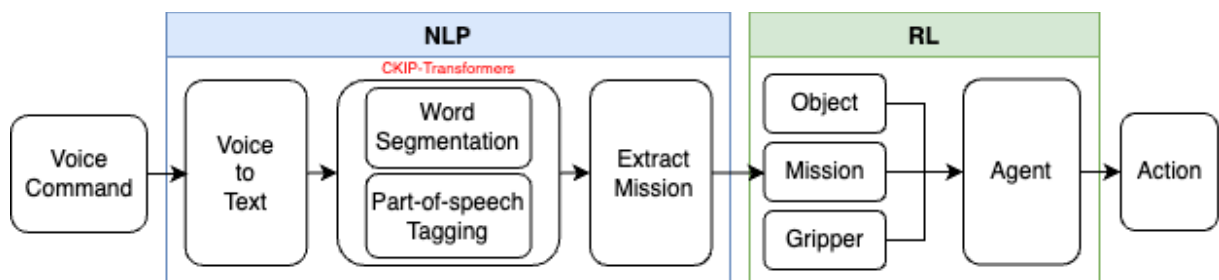


圖 6、協作機器人系統架構圖

第四章 實驗場景與結果

4.1 硬體設備與軟體

本實驗使用的硬體設備為一台用於進行深度強化式學習訓練的伺服器。實驗所用的軟體包括Nvidia Isaac Gym，實驗場景內所使用的協作機械手臂為TM5M-700，夾爪則是Franka Hand。

● 伺服器

伺服器用來進行協作機械手臂的深度強化式學習訓練，其規格如表2。這些高性能配置確保了訓練過程中的計算效率和穩定性，從而提升模型的訓練效果。


表 2、伺服器規格表

處理器型號	13th Gen Intel(R) Core(TM) i7-13700
處理器速度	5.10 GHz
記憶體	64GB
獨立顯卡型號	Nvidia GeForce RTX 4090
獨立顯卡記憶體	24GB

● 協作機械手臂

協作機械手臂採用的是由達明機器人(Techman Robot)所製造的TM5M-700六軸機械手臂，其規格及外觀如表3。TM5M-700是達明機器人最緊湊的協作機器人，可以輕鬆整合到任何生產線中。該機器人設計有內建視覺系統，專門用於滿足小零件組裝和消費電子產品及消費品生產過程中所需的靈活生產需求。我們的機器人為中小型企業提供了極大的多功能性。TM5M-700的尺寸也使其能夠快速部署，並輕鬆適應現有的工廠環境[29]。

表 3、TM5M-700 規格表[29]

		
重量	22.1kg	
最大容許負載	6kg	
可達範圍	746mm	
關節活動範圍	J1, J6	+/-270°
	J2, J4, J5	+/-180°
	J3	+/-155°

- 夾爪

裝載在協作機械手臂上的夾爪採用的是Franka Robotics生產的一種電動兩指平行夾爪，其規格如表4。

表 4、Franka Hand 規格表[30]

夾持力範圍	30~70(N)
開閉行程	80(mm)
開閉速度	50(mm/s)
重量	730(g)

