三、研究計畫內容(以中文或英文撰寫)：

(一) 研究計畫之背景。

背景

在科幻小說和電影的描繪中，未來的機器人已具備成熟的認知與行為智慧，能夠在人類社會中以「類人」的外表和行為模式與人類共處，同時又有超越人腦的算力和智力。儘管此種想像目前仍然是遙遠未來的願景，機器人成為人類社交空間中的自然存在還有很長的路要走，但在認知科學、社會科學、與資料科學等跨領域協作發展的推動下，社交機器人研究的快速進展正在幫助我們更接近這一現實。

「人機共生」並不是一個新鮮的議題，這個概念最早可追溯到 1960 年計算先驅 J.C.R. Licklider 以此為題的開創性論文[1]， 他認為"人機共生是人與電腦合作互動的必然發展，人類和電腦之間將產生非常緊密的夥伴關係，目的是使人和電腦能夠合作決策、控制複雜的情況，而不是僵化的依賴電腦預定的程序.....共生夥伴關係將比單靠人類更有效地執行智力操作"，這一觀點預言了認知計算將是可編程計算的必要和自然演變。隨著 AI 技術蓬勃發展，大規模並行計算以及海量結構化、非結構化資料的積累為認知計算奠定下基礎，2011 年，首個認知服務系統 IBM Watson[2]問世，認知計算的研究開展進入了一個嶄新的時代。

近年來，機器人製造、感測、控制技術等不斷精進，機器人不僅能夠因應外在環境的變動而做出回應，還能夠通過觀測人類情緒狀態而作出安撫的行為，此類關注社交魅力的個人化輔助型機器人被稱為社交型機器人，人們開始嘗試將其投入零售、教育、醫療等應用場域、完成更多樣型之任務，如提供人機界面諮詢[3-5]、教導學習[6]、提供生活協助或是陪伴功能[7-9]。這些應用有助於改善社會少子化、老齡化導致勞動人力漸不足之問題，同時極大地改善了人們對 AI 機器人的接受度。未來社會，機器人將進入人們生活的方方面面，不只是作為生產協作的自動化工具，各種型態的社交型機器人將逐漸扮演個人管家、陪伴者的角色。因此，人們越來越需要能夠在日常情況下與人安全互動的機器人，保障使用者的生命、財產安全，亦須在互動中關注個人化的心理需求，不侵犯個人隱私，不勾起創傷回憶等。然而，設計一種能夠贏得人類信賴的社交型輔助機器人，使其在人類社會中學習多元知識、社交行為，成為人類的陪伴者或專屬顧問，是一項極具挑戰性的任務，這要求機器人具備深層認知能力，能夠預測自己行為的影響以及周圍人的行為和需求。為了實現這一目標，亟需開發一個整合了物理系統與經驗知識的認知決策架構，以實現在無約束環境的自然互動、獲取和使用經驗的需求控制、反饋學習等目標。此外，一般的服務者或照護者很少具備廣泛的知識背景以回應複雜多變的需求，亦少有能夠關注使用者心理狀態，因此在機器人的認知方面引入自然科學、社會科學、心理學知識，是一個必然的趨勢。

目的及重要性

「**我們在製造人造人類並與之互動的道路上走得越遠，我們對自己的了解就越多。**」Broadbent 在心理學領域第一篇與社交機器人相關的綜述[10]中這樣寫道。人類有一種根本的創造傾向，而最終的創造產物是另一個(超級)人類。在眾多技術中，以人工智慧(AI)技術構建模擬人類行為、能夠與人類進行自然互動的類人機器人顯示出其廣泛的應用前景。然而，只有更多地了解人類的行為、認知與情緒，才能製造出能夠被人類社會所接納的「類人」機器人，心理學家因此參與並引入認知理論到機器人與人互動(HRI)研究中，幫助工程師更好地理解和模擬人類，用適當的方法進行實驗，並開發認知治療輔助機器人。儘管大量研究還處於探索階段，這一跨學科合作主題已顯示出巨大的潛力為 HRI 領域做出貢獻，這將是一個

雙贏的局面：人類行為的研究能夠為機器人的構建提供參考，而機器人的應用與測試則讓我們得以通過互動更了解人類的認知、情感和行為。

此計畫擬構建一個「**具有深度(層)認知能力之 AI 社交型機器人系統**」，設計並實現此「類人」高度智慧的機器人之自主認知與自主行為能力，通過觀察和實踐蒐集之多模態資料，學習和推理如何在複雜世界辨識多元的服務或情感需求、實現複雜目標，期望機器人能夠自主拓展知識邊界、深入理解環境和人類社會，主動偵測人們的行為表現及身心狀態並提供適切的服務，同時，與人類建立信任關係，在提供貼心的輔助時，亦引導人們對事件做出正確的決策、鼓勵人們以積極的方式行動。

國內外有關本計畫之研究情況

國內機器人研究涵蓋眾多研究領域，包含如服務型機器人系統、機器人定位、路徑規劃、自主導航、視覺感知、人機互動、運動控制等，近幾年研究有羅仁權教授建立的圖像描述服務型機器人[11]來輔助有視覺障礙人士，宋開泰教授透過實現內窺鏡自主定位來改善輔助手術的人機協作過程[12]，並開發機械手臂即時避碰系統[13]和人機協作導引車[14]，而胡竹生教授有磁性定位的相關研究[15]與使用深度學習技術應用於工業影像瑕疵檢測[16]，以及王文俊教授透過障礙物辨識與距離偵測設計的穿戴式導盲裝置[17]，翁慶昌教授探討雙臂機器人運動規劃[18]與物件重新定位[19]，和蔡清池教授對於人機協作的機械手臂速度控制策略[20]等。國內機器人的研究多側重在機器人透過環境理解來完成特定任務，目前較少有著重建立同時具備環境的空間認知和人類的社交認知的社交機器人系統，以實現「類人」的深度(層)認知能力。

近年來，社交機器人被應用到多個場域，包括老年人和認知障礙人士的陪伴者、教育環境中的機器人，以及作為支撐認知和行為改變介入療法的工具[21]。將機器人應用在人口結構老化的議題上，醫療照護機器人與家用陪伴型機器人市場也是臺灣產業發展的目標。但是，迄今我們過度直覺地把機器人及其他 AI 技術當作人類一樣互動，試圖感知它們的人類特徵，包括思想和情感，卻發現他們並不像人類一樣思考，也尚未被賦予與人類相同的道德權利和責任。此外，機器人的社會情感互動功能遠低於人類水平，互動過程中不可避免地會出現錯誤，這些錯誤可能會讓人們對機器人的感覺產生重大影響。在社會情感維度上的錯誤方面，先前對 HRI 發生錯誤的研究主要集中在機器人被要求的功能出了錯誤，例如導航。然而，遵守和適應共同商定的社會規範對於建立社會關係至關重要。因此，當機器人違反社會規範時，它可能會顯著影響人們對機器人的好感與信賴。了解這種影響可能是推進當前 HRI 研究的關鍵，尤其是在需要積極的人機關係的縱向場景中[22]。關於如何克服此類錯誤，我們整理了與以下主題相關的社交機器人研究：

(1) 多層級地圖與意圖偵測(空間認知)

環境三維結構理解和導航是行動機器人的關鍵認知功能，使其能夠學習環境的表示並通過及時感知資訊(例如視覺觀察)在空間中移動。至今為止，主流的建圖與定位(SLAM)方案[23]多是基於像素層級的特徵點來提取路標，與之不同的是，人類是通過物體在圖像中的運動來推測相機的運動，而非特定像素點，由此啟發，部分研究者試圖利用基於物件資訊的方案來實現 SLAM，通過引入語義資料生成高層級的地圖(Semantic Mapping)，同時，將語義特徵用於優化閉環檢測、Bundle Adjustment 以提高定位精度[24]。

此外，人們向機器人尋求幫助時，並不會使用技術手段設置精確目標位置，而是使用自然語言指令以提示目標所具有的語義特徵[25-27]。因此，機器人必須能夠將自然語言中的語義與環境地圖建立對應關係，即建立多層級地圖，並將指令語句解構為一系列導航任務。多層級地圖所涵蓋的通用知識，亦能幫助機器人在陌生環境中進行轉移學習。

利用語義空間的簡潔表示，也能夠克服現有基於視覺的地圖繪製和導航模型之內存需求的問題，而內存需求隨著探索時間的持續和與探索策略相關的迂迴路徑線性遞增。Chen, Kevin 等人提出了一種基於自組織(self-organizing)神經網路的增量拓撲映射和導航框架(framework) e-TM[28]，將探索軌跡明確建模為以“事件”序列作為地標情境記憶，在回溯記憶時將環境佈局的嵌入知識轉移到空間記憶中，即編碼地標之間的拓撲關係。融合自適應共振理論 (ART) 網路作為兩個記憶模塊的構建基石，可以將多個輸入模式概括為記憶模板，因此提供簡潔的空間表示並支持通過推理發現新的捷徑。對於導航，e-TM 應用轉移學習範式將人類演示整合到預訓練的運動網路中，以實現更流暢的運動。基於模擬 3D 環境 VizDoom 的實驗結果顯示，與最先進的模型半參數拓撲記憶 (SPTM) 相比，e-TM 顯著降低了導航的時間成本，同時學習了更稀疏的拓撲圖。

(2) 自主行為符合社交準則、機器人與人建立互賴關係(社交認知)