- Nvidia Isaac Gym

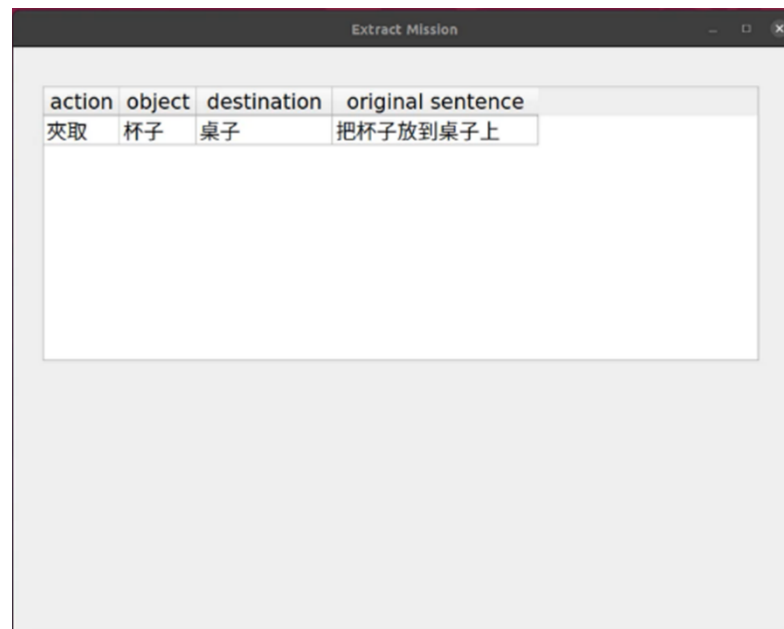
Nvidia Isaac Gym是一個高效能且靈活的仿真平台，專為加速強化式學習算法的開發而設計。該平台利用GPU加速技術，能夠並行運行數千個仿真環境，大幅提升仿真效率，縮短模型訓練時間。Nvidia Isaac Gym內置高真實感的物理引擎，能夠模擬複雜的剛體動力學、柔性物體及流體動力學，提供真實的仿真結果。平台提供多種預設機器人模型和環境，支持自定義環境，並與主流強化式學習框架（如PyTorch和TensorFlow）整合，簡化算法開發流程。作為Nvidia Isaac SDK的一部分，Isaac Gym與其他模組（如Isaac Sim）聯動，提供從仿真到實際應用的一體化解決方案。其豐富的API允許用戶進行深度自定義和擴展，適應不同研究方向和應用需求。Nvidia Isaac Gym憑藉其高效能仿真、真實物理模擬、強大的整合性和靈活性，為強化式學習和機器人研究提供了堅實的技術支持，是該領域的重要工具[31]。

4.2 模擬環境設置

本實驗自然語言處理部分開發了一個使用者介面，如圖7所示，此介面是為了讓使用者能夠直接進行語言指令操作，並能夠即時查看處理後的指令是否準確。

介面分為四個欄位：

1. 動作
2. 物件
3. 目的地
4. 使用者所講的指令



action	object	destination	original sentence
夾取	杯子	桌子	把杯子放到桌子上

圖 7、自然語言處理介面

本實驗強化式學習訓練部分利用Nvidia Isaac Gym搭建了一個大規模並行訓練環境，如圖8所示。場景包含16,384個工作單元，每個單元由桌子、三個物件、機械手臂及其底座、兩個當作目的地的長方體所組成。

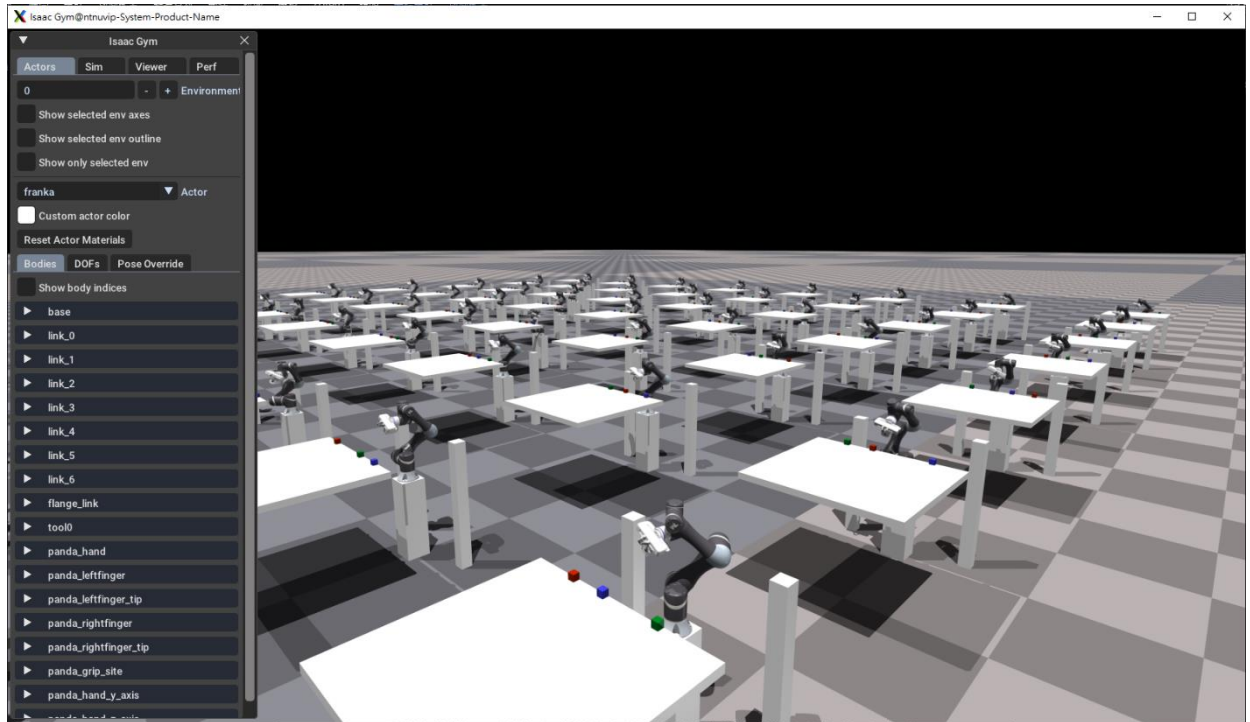


圖 8、Nvidia Isaac Gym 訓練環境

在每個工作單元中，桌子用於放置操作物件，提供機械手臂的工作平台。這些物件的位置和姿態在訓練過程中會隨機變化，這種設計旨在模擬真實工作場景中的多樣性和複雜性。通過這種方式，訓練過程可以有效地解決模擬環境(sim)到真實環境(real)轉移中的挑戰，確保機械手臂在真實場景中的穩定性和準確性。

- **機械手臂**

機械手臂的位置會隨機變化，這取決於其底座的隨機位置變化。姿態的生成是通過將預設的姿態乘上一個隨機範圍內的浮點數來實現的，這樣可以生成不同的姿態。此外，每個關節的初始角度也會進行調整，使得機械手臂能夠從不同的初始位置和姿態開始運行。

- **其他物件**

桌上的三個待夾取物件、桌子、手臂底座以及兩個目標底座的位置都是以預設位置為基礎，然後在三維座標x、y、z上分別加上隨機誤差值。這些誤差值的範圍在-0.05m到0.05m之間變動。除了桌上的三個待夾取物件外，其他物件的姿態

都是固定的。待夾取物件的姿態僅在z軸上隨機生成，其姿態的計算方式是將預設姿態乘上一個隨機範圍內的浮點數來生成。

這種隨機初始化方法有助於強化式學習模型更好地適應各種不同的場景和操作要求，提高其在實際應用中的泛化能力和靈活性。

4.3 模擬環境訓練結果

本實驗強化式學習所採用的神經網路架構為MLP，激勵函數採用的是ELU(Exponential Linear Unit)。ELU常用於神經網路中，其設計特點包括：平滑性，有助於網路的學習和收斂；允許負輸出，幫助在負激勵情況下的學習；平均值接近零，能加速神經網路的訓練，特別是在深層網路中。這些特性使得ELU在某些情況下比ReLU或Leaky ReLU更具優勢。詳細參數設置如下表5所示。

表 5、神經網路參數

項目	數值
Learning rate	0.0005
Minibatch size	32768
Activation function	ELU
Units	[1024, 1024, 512, 512]

圖9為強化式學習的訓練結果圖，這張圖展示了強化式學習訓練過程中累積獎勵隨時間（步數）變化的趨勢。在訓練初期（0到約2000步），累積獎勵迅速上升，表明代理在這段時間內學習到了有效的策略並顯著提高性能。從約2000步到約4000步，累積獎勵繼續上升但速度變慢，並開始出現波動，這表明代理正在微調策略以應對更複雜的情境。從約4000步到約6000步，累積獎勵趨於穩定並在一定範圍內波動，顯示代理的策略已基本成熟並且在大多數情況下表現良好。這張圖清晰地展示了代理在訓練過程中的學習效果和性能變化，表明訓練基本順利且代理已學習到穩定有效的策略。最終，在模擬環境中進行了500次測試，訓練結果的成功率達到了70.06%。