實際上，在認知機器人學中，機器人的身體不僅僅是用於物理操作或運動的工具，它還是認知過程的一個組成部分。因此，認知機器人是一種具形的認知形式，它利用機器人的物理形態、運動學和動力學，以及它的運行環境，來實現其自適應預期互動的關鍵特徵。認知機器人學是一門多學科的科學，借鑒了自適應機器人以及認知科學和人工智能的研究，並經常利用基於生物認知的模型[29]。認知機器人通過感知環境、關注重要事件、計劃做些什麼、預測自身及他人行為的結果以及從由此產生的互動中學習實現複雜的目標，在自主活動方面，通過不斷學習、推理和分享知識來應對自然環境固有的不確定性，在與人互動方面，能夠從他人的角度看待世界，因此可以預測該人的預期行為和需求，從而更好地實現互動。國外研究中，已有具有認知能力的機器人用導航與建圖和取物運輸[30]、健康照護[31]等。最新研究結果顯示陪伴型機器人作為認知機器人的一支有極大潛力應用於兒童自閉症診斷[32]、老人認知治療[33]等課題，隨著自然語言技術的發展，機器人得以運用語音辨識、情緒辨識、對話生成等手段達成與使用者進行自然對話的目標，以自然語言指令為導向的機器人控制方法[34]，亦有研究者將聊天機器人應用於心理學測試[35]。多模態機器學習的發展更賦予了機器人通過大量資料學習通用知識的機會，例如，從產品開箱測試中學習物件可用性知識、從圖文資料中學習通用知識[36]等。可以預見，未來科技高度發達的社會，人們對 AI 技術與機器人的預期不會只停留在完成特定任務的工具，而是一個可以自然互動的個人化的 AI 助理。

機器人與人的互動過程通常都是十分被動的，然而實驗中我們發現，具有情感障礙的人群更傾向於給予被動的回應，此時，缺乏主動性設計的機器人會表現出無響應或錯誤響應的行為，導致不良的互動體驗，因而無法與人建立起互賴關係[37]。因此，為社交型機器人構建理解、學習人類社交行為並作出積極引導的機制，是亟待解決的課題。另一方面，人類行為包含了情緒、動機與行為表現等多個面向，隨著與人互動的機器人數量逐步增加，其中一些需與社會敏感族群互動(例如身障者、老年人、癡呆症患者和自閉症患者)以提供諮詢、協助和關懷等。先前的研究表明，人類較喜歡能夠表現出符合人類社會期待(例如同情心)的機器人，使機器人與人互動上更有溫度、更具同理心、更具社交性及親和力，因此機器人應能在移動、對話、特定輔助功能中顯示其社交友善性。未來少子化、老齡化社會中孤立的情形會更加嚴峻，此時機器人有機會充當陪伴者、傾訴者，因此需要感知人類情緒、理解人類行為、並能基於同理心做出回應，這也是最近被越來越多研究者關注的課題，儘管在這一領域尚未獲得突破性的進展。

(3) 提供個人化的記憶、心理輔助服務(人物印象、基於生物學/心理學模型的引導輔助)

人類具有獨特的記憶模型包含感覺記憶、短時記憶、與長期記憶，其工作機制能解釋很多人類記憶過程中的心理現象，儘管至今尚有部分爭議，但經典模型[38]的主體思路仍得到

許多實驗的證明，並且開發了後期大量的記憶領域的研究。後期在短時記憶基礎上拓展出的工作記憶理論[39]，在人類高層次的認知行為中，如閱讀、理解和推理，扮演著重要的角色。機器人作為一個智慧助手，對人類行為的理解不能僅侷限在表觀現象的覺察，更應關注心理過程的推演，從而找到問題的癥結，或是以符合人類習慣模式的方式，引導個人化的自我覺察、自我認知，使得機器人在未來能夠滿足個人化的情感諮詢的需求[40]。

(4) 認知計算

然而，儘管已有大量研究投入到社交型機器人的研究，少有能夠於複雜環境中實現自主管理以覺察、規劃、執行複雜任務的機器人，原因在於缺少具有可拓展性的自主認知與決策架構，而這一架構又仰賴於認知計算(Cognitive Computing)的發展，目前已有與元認知相關的研究[41,42]發表，並應用於教育領域以探索認知機器人如何幫助人類訓練思考[43-45]。認知計算(Cognitive Computing, 2011-)作為第三個計算時代[46]，與前兩個時代，製表時代(Tabulating Computing, 1900s-1940s)、可編程計算時代(Programming Computing, 1950s-present)，有著根本性的差異。因為認知系統會從自身與數據、與人的互動中學習，所以能夠不斷自我提高。因而，認知系統絕不會過時。它們只會隨著時間推移變得更加智能，更加寶貴。這是計算史上最重大的理念革命。隨著時間推移，認知技術可能會融入許多 IT 解決方案和人類設計的系統之中，賦予它們一種思考能力。這些新功能將支持個人和組織完成以前無法完成的事情，比如更深入地理解世界的運轉方式、預測行為的後果並制定更好的決策。

原創性、預期影響性

基於上述研究進展之討論，本計畫擬構建一個工作於無約束環境的「具有深度(層)認知能力之 AI 社交型機器人系統」，嘗試解決目前與未來社交型機器人利用自主認知與自主行為規劃，自處於複雜環境、實現複雜目標的問題。本計畫旨在以探索人類認知、心智、行為等模式為切入點，設計並實現社交型機器人的自主認知與行為，包含空間認知與自主移動、人類社交行為認知、體現同理心之行為設計、因果推論學習、知識邊界的自主拓展等。

探索「類人」高度智慧機器人與人類的互動，能夠推動認知科學的發展，進而人們能夠創造出更符合人類認知行為模式的認知機器人之設計。

(二) 研究方法、進行步驟及執行進度。

本計畫基於機器人的自主認知與決策架構所面向的不同認知範疇，訂定了四階段性的里程碑目標，分成四年執行，依循序漸進的方式陸續開發關鍵的功能模組，期能最終完成研發「具有深度(層)認知能力之 AI 社交型機器人系統」。

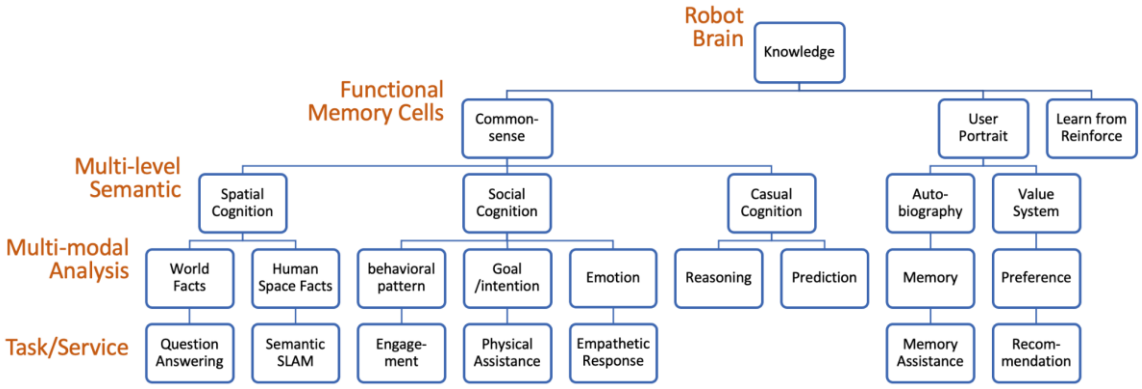


圖 1-1、認知模組框架

圖 1-1 描繪了一個完整且具深度認知能力的機器人應該具備有的功能模組架構，此功能框架(framework)是剖析人類(human agent) 認知能力[47]後所啟發而得，換言之，若能成功地在機器人上重置這框架，那所設計之社交機器人(robot agent)將會好比是 human agent 的數位孿生(digital twin)，屆時機器人與人類就有可能以自然方式在人類的社會中互動。

上述架構囊括了三個方面：

1). **通用知識 (commonsense)**：機器人首先需要能在無約束的複雜環境中建立一多層級概念圖，使得機器人能基於該地圖在人們熙來攘往的場域進行精準的社交導航而不會迷航，其次，機器人能透過多模態(modality)感測器(如 LiDAR、RGB-D 相機、距離感測器、麥克風陣列等)利用深度感知原理自主理解環境及周遭人群以便進行與人溝通或提供服務；

2). **人物印象 (user portrait)**：機器人作為人類陪伴者、照護者，需要通過長期的互動瞭解使用者的過往與偏好，基於這理解，人類與社交機器人才能持續和諧交往，且建立所謂的信任、友誼，進而建立深化的交情，如此機器人才能勝任伴侶與照顧者的角色；

3). **反饋學習 (learn from reinforcement)**：為使社交機器人能更融入人類的社會、其所表現的行為與和人的對談能更被人類所接受，則機器人要能自與周遭環境或人類互動過程中得到的反饋進行學習，雖不見得可達到人類社會中"觸類旁通、舉一反三"的最高境界，但至少與人在一環境下共同"生活的智慧"能與日俱進。

根據心理學的社會學習理論(Social Learning Theory)，個人認知、環境和與他人的互動影響著個人的學習和行為，圖 1-2 中的 "Behaviour"代表(與他人互動時)他人行為，"Personal"為個人認知，"Environment"則就是環境；基於此，本計畫規劃成四階段、分四年執行，而分年主題分別為: 第一年研究對應為機器人與環境互動和人類與環境的互動之研究，第二年研究對應為機器人與人類互動之研究，第三年結合第一年與第二年的研究成果探究機器人、人類與環境三者的互動關係，第四年研究則是加入自主知識更新與拓展的功能。以下就每一年所將研究的內容及將採用的研究方法、步驟詳加描述。

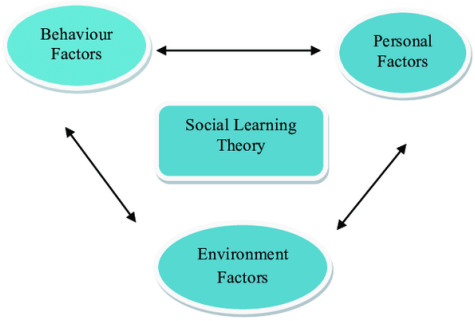


圖 1-2、社會學習理論(來源[48])

第一年

在此年度，計畫目標除了開發一個具深度認知能力的社交機器人系統之外，更重要的是研發出環境 3D 空間認知的模組架構以及在理解運動任務或需求後自主行為的形成模式。在機器人 3D 空間認知方面，過往的研究偏重在感知(sensing)層級，及利用一種或多種感測器來掃描環境以獲得精細的訊號圖，在轉成特徵圖之後據以建構環境空間的二維或三維的