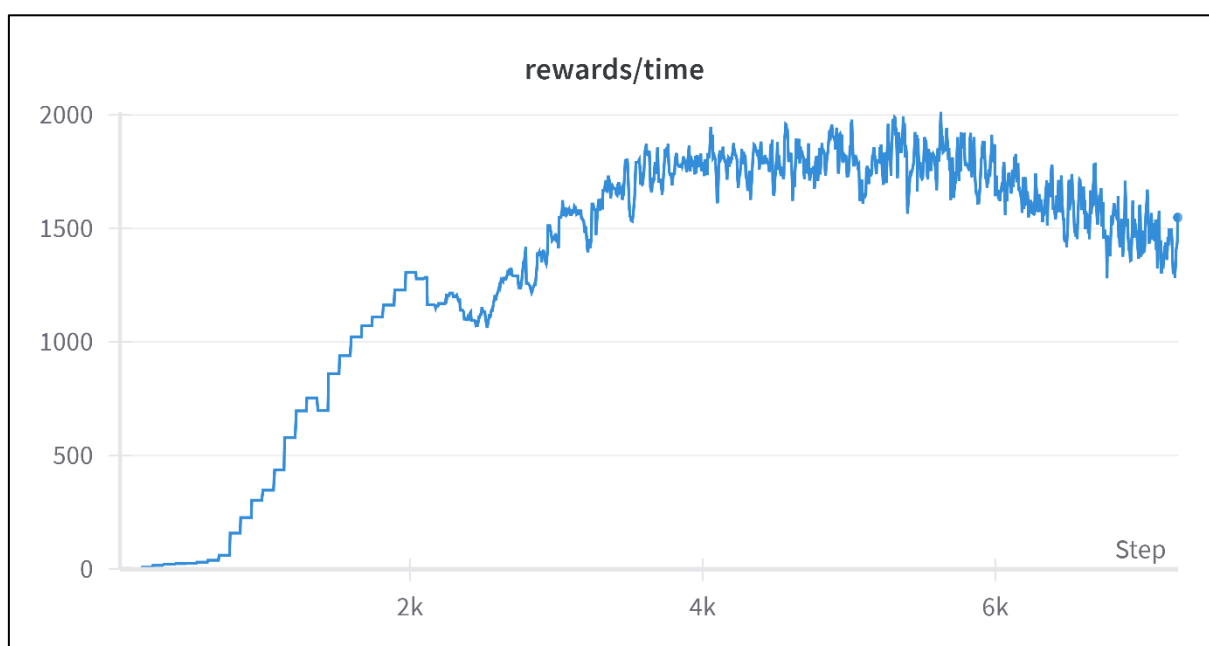


圖 9、強化式學習訓練結果

下列圖片為強化式學習訓練過程，從圖中可以看到代理確實有學會讓機械手臂執行所指定的任務內容。

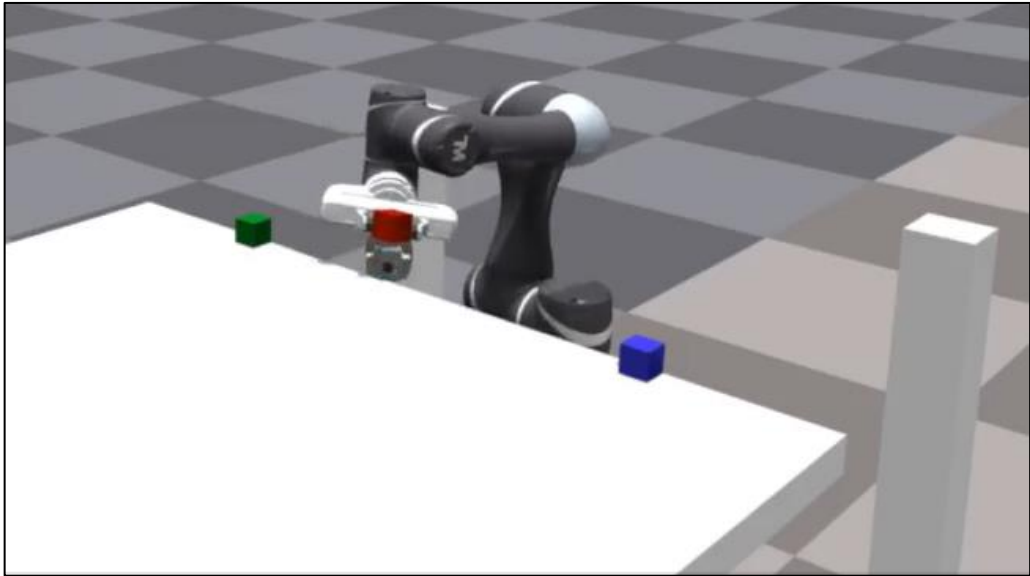