幾何地圖，而演算法 SLAM 即是典型的代表，最後在這地圖中為完成運動任務而進行軌跡規劃或導航；但反觀人類在處於一個新的環境中，其能輕易地移動或前往某特定的目的地時，並非採用上述的方法或模式，而是由上而下的認知概念出發，進而形成運動策略，最後再解構成細部的軌跡或導航，換言之，機器人在無約束的複雜環境中能透過多模態(modality)感測器(如 LiDAR、RGB-D 相機、距離感測器、麥克風陣列等)利用深度感知原理建立一多層級概念圖[25-28]，使其能基於該地圖情形下即便在人們熙來攘往的場域仍能進行成功的社交導航而不致迷航，其次，機器人自主理解環境及周遭人群後可進行與人溝通或提供服務等(如圖 1-3 所示)。

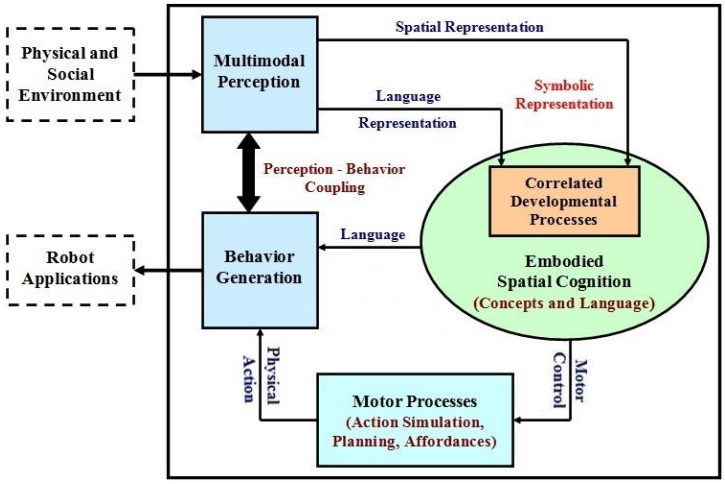


圖 1-3、機器人利用多模態感測理解環境、周遭人們與互動(來源[49])

1. 空間認知：構建用於室內環境的多層級概念圖

本計畫擬研發具認知能力之社交機器人系統搭載多元感測器，以滿足不同環境下自主推理與自主行為目標，亦期望展現符合社交準則、認知情境的行為以滿足人類複雜的需求。以往研究中，一旦機器人理解當前狀態，進而決策出所要執行的任務，大多先以理解整體環境為目標進行導航與建圖(SLAM)，即利用感測器資料建立環境特徵地圖、對照當下特徵以取得目前位置。然而，使用像素層級的資訊進行推斷十分消耗運算資源，在此計畫中，我們將實現一個多層級的空間認知架構，使用高層級的語義特徵以實現輕量化的移動運算。

多層級空間認知主要包含三大模組：多層級概念圖構建與定位、多目標行人追蹤、人群行為/移動模式認知以及社交導航。

- 多層級概念圖構建與定位

本計畫將基於語義分割/全景分割網路提供的語義信息，建立同時定位與語義地圖構建的基礎框架，並基於該框架提供的密集點雲語義分割地圖建立考慮環境理解的拓撲圖。我們將使用裝載於機器人的 RGB 相機及 LiDAR 作為主要感測器，前者可用於擷取機器人前方的清晰影像，後者則具有水平方向 360°、縱向正負 15°的視野，可獲取周遭環境之全向的三維點雲。

影像的語義分割/全景分割近年已成為電腦視覺領域中重要的研究課題，從初期的分類問題、演進到語義分割、實例分割的問題。一般而言，影像分類問題僅需將一整張照片做一次分類，而語義分割，卻主要是針對圖片上各個像素分別做分類，至於實例分割則是將圖片中的前景物體進行分類，並預測其對應的遮罩(mask)。近期出現了結合兩者的全景分割問題，目標是將每個像素上做分類，若其屬於前景物件，則需同時分類至對應物體之遮罩中，如圖

1-4所示範例。本計畫針對全景分割問題提出利用深度學習網路之有效解決方案，整體架構結合了由下而上分析法，以及由上而下分析法，不只在結果上可以有更佳的表現，在時間表現上也能符合實用需求。另外，針對傳統上遮罩在邊緣輪廓上較容易出現不平整的問題而降低分割的品質，因此本計畫亦擬將針對物件、背景之對應輪廓提出了加強特徵及損失函數，使系統可以對輪廓加強學習，從而提高遮罩的完整度。同時，在信心分數的估算上，亦擬結合遮罩品質做預測，使分數更符合在分割問題上的使用需求。

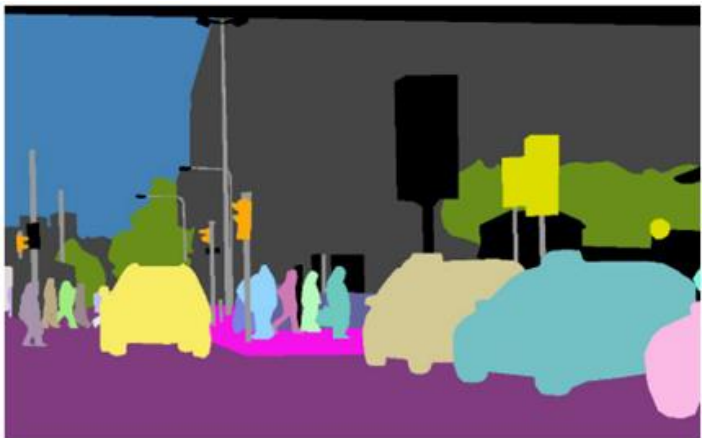


圖 1-4 、影像全景分割範例

在3D點雲的語意分割/全景分割方面，現有演算法大多使用 LOAM(LiDAR Odometry and Mapping)為基礎框架，以實現三維建圖與定位系統。LOAM 系統框架包含兩個主要演算法，其一為執行高頻率的里程計但獲得較低精度的運動估計之定位演算法，其二，以比定位演算法低一個數量級的頻率進行點雲匹配與註冊，即建圖和校正里程計之演算法，將這兩個演算法結合可獲得高精度、及時性之 LiDAR 里程計。當點雲清晰可靠時，上述方法效果較佳。近年來，基於 Range View 的方法進行 LiDAR 語義分割顯現出一定的優勢。在處理三維點雲資訊上相較之前的三維卷積的方法如 PointNet[50]，基於 Range View 的方法可以實現一般 LiDAR 高工作幀率(10-25 fps)的預測速度。而結合語義分割的 SLAM 方法除了可以建立語義地圖，同時還可以避免在複雜環境下動態物件對建圖與定位的影響。

考慮到效能、精度以及其他感知子系統的整合，本計畫將參考 SuMa++[51]建立同時定位與語義地圖構建的基礎框架，並在其已有的語義分割網路的基礎上，引入全景分割算法，通過對物件/人物個體的單獨分割，進一步提升語義地圖的認知層次。SuMa++的系統結構如圖 1-5 所示，其使用 Range Net++[52]作為語義分割網路，並使用當前幀(frame)之語義訊息進行地圖的校準，並過濾動態物件，其結構如圖 1-6 所示。

該網路使用修改過的 DarkNet-53 作為骨架網路(backbone net)，整體網路為 Encoder-decoder 結構，使用具有大幅度降採樣(down sampling)能力的卷積核(convolution kernel)編碼語義訊息。其輸入為 5 個通道的 Range Image，即 *x, y, z,* 距離以及反射率，並且只在橫向做降採樣。此外，骨架網路的前四層作為前景-背景分割特徵比對網路的輸入，使用 1x1 卷積與 3x3 卷積擷取特徵，如圖 1-7 所示。在使用中我們將分別輸入前一幀(frame)與當前幀(frame)的骨架網路特徵，以及前一幀的目標遮罩，並使用前一幀的目標遮罩區分前一幀的前景與背景，並與當前幀進行比較，以得到在當前幀區域目標的最大響應圖，進而進行反捲積上採樣(up sampling)，以獲得目標的當前幀分割遮罩。

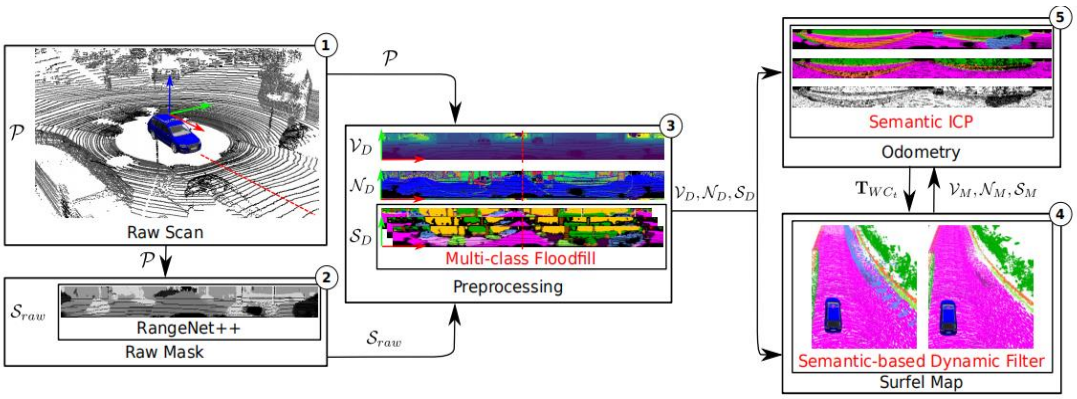


圖 1-5、SuMa++的系統結構(來源[51])

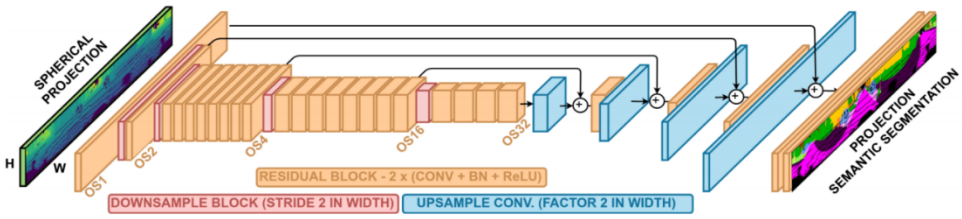


圖 1-6、3D 點雲語義分割網路(來源[52])