圖 10、夾爪靠近物件

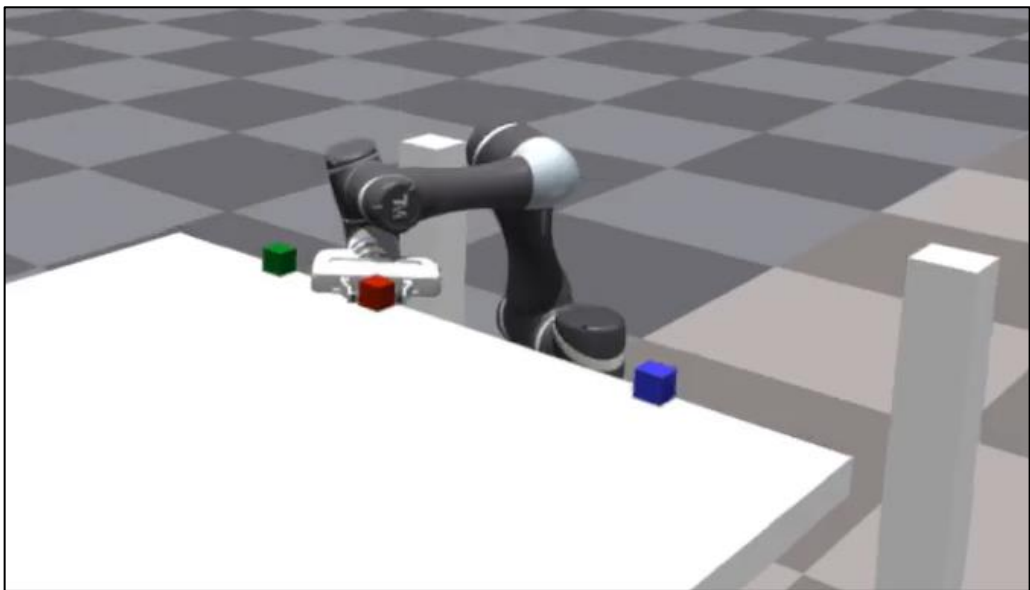


圖 11、機械手臂夾起物件

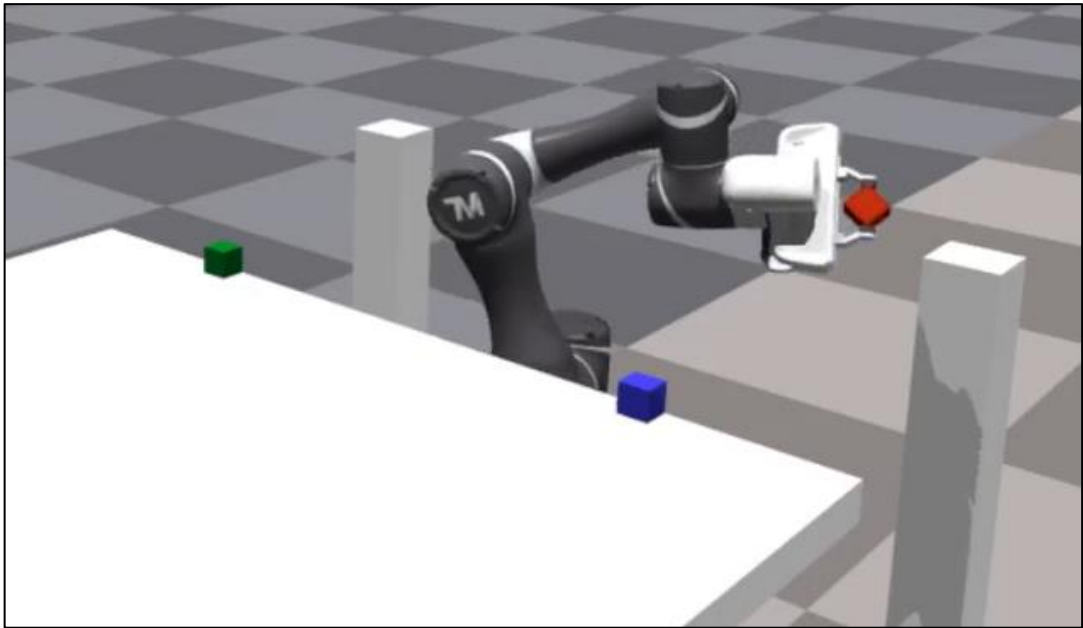