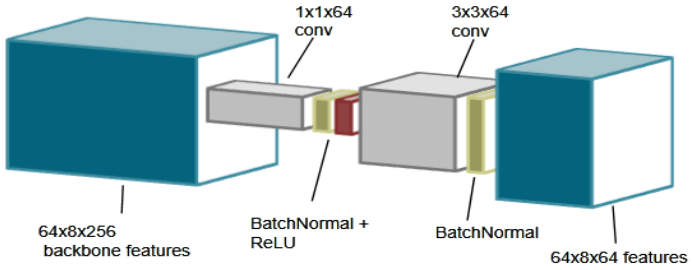


圖 1-7、前景-背景分割特徵比對網路

根據前述 2D 影像及 3D 點雲之語意分割(semantic segmentation)之後，根據前述提到的密集點雲語義分割信息，並結合 RGB 相機的物件偵測方法獲取的更詳細的物件訊息，了解環境中的物件分布狀態，進而將環境資訊壓縮成拓譜圖，並推論出相對應空間名稱，用符號簡化所記錄的 2D 拓譜資訊。相比直接使用密集點雲，增加的拓撲圖訊息可大幅減少當前空間定位複雜度，即只需調用當前拓撲圖節點相關的地圖;與此同時，拓撲圖訊息允許我們將環境感知與人物行為辨識之間的交互關係進行結合，形成更完整的空間認知，讓機器人系統可以在複雜的情境中，更加靈活的應用。

本計畫團隊目前在利用 LiDAR 進行同時定位與建構語義地圖已有初步成果，基於上述，本計畫將訓練語義分割/全景分割網路，以及開發行人分割追蹤網路，提供未來導航、人群移動模式開發的現實數據參考，故擬擴充實驗室已有點雲數據集 NTU-ACL-Semantic 數據集。此數據集將包含室內房間、室外以及室內長距離公共區域環境，並具有語義分割(Semantic segmentation)與行人的實例分割(Instance segmentation)，如圖 1-8 所示。該數據集將使用 Velodyne VLP16 LiDAR 於校園環境、醫院環境以及照護機構環境進行拍攝，以配合計畫所規劃的機器人使用場域。

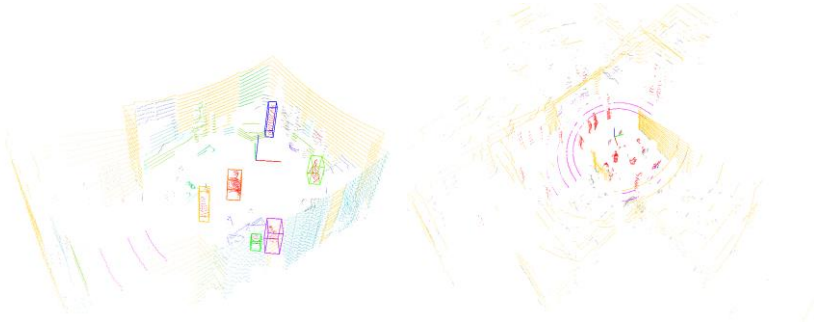
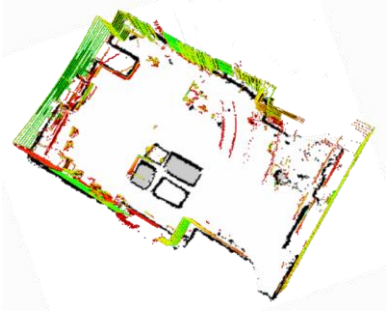


圖 1-8、 NTU-ACL-Semantic 語義/實例分割數據集
左圖為永齡生醫工程館 412 實驗室拍攝，右圖為臺大學生第一活動中心一樓拍攝

(a). 空間實景圖

(b). 3D 空間示意圖

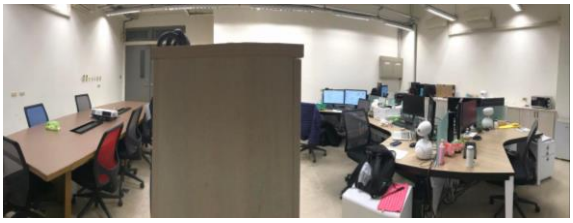


(c). 3D-LiDAR 語義分割

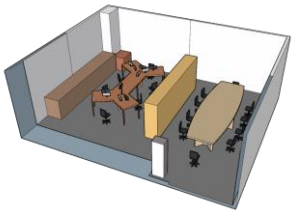
(d). 3D 符號拓譜圖

(e). 3D 空間示意符號拓譜圖

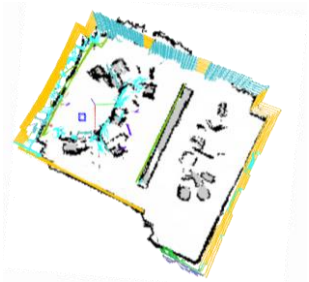
圖 1-9、生活區多層級空間認知概念圖



(a). 3D 空間實景圖

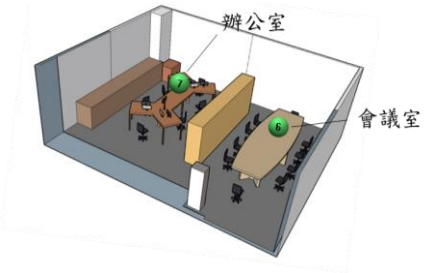


(b). 3D 空間示意圖



c) LiDAR 語義分割

d) 3D 符號拓譜圖



e) 3D 空間示意符號拓譜圖

圖 1-10、工作區多層級空間認知概念圖

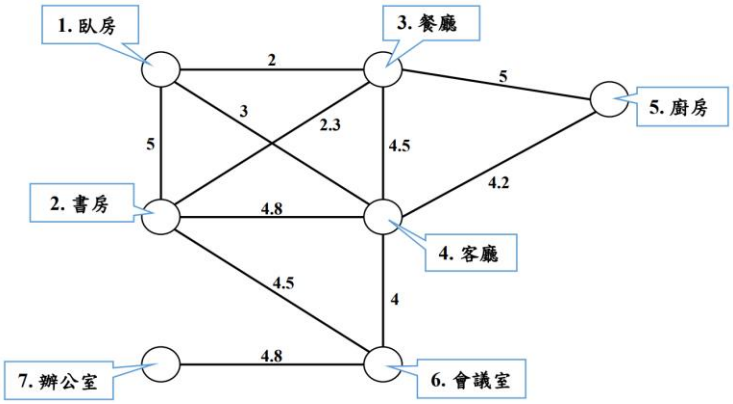


圖 1-11、全區 2D 平面符號拓譜圖

基於此系統提供的密集點雲語義地圖，並結合其他感知子系統(例如相機)的預測結果，我們便獲得可用以理解環境的多層級概念圖。詳言之，第一層的概念圖即會是場景的種類，例如家中的客廳、臥室、廚房等，第二層的概念圖即會是場景中 3D 空間的佈局與陳設、物件與人員的分布等，第三層的概念圖便是物件、人員以符號代表的拓譜圖(關聯圖)，第四層就是最細膩的空間幾何地圖，在機器人新處於一個陌生空間時，較難具備第四層級的概念圖，但基於前三層級之概念圖，一個具有深層認知的機器人應即能如同人類一般，被交付運動任務後或有行動需求時，即可在理解環境之後推理任務或需求與空間(環境)之關係，自主決定行為模式，循序或逐步達成任務或完成設定之目標。以下為幾個例子：例一、機器人在客廳某一角落，客廳內堆放了一些臨時置放的雜物或箱子，當家庭主人告知機器人須前往廚房幫忙，機器人應瞭解客廳與廚房連結之通道與客廳房門的位置(基於語意概念圖、符號拓譜圖)，自行導航至客廳房門、進入通道，緊接著前行至廚房門口，最終達成運動任務或行動目標；例二、機器人在審視一個大型機構的樓層平面圖後，被告知須從一間辦公室須前往另外一處會議室，但機器人初次來到這機構，在無引導之情形下，機器人透過多層級的空間概念圖(場景分類圖、語意概念圖、符號拓譜圖)即可自行推論在該樓層的所在位置、走出辦公室的房門、進入適當的走廊、經過幾個轉向之後進入會議室的大門，順利完成任務；例三、機器人在醫院依序巡視病房時，在中途被打斷行程、被告知須前往管理室協助，但機器人在審視原未完成任務的急迫性之後，一旦管理室工作完成即立刻以最短路程急速趕回執行原任務。

- **多目標分割追蹤(Human tracking)**

當機器人執行人機互動(Human-robot interaction)或人機協作(human-robot collaboration)任務時，機器人需要追蹤互動人物的動作、識別周圍環境，並預測人類操作者的行為，即具有一定的認知能力。而在執行這些任務的過程中，機器人移動所涉及到的社交導航系統不僅需要考慮固定的障礙物，還需要當前場景中的行人移動訊息，來達成有效率又同時尊重他人私有空間的移動方式。如前述，我們可經由語義地圖基礎框架得到當前的語義地圖，以獲得對當前環境的語義訊息。而對於當前環境中的人，我們計畫設計一種多目標追蹤系統，來對行人進行分割追蹤。

在之前的大部分社交導航(social navigation)的系統中，對行人的追蹤使用 RGB-D 相機，其視野範圍有限，當機器人執行人機協作任務時，無法同時兼顧周圍環境的行人探測與對目標人物更高級別的追蹤需求(如對目標人物進行動作預測)，這一點在可全向移動的機器人上尤為明顯；當機器人與目標人物相伴而行時，與目標人物的持續互動使得機器人的朝向與行

進方向不一致，從而導致行進方向上不在 RGB-D 相機的視野範圍，使得社交避障無法正常進行。

我們計畫設計的多目標追蹤系統將使用 3D LiDAR 的 360 度點雲訊息進行追蹤。相比於 RGB-D 相機，3D LiDAR 的 360°FoV 解決了 RGB-D 相機視野受限的問題，並且 3D LiDAR 探測距離可以涵蓋 1m~50m 的距離，這使得我們的系統不僅可以在室內環境中使用，在一些長走廊、機場、地鐵站等較大空間的環境以及室外環境都能夠正常工作。研究方法將參考 Range Net++[52]對點雲進行球面座標轉換生成 Range Image，並使用其生成語義分割遮罩，提取其中屬於人的遮罩(mask)作為當前場景的行人目標，整個過程如圖 1-12 所示。