圖 12、物件靠近目標位置

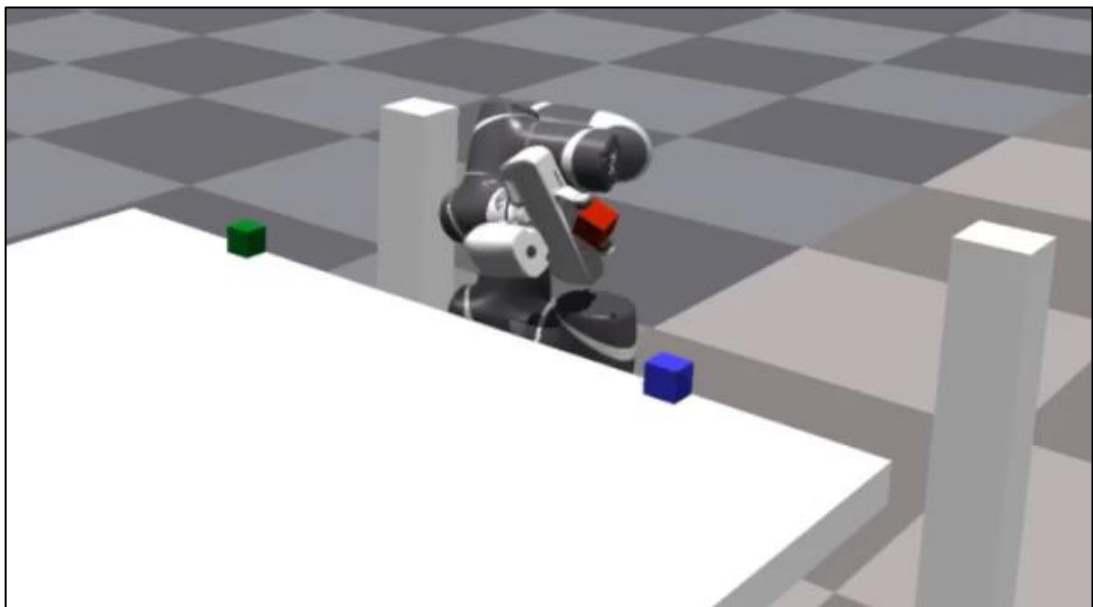


圖 13、機械手臂翻轉物件

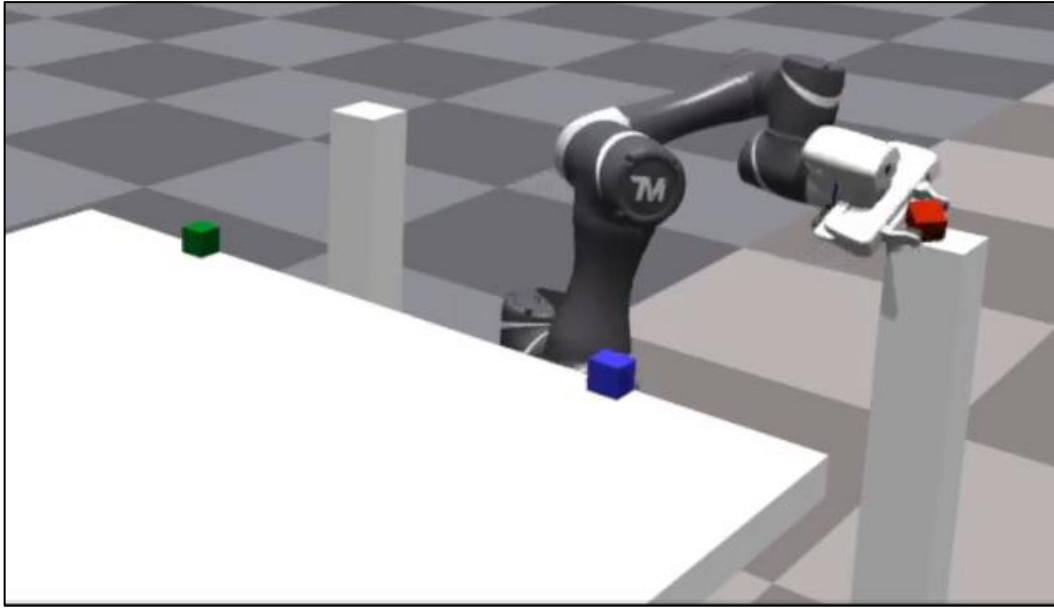


圖 14、物件到達目標位置及姿態(完成任務)

第五章 結論與未來展望

5.1 結論

本論文通過整合自然語言處理和強化式學習技術，成功開發出一套可以利用語音指令去控制協作機械手臂執行指定任務的系統。該系統能夠理解人類語言指令，並在動態環境中自主學習和執行複雜任務。研究成果顯示，透過語音轉文字技術、CKIP Transformers語言模型進行詞性分析和任務提取，以及Nvidia Isaac Gym強化式學習平台進行訓練，協作機械手臂能夠達到執行不同的任務需求。

在模擬環境中，本論文分別對自然語言處理以及強化式學習開發，在自然語言處理方面，開發了一個介面，能夠讓使用者清楚知道語音指令經過處理後的任務內容，在強化式學習方面，透過將隨機指定物件及任務參數送進模型，使強化式學習代理能夠學會讓協作機械手臂執行不同的任務。

5.2 未來展望

在自然語言處理方面，目前的任務提取模組雖然能夠提取並整理任務內容，但僅能針對相對簡單的語句進行處理並正確提取任務內容。這種限制使得系統在面對日常生活中更為複雜和多樣化的語句時，表現出一定的局限性。未來希望能夠設計並優化此任務提取模組，使其能夠處理更加複雜的語句結構，涵蓋更廣泛的語言模式，從而讓使用者能夠用更貼近日常生活的說話方式來操作系統。這樣一來，不僅能夠提高系統的實用性和易用性，也能夠讓更多的使用者受益，實現更自然和人性化的互動體驗。

在強化式學習方面，目前的任務內容僅包括選擇物件夾取或翻轉到指定目的地，且物件的形狀也非常相似。這樣的設定在某種程度上限制了協作機械手臂的應用範圍和靈活性。未來希望能夠擴展和多樣化訓練任務內容，例如訓練協作機

械手臂能夠夾取形狀各異、材質不同的物件，或者執行更加複雜和精細的動作，諸如組裝、分類、排序等。這些改進將使協作機械手臂能夠應用在更多不同的環境和場景中，例如工業自動化、醫療輔助、家庭服務等，從而大大提升其實用價值和市場競爭力。透過這些努力，最終實現協作機械手臂的智能化和多功能化，滿足更多元化的需求。

參 考 文 獻

- [1] L. Rabiner and B. Juang, "An introduction to hidden Markov models," *IEEE ASSP Magazine*, vol. 3, no. 1, pp. 4-16, Jan. 1986.
- [2] C. Sutton and A. McCallum, "An introduction to conditional random fields," *arXiv preprint arXiv:1011.4088*, 2010.
- [3] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intelligent Systems and Their Applications*, vol. 13, no. 4, pp. 18-28, July-Aug. 1998.
- [4] F. Almeida and G. Xexéo, "Word embeddings: A survey," *arXiv preprint arXiv:1901.09069*, 2023.
- [5] S. Sivakumar, L. S. Videla, T. R. Kumar, J. Nagaraj, S. Ithal, and D. Haritha, "Review on Word2Vec word embedding neural net," in *Proc. International Conference on Smart Electronics and Communication (ICOSEC)*, Trichy, India, 2020, pp. 22-26.
- [6] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *Proc. Conf. Empirical Methods in Natural Language Processing (EMNLP)*, Doha, Qatar, 2014, pp. 1532-1543.
- [7] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," *arXiv preprint arXiv:1607.01759*, 2016.
- [8] Y. Kim, "Convolutional neural networks for sentence classification," *arXiv preprint arXiv:1408.5882*, 2014.
- [9] R. M. Schmidt, "Recurrent neural networks (RNNs): A gentle introduction and overview," *arXiv preprint arXiv:1912.05911*, 2019.
- [10] S. Wang and J. Jiang, "Learning natural language inference with LSTM," *arXiv preprint arXiv:1512.08849*, 2016.
- [11] B. Zhang, D. Xiong, J. Xie, and J. Su, "Neural machine translation with GRU-Gated attention model," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 11, pp. 4688-4698, Nov. 2020.
- [12] G. Brauwiers and F. Frasincar, "A general survey on attention mechanisms in deep learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 4, pp. 3279-3298, Apr. 2023.
- [13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *arXiv preprint arXiv:1706.03762*, 2023.