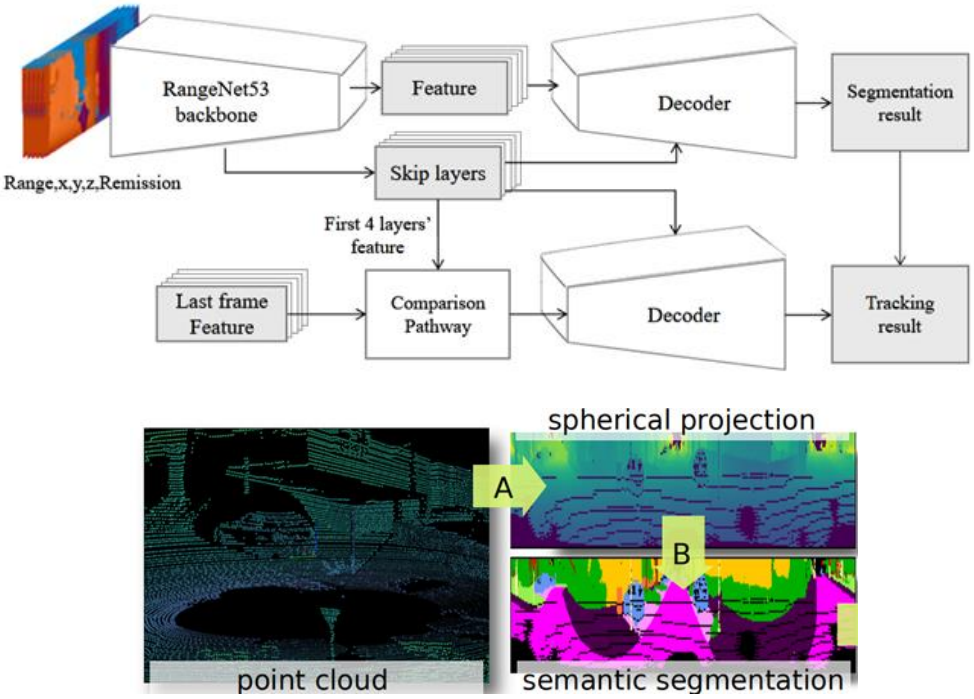


圖 1-12、利用 Range Image 進行人物追蹤網路

– 球面座標轉換：

球面座標轉換主要將 3D LiDAR 產生的點雲資訊根據其(X,Y,Z)空間座標，從 LiDAR 三維座標系投影到二維圖像座標系，產生一個類似深度圖的圖片，稱之為 Range Image，其中的每個像素包含了對應投影位置點的 x, y, z, 距離以及反射率資訊，其轉換公式如下：

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{2}\left[1 - \arctan(y,x)\,\pi^{-1}\right] & w \\ \left[1 - \left(\arcsin(z\,r^{-1}) + \mathrm{f_{up}}\right)\mathrm{f}^{-1}\right] & h \end{pmatrix}$$

其中(u,v)代表攝影機坐標系坐標，(x,y,z)代表 LiDAR 三維坐標系坐標，r 代表點到 LiDAR 原點的距離，$\mathrm{f_{up}}$ 與 $\mathrm{f_{down}}$ 代表 LiDAR 的垂直視野最大與最小角度，h 與 w 代表了 Range Image 的期望高度與寬度。Range Image 作為一種點雲的中間表示，其使得點雲可以使用二維卷積進行處理，在大幅減少計算量的同時可使用在圖像處理領域已成熟的多種二維卷積神經網路作為骨架網路，進行特徵擷取。此算法實際上可以理解為一種降採樣(down sampling)的方法：在轉換過程中會丟失大量的點，如掃描的 130,000 個點轉換到 64x512 的 Range Image 會只剩下 32,768 個點。因此，在轉換過程中建立點-像素點對應關係，可以避免 3D 空間中點的過度丟失。

- **人群行為/移動模式認知與社交導航**

“社交式導航”表現在機器人各式運動任務中，使其即便在人群的環境中活動也能讓人感受其類人的友善度、不致讓人有空間的壓迫感，整體而言，機器人能夠更友好地實現與使用者、行人的互動。其次，機器人能夠在對環境理解、行人需求辨識的基礎上，主動提供適切的服務。

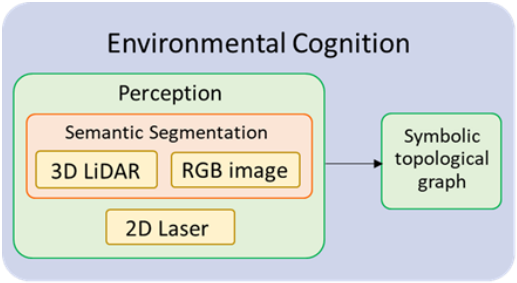


圖 1-13、環境空間認知架構

在本計畫中擬採用的方法，如圖 1-14所示，主要分為三個步驟：人物追蹤、軌跡預測及路徑規劃。在社交導航中，三維點雲可用於實施周遭行人及時偵測與追蹤，結合軌跡預測[53,54]之演算法，可避免機器人妨礙行人去路或與行人發生碰撞而造成他們的不悅。由於 LiDAR 縱向視角較小，僅有 30° 範圍，若僅採用 LiDAR 重建地圖，難以獲得近距離障礙物資訊，導致行走時避障的困難。因此，另在機器人較低位置安裝距離感測器(range sensor)，對一固定高度進行平面掃描，以檢測較低處的障礙物及行人雙腿。LiDAR 所獲取的三維點雲資料，可表示為：

$$P_i = \{x_i, y_i, z_i\}, i \in N$$

$$Point\ cloud = \{P_1, P_2, P_3 \ldots P_n\}$$

– 人物追蹤與軌跡預測

在感知模組的部分，擬嘗試使用以下兩種方式：第一種方式使用 RGB 相機以及距離感測器對人群進行感知與追蹤，此方式的優點在於誤判的可能性較低，但是可感知的視野有限；另一種則是採用 LiDAR，此方式雖擁有 360 度的視野範圍，但是由於解析度與影像相比較低，偵測上可能較為困難，因此需要對各種不同的人群追蹤模組進行測試，尋找最為穩定且有效的方式。而軌跡預測的部分，擬採用深度學習方法，將先前追蹤模組得到的歷史資料作為輸入，利用神經網路模型或貝氏概率進行未來軌跡的預測，此部分除了需要考慮預測的準確度以外，也需同時考慮機器人上的運算資源是否能夠及時完成預測結果。

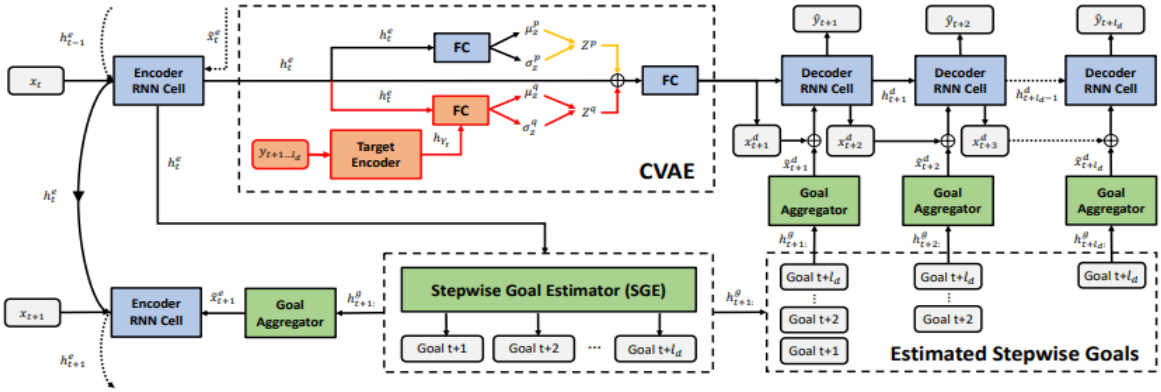


圖 1-14、行人軌跡預測(來源[57])

– 社交式導航演算法

通過相機和 LiDAR 資訊，機器人可取得周遭行人位置和動作資訊，本計畫擬將預測之行人軌跡、行人回饋納入路徑規劃考量，從而提高於多人環境中移動之社交接受度和社交友好度。在實施社交式導航時，根據當前位置利用下一個位置來計算其是否無碰撞，因此假設機器人在短時間內 Δt 弧形行走，並且機器人的當前位置為(i,j)，另外，θt 表示航向與水平線之間的角度。如果機器人以線速度v和角速度w移動，則可以將機器人的下一個位置導出為以下函數：

$$\hat{\imath} = i - \frac{v}{w}\sin(\theta_t) + \frac{v}{w}\sin(\theta_t + w \cdot \Delta t)$$

$$\hat{\jmath} = j - \frac{v}{w}\cos(\theta_t) - \frac{v}{w}\cos(\theta_t + w \cdot \Delta t)$$

$$\widehat{\theta_t} = \theta_t + w \cdot \Delta t$$

在上述函數中，$(\hat{\imath},\hat{\jmath})$ 和 $\hat{\theta}_t$ 表示當前位置和航向的座標通過計算出的下一個位置和方向，可用以驗證所選擇的速度和加速度是否會產生碰撞。

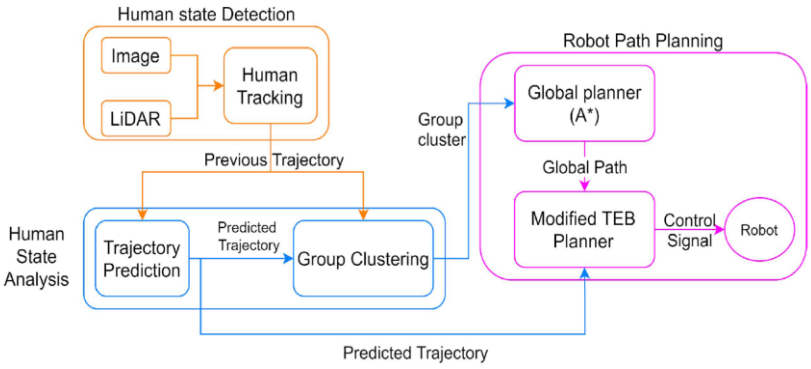


圖 1-15、人群中之機器人社交導航系統

– 人物活動辨識

認知環境其實亦包含認知某些特定人物的刻正進行的活動，因此本計畫亦將利用對周遭人物的偵測、追蹤之後，擬採用 RGB-D 相機辨識環境中當前特定人物的行為，以利社交機器人做出最佳互動行為或最合適之反應。

2. 社交機器人硬體平台

本計畫旨在提升機器人於空間及社交上之認知能力，使其能夠理解周遭環境並做出有意義的互動，以及達成與人類間更自然的互動及推理。惟現時市面上之硬體皆有諸多限制，例如對於系統之存取限制導致其難以做為開發平台使用，或感測器、平台之尺寸規格不符需求，且通常皆難以進行後續擴充等等。故本計畫擬提出一全向輪機器人作為合適的開發平台，以符合計畫之需求並提供良好的擴充性，使機器人能夠更容易實現與環境的互動及探索，並充分展現其認知能力。

認知之感知需求：對於環境認知來說，如何有效接收環境之語意訊息是相當重要的一環。其中，環境之 RGB 影像提供了豐富的基礎資訊，對於了解事物之「意涵」至關重要，而環境之幾何資訊則是在複雜場景中順利移動、追蹤物體的重要關鍵。如較複雜之生命體

(哺乳類)，即是透過雙眼(雙目攝影機)，產生 RGB 影像及深度視圖，透過場景之顏色、幾何認知物體，最後整合高級訊息達成對環境的認知。故在本計畫中，我們將會使用 RGBD Camera 接收環境之 RGB 影像及深度圖，用來對環境進行較為細緻的認知。並配合 Lidar 接收四周粗略的空間訊息，這些資訊將用於認知機器人周圍的空間狀態，使機器人在複雜環境順利移動、認知並追蹤不同之目標等等。

社交互動需求：為展現本計畫提出之社交認知能力，機器人不僅具備認知人類情緒等資訊的能力，也能做出有意義的回應。相較以往僅有最基本的語音回覆或簡單表情變換，本計畫擬提出之系統具備更自然且複雜的表情及肢體語言，使人機間的交流不再是單向的，而能提供具有「回饋感」的互動。因此，本計畫擬提出之平台將具備顯示螢幕、揚聲器及手臂，作為機器人表達表情及肢體語言之媒介，展現其於社交方面的認知能力。

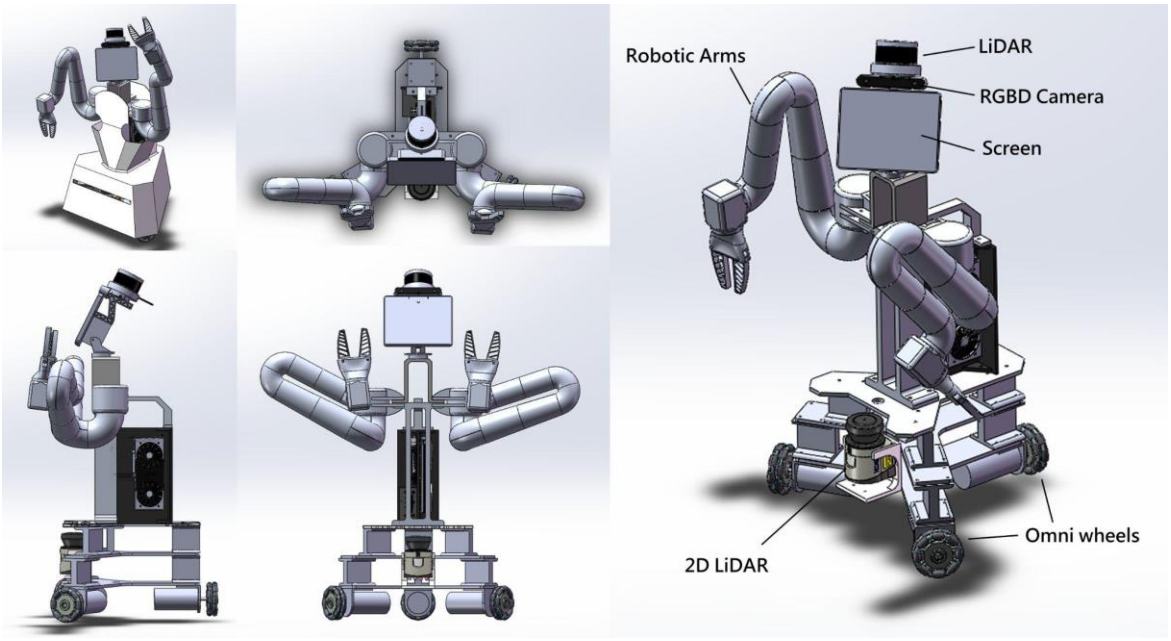


圖 1-16、機器人之初步設計圖

因此，本計畫擬提出之平台將具備(a)一全向輪底盤，使機器人能在較狹窄的室內環境下靈活移動；(b)可動之頭部，含一觸碰顯示螢幕，作為機器人表情顯示或更複雜的控制 UI 介面。及兩隻手臂及夾爪，使機器人能夠與環境互動並賦予機器人肢體語言；(c)各式感測器，提供達成對環境之深度認知所需之基本資訊(如前方之 RGB-D 影像、立體資訊(點雲)、聲音、觸覺等)。整體如圖 1-16 所示，各項細節如下:

- **移動平台方案**

底盤將採用具有三輪全向輪的移動平台，其機動性可達三個自由度，即速度控制上不受約束，可達成任意方向之平移及旋轉，使機器人的移動更加靈活。相較於採用麥克納姆輪之移動底盤須放置四輪及四顆馬達，本方案能節省空間使用並使尺寸更加精簡，以利機器人在較狹窄的室內環境中順利移動。如圖 1-17 所示，全向輪移動平台的輸入為[ω1 ω2 ω3]，分別為三輪的角速度。可輸出平台之[v_x v_y ω]，即切線速度、平移速度、角速度，且三者相互獨立，達成高自由度之平面移動。

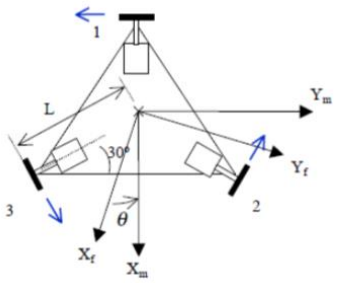


圖 1-17、全向輪底盤示意圖

本計畫並擬於底盤上配備高精度 IMU，配合馬達 Encoder 計算之 Odometry，以及前述之 semantic SLAM，最後可透過 fusion 提高定位的準確度及可靠性。於人潮較壅擠的場所，部分方法可能失效的情況下，仍可透過 IMU 及 Odometry 維持一定的定位能力。

- **頭部及手臂方案**

頸部具有兩個旋轉關節，分別控制頭部的 pitch 及 yaw 角，使頭部的感測器(深度相機、LiDAR、麥克風陣列)能夠更靈活的讀取環境資訊。並使顯示螢幕(即機器人面部)能夠面向使用者，或做出肢體語言(注視、搖頭、點頭等等)。為使機器人能夠與環境進行互動，並擁有與人類相仿的肢體語言，本計畫擬於機器人兩側各設計一具 7 個自由度的手臂，其中多餘的一個自由度(Redundancy)使其能夠更靈活的動作，並以更複雜之路徑規劃繞過障礙物與物體互動。兩隻手臂的末端各裝備具一個自由度的撓性夾爪，足以讓機器人抓取大部分的物品，且不致人類受傷。

- **感知方案**

本計畫提出之平台於頭部配有一深度攝影機及一 360 度 LiDAR，而如前段(1-b)所述，頭部具兩個自由度的旋轉，使兩者的偵測範圍更加自由，減少 FoV 不足產生視覺死角的狀況。深度攝影機之 RGB 資訊輔以其深度影像，提供機器人更多前方資訊，認知需高度關注之事物。而頭頂之 LiDAR 則能產生大範圍點雲，除建構 3D 地圖外，亦能使機器人認知四周空間之概況，透過系統之認知能力，追蹤目標(人類)、軌跡預測、分析障礙物，及閃躲接近之物體，達成於室內軌跡自然之移動且不致人類不適的室內導航能力。而底部之 2D LiDAR 則能進一步輔助機器人之視野死角問題，提供更強健的導航能力。

為達成人機間的自然語言溝通，本計畫擬於頭部亦配置揚聲器及麥克風陣列，配合 TTS(Text to Speech)及 STT(Speech to Text)技術，將人類語言轉換成文字，並透過整合認知系統之 NLP (Natural Language Processing)產生回應的內容，通過麥克風說出，達成人機在語言上的主動及被動的交流。而麥克風陣列亦可達成聲源的辨識，使機器人能以聲音呼喚。更甚將聲音與影像資訊進一步關聯，擴張其感知能力。

而對於人類來說，身體的觸碰(如肩或手等)亦是一常見且表示親密友好的交流方式，為使人類能夠透過觸摸來呼叫或感測碰撞，機器人將在外殼支撐處設有一組感測器，檢測觸碰來源，提供人機間不同的溝通及交流方式。

第二年

構建人物印象：機器人要想成為貼心的陪伴者、照護者，需要能通過長期的互動了解使用者的記憶線索、認知水平與偏好資料，於日常生活中承擔遠端醫療的輔助角色，提供病人期盼的資訊或諮詢，或主動發掘可能的危急狀況；另一方面，透過與醫療照護機構的合作，將部分專業知識轉化為機器人可讀取的資料，並在未來承擔繁雜流程中多步驟的輔助任務，改善未來社會高齡化、少子化、專業人才不足且培育成本高的問題。在第二年中，我們將從人類行為認知的角度，實現對人類意圖、行為模式以及情緒狀態的辨識，設計符合人類社會準測或人類預期的行為或提供具有同理心的對談或共情行為。