- [14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2019.
- [15] G. Yenduri, R. M. C. Selvi, S. Y. G. Srivastava, P. K. Reddy, D. Raj, R. H. Jhaveri, B. Prabadevi, W. Wang, A. V. Vasilakos, and T. R. Gadekallu, "GPT (Generative Pre-Trained Transformer)— A Comprehensive review on enabling technologies, potential applications, emerging challenges, and future directions," *IEEE Access*, vol. 12, pp. 54608-54649, 2024.
- [16] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text Transformer," *Journal of Machine Learning Research*, vol. 21, no. 140, pp. 1-67, 2020.
- [17] "GPT-4," OpenAI. [Online]. Available: <https://openai.com/index/gpt-4/>. [Accessed: June 30, 2024].
- [18] "Llama2," Meta AI. [Online]. Available: <https://llama.meta.com/llama2/>. [Accessed: June 30, 2024].
- [19] LangChain Documentation. [Online]. Available: <https://python.langchain.com/v0.2/docs/introduction/>. [Accessed: June 30, 2024].
- [20] C. Yu, K. Li, Y. Zhang, J. Xiao, C. Cui, Y. Tao, S. Tang, C. Sun, and C. Bi, "A survey on machine learning based light curve analysis for variable astronomical sources," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 12, no. 5, 2021.
- [21] Y. Duan, X. Chen, R. Houthoofd, J. Schulman, P. Abbeel, "Benchmarking deep reinforcement learning for continuous control," *arXiv preprint arXiv:1701.07274*, 2017.
- [22] H. Tan, "Reinforcement learning with deep deterministic policy gradient," in *Proc. International Conference on Artificial Intelligence, Big Data and Algorithms (CAIBDA)*, Xi'an, China, 2021, pp. 82-85.
- [23] T. Haarnoja, A. Zhou, P. Abbeel, and S. Levine, "Soft Actor-Critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor," *arXiv preprint arXiv:1801.01290*, 2018.
- [24] J. Schulman, P. Abbeel, and X. Chen, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [25] "Proximal Policy Optimization," Videh Raj Nema. [Online]. Available: <https://aarl-ieee-nitk.github.io/reinforcement-learning,/policy-gradient-methods,/sampled-learning,/optimization/theory/2020/03/25/Proximal-Policy-Optimization.html>. [Accessed: June 26, 2024].

- [26] A. Angleraud, A. M. Sefat, M. Netzev, R. Pieters. "Coordinating shared tasks in human-robot collaboration by commands," *Frontiers in Robotics and AI*, vol. 8, 2021.
- [27] C.-C. J. Hsu, P.-J. Hwang, W.-Y. Wang, Y.-T. Wang, & C.-K. Lu, "Vision-based mobile collaborative robot incorporating a multicamera localization system," *IEEE Sensors Journal*, vol. 23, no. 18, pp. 21853-21861, Sept. 15, 2023.
- [28] "CKIP Transformers" CKIP, Institute of Information Science, Academia Sinica.
[Online]. Available: https://ckip.iis.sinica.edu.tw/project/language_model. [Accessed: June 30, 2024].
- [29] TM5M-700. [Online]. Available: <https://www.tm-robot.com/zh-hant/tm5-700/>.
[Accessed: April 26, 2024].
- [30] Franka Hand Product Manual. [Online]. Available:
<https://www.toyorobot.com/Product/Series/CHG2>. [Accessed: April 26, 2024].
- [31] Nvidia Isaac Gym. [Online]. Available: <https://developer.nvidia.com/isaac-gym>.
[Accessed: April 26, 2024].

自 傳

我是駱忠湧，大學就讀國立屏東大學電腦與通訊學系，在大學期間，我對程式設計有著濃厚的興趣，並且專注於VR和AR相關技術的研究與應用。我積極參與各類比賽和展覽，並在課餘時間學習和實踐各種程式相關技術，從而進一步提升了自己的技能和知識。碩士就讀的是國立臺灣師範大學電機工程學系，在碩士期間，我進入了CIRLAB實驗室，在實驗室中主要是以協作機械手臂開發及整合為主要的研究，同時也學到了深度學習、強化式學習等多種知識。

學 術 成 就

1. 論文發表

- (1) C.-Y. Lo and C.-C. Hsu, "Development of collaborative robots by integrating natural language processing and reinforcement learning," in *Proc. of the 2024 National Symposium on System Science and Engineering (NSSSE 2024)*, June 2024.