1. 人類意圖與模式辨識

目前市面上的協作型機器人多半是被動式驅動去輔助人類完成任務，像是藉由接收語音去紀錄提醒事項，或藉由手動選擇讓機器人去執行想要的任務，但在認知機器人中，我們希望跳脫這個框架使機器人能夠自我偵測任務並自我完成或輔助人類完成該任務。在此之前機器人必須有理解環境的能力，並能夠從環境中的特徵來判別特殊的場景如醫院或車站等，而從環境的認知從知識庫中(knowledge base)篩選與環境密切相關的任務。在有人類參與的任務中，機器人必須有能力辨識人類當下所進行的活動，並從當下的行為預測未來的行為，在根據每一次偵測的行為更新知識庫中該行為與目標的關聯性，這項技術包含了活動辨識、目標辨識、計畫辨識[55,56]的建立與知識庫的存取。

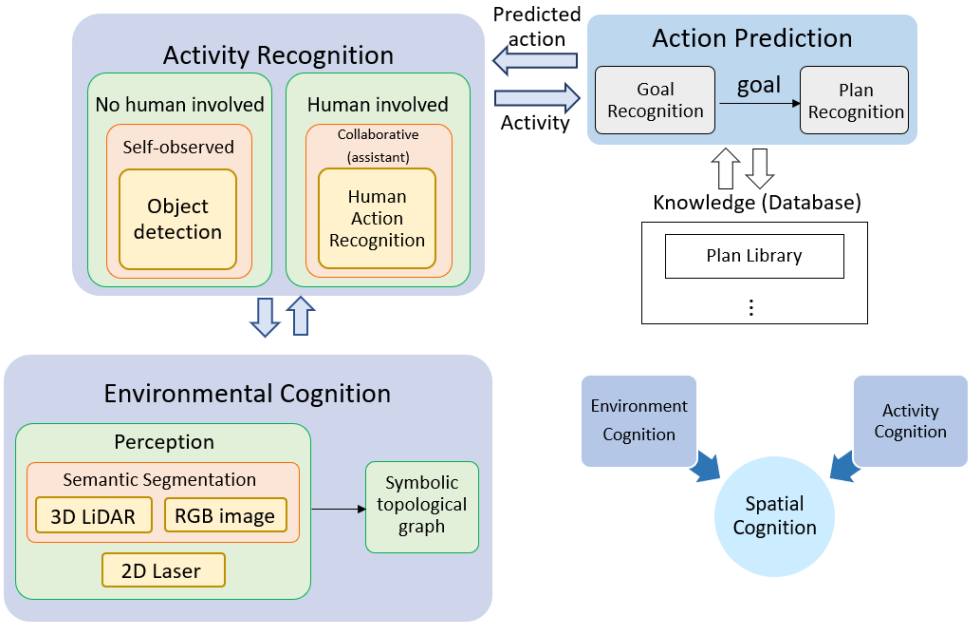
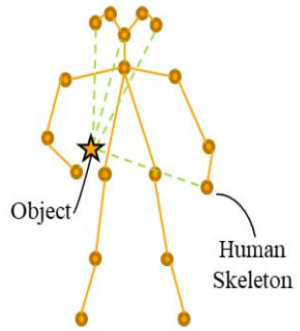


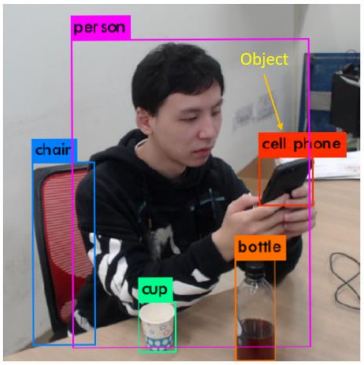
圖 2-1、環境結合人類活動感知之空間認知架構

1-a)　活動辨識(Activity Recognition)

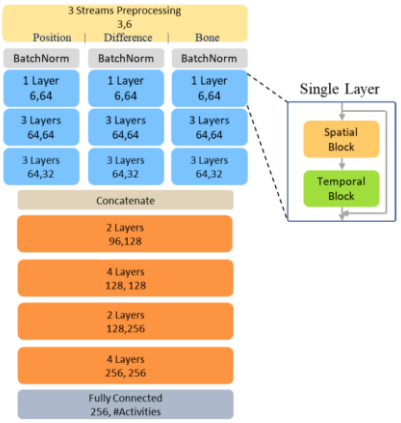
人類活動辨識包含了人類行為辨識(human action recognition)，藉由偵測人當下的行為來判斷人在進行什麼樣的活動，我們藉由 RGB-D 相機與 Openpose 做人體骨架(human joints)的偵測和用 YOLO 對周圍物體(nearby object)的偵測當作輸入，並基於這兩項資訊來判斷人類當下的進行的活動。



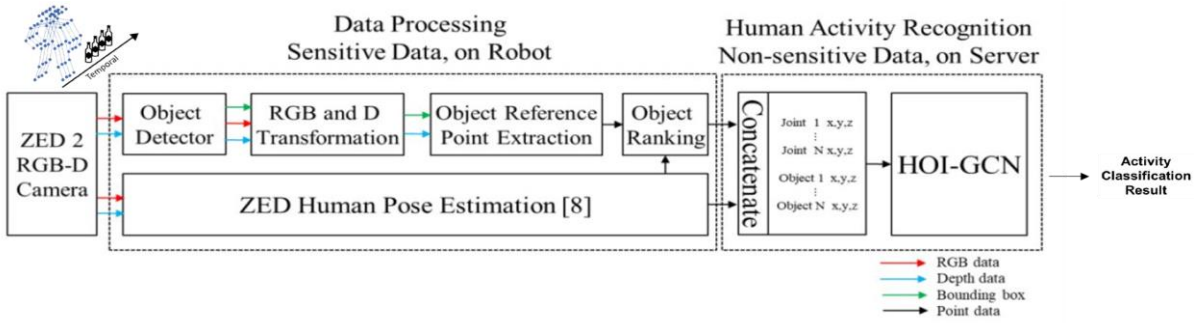
(a). 人體骨架與物體之關係



(b). YOLO 物體辨識



(c). 人物互動之圖卷積神經網路



(d). 系統流程圖

圖 2-2、人體骨架與物體資訊偵測之流程圖

人體骨架適合用在 GCN 模組上做空間上的學習，GCN 的優點為分析與計算的參數較為少，同時結合物體的資訊能達到較高的準確度與判斷行為的即時性，而在時間上的學習 LSTM 的運用能夠獲得序列中不同幀的重要性。反之在無人類參與的情況下，機器人要有能力在環境中自我找尋任務去完成，透過物體的偵測與環境的分析，在搭配知識庫的資訊來篩選適當任務去執行，而辨識後的 activity 將送入至目標辨識模組或計畫辨識模組進行更深層的分析。

1-b) 目標辨識(Goal Recognition)

目標辨識為計劃辨識的一部分，根據所辨識出一系列的人類行為並透過知識庫裡所建立的計畫樹狀圖(plan library)[56]來判斷最終的目標。計畫樹狀圖的結構是由目標、次目標、動作、生產規則等所組成，含有在任務下各個所需達成的目標與先後順序，如圖 2-3 所示，並透過在有效假設的任務上用貝氏推理(Bayesian reasoning)[57]找出最高機率的目標，最後將此目標與當下辨識到的行為作為計畫辨識模組的輸入。

1-c) 計畫辨識(Plan Recognition)

計畫辨識是基於所得到的人類行為與目標資訊進行未來行為的預測[58]，並根據此預測給予適當的協助，如圖 2-4 所示，在醫院裡當機器人偵測到醫生在為病患處理傷口時，辨識到醫生在消毒傷口之後要能認知下一步是為傷口上藥，並且及時的提供相關藥品或上藥的協助，這也牽涉到機器人需要了解藥品在環境中的位置並移動至此，我們透過機器人初步對環境與物體位置在知識庫上的建立來攝取該資訊，並使用既有的導航系統與必障功能在環境中進行移動。

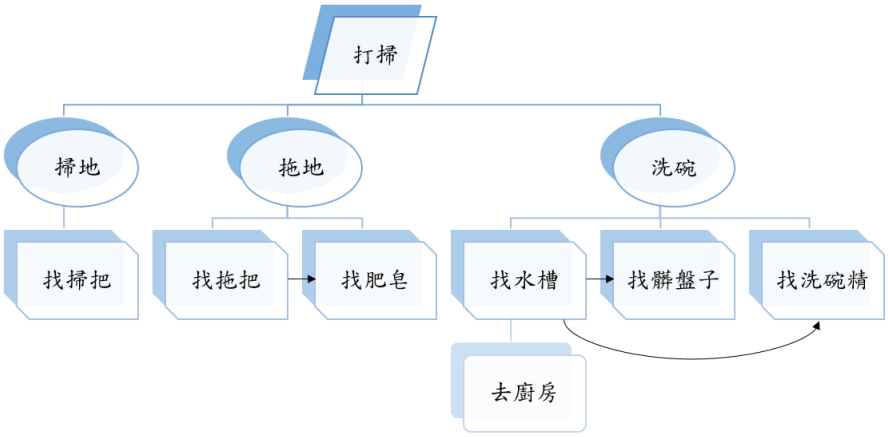


圖 2-3、計畫樹狀圖，例：室內清潔活動

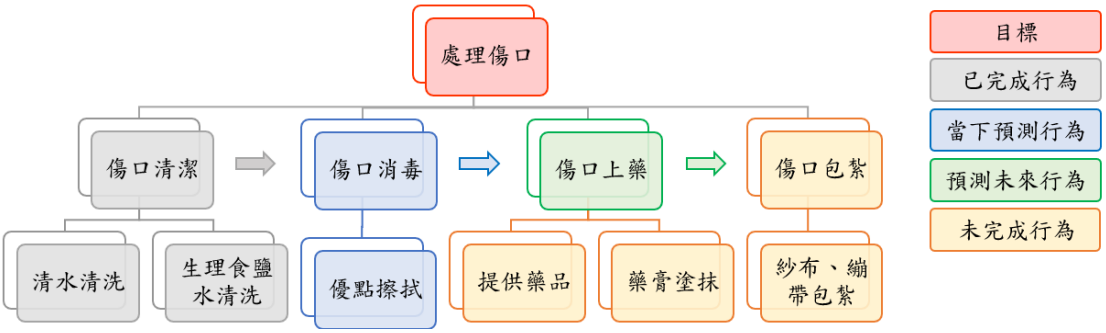


圖 2-4、行為偵測與未來行為預測、執行之計畫樹狀圖

2. 人類情緒辨識與安撫

現有之人機互動(Human-Robot Interaction)大多利用影像辨識及語意分析等方法，先透過相機辨識人類表情是為開心、生氣、難過等情緒[59]，同時機器人與人聊天互動蒐集語意，並且分析語句，最終呈現的表情也是只有固定幾種[60-62]。

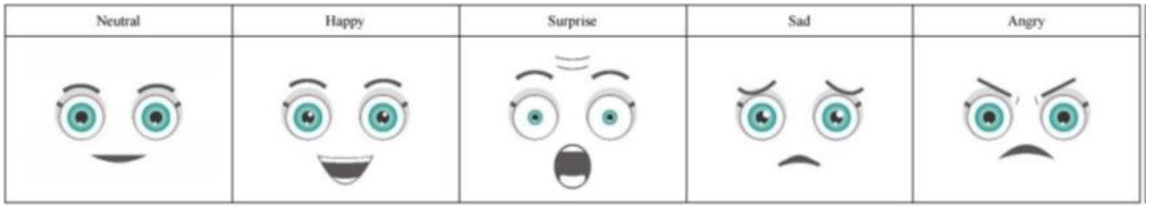


圖 2-5、目前多數機器人呈現的表情較為固定單一(來源[60])

然而，為使機器人在社會情感維度上更貼近人類的需求、正確地辨識人類的情緒、並能夠遵守和適應共同商定的社會規範以建立信任關係，賦予其對人類情緒、情感的認知能力，遵從心理學討論的社交行為及同理心策略，實現生成具有同理心對談內容之演算法。人與人的互動(Human-Human Interaction)中其實會展現豐富的表情，具統計人類至少擁有超過 21 種表情(包含憤怒、專心、迷亂、蔑視、期待、厭惡、興奮、恐懼、快樂、悲痛、驚訝、怒

視、咆哮、失敗等)，在人與人交流時，同理心被描述為準確識別他人的情緒並表現出來基於情況的適當行為，但人善於掩飾內心想法和情緒，且人的表情與肢體語言可能代表很複雜的情緒(不是單純的開心或傷心，而是由許多情緒所組合而成)，因此偵測微表情及分析人真正的話語意思與肢體動作背後的真正情緒涵義是本計畫想研究的方向，最終機器人將學習如何判斷面部的呈現方式(眉毛上揚的角度、眼睛大小、嘴的形狀等)及肢體的表達方式(手臂各軸的轉動角度等)，而這也是本計畫中對於同理心行為設計的目標。

2-a) 情緒辨識-語義情緒辨識

本計畫擬利用自然語言處理技術，透過使用者說話的內容來辨識其情緒。過去傳統的研究多是使用預先定義好的辭典，搭配關鍵字偵測來辨識整句話的情緒內容。此種方法往往會受限於辭典的大小以及文字的表現方式，因此在實務上的表現不盡理想。近年來，隨著深度學習的發展，透過對文字的分析來辨識情緒的準確率已經越來越高。BERT[63]藉著對含有大量參數的模型在大量的文本上進行預訓練的方式，讓模型在某種程度上像是讀過了這些文本一樣，儲存了這些文本隱含的知識，之後只需要將模型在相對少量的資料集上進行微調之後即可應用在包含情緒辨識在內的各種下游任務上。

在模型的微調訓練過程中，一個句子中的每個字會先被轉換成向量，這些向量會和模型中每一層的參數進行運算，最終我們按照 BERT 模型的預訓練過程，取最後一層的第一個向量，將該向量通過一個線性層和歸一化指數函數，便能取得輸入句子在每個情緒類別上的機率分布，機率最大的類別即為該句子的情緒類別。

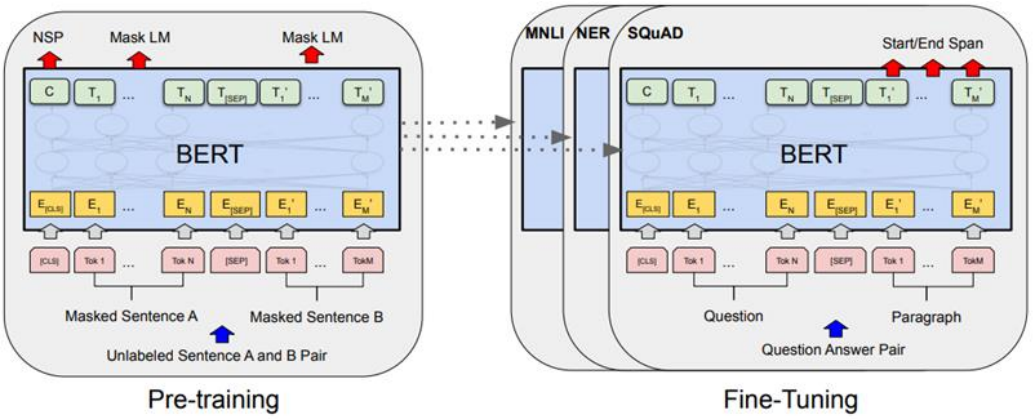


圖 2-6、BERT 模型的預訓練、微調過程示意圖(來源[63])

我們將使用 GoEmotion[64]資料集來對模型進行微調，該資料集在 Reddit 論壇上蒐集了五萬八千段文字，每段文字的長度約為 12，這些文字被人工分類成開心、難過、憤怒等 27 種不同的情緒。我們將文字作為模型的輸入，文字對應的情緒類別作為資料的標籤來對模型進行微調訓練。訓練完畢後將模型儲存起來，之後在實務應用時，將使用者所說的話輸入到模型進行預測，便可達到即時的情緒辨識。

情緒辨識的結果將會與本計畫中的其他技術，如同理心等做結合應用，讓機器人在與使用者互動時能夠即時的認知到使用者的情緒，並進一步利用該資訊來和使用者互動，以期增進機器人在社交互動上的認知能力。

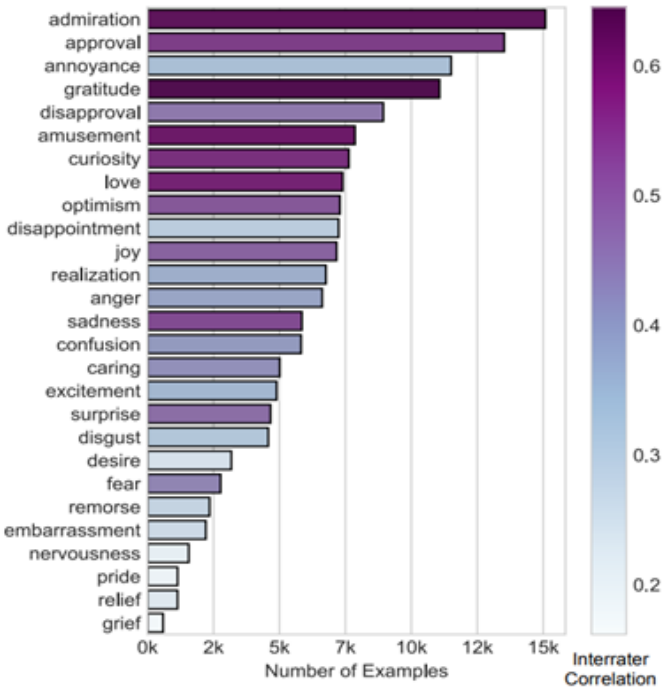


圖 2-7、GoEmotion 資料分布狀況(來源[64])

2-b)情緒辨識-影像情緒辨識

判斷人類情緒有許多方法，其中，人類之微表情是一種人類在試圖隱藏某種情感時無意識做出的、短暫的面部表情，通常對應七種情感：厭惡、憤怒、恐懼、悲傷、快樂、驚訝和輕蔑，而二十世紀最傑出的100位心理學家之一的保羅·艾克曼聲稱，任何人經過訓練都可以更容易地識別微表情;判斷手勢及人體骨架得知人類情緒[65]在目前已有少數研究論文發表；而本計畫最特別之處為加入了腳尖朝向資訊，有研究指出，腳或腳尖的動作容易表現出人的真心，不容易欺騙他人(例如:對話時，膝蓋或腳尖若是對著對方，表示對對方有好感，或對談話興味盎然;相反地，膝蓋或腳尖若是偏離對方，則表示拒絕對方，或對談話興致缺缺)。故本計畫預計使機器人能透過微表情、手勢、肢體動作、腳姿，判斷人類真實情緒。

故本計畫將利用深度影像辨識與音頻語意分析人類情緒，透過相機分析人類微表情，結合手勢行為特徵的辨識分析，並分析人類腳尖朝向隱藏之涵義，再透過骨架擷取模型得知人類行為並分析人類真實情緒，經由這四個特徵輸入，透過深度神經網路，最終輸出能代表肢體語言的 feature vector。

3. 展現同理心、情緒之設計

3-a)機器人同理心表情與動作設計

目前現階段的機器人產業中，完全沒有表情或表情較為固定單一的機器人居多，見圖 2-8，本計畫為了打造具有深度認知能力之社交型機器人，故機器人與人互動中，機器人回應的表情及肢體動作也尤為重要。

機器人能藉由學習「學習對象」面對「目標對象」所產生之反應的表情與肢體動作及語言，以產生相似的表情與行為。故機器人最終將學習結合影像及語意辨識產生之結

果(目標對象之真實情緒)，回饋相應的表情與手勢(非固定表情或動作)，改善機器人面部表情及手勢固定單一之狀況。

圖 2-8、完全無表情手勢機器人、表情及肢體單一固定機器人

本計畫對於機器人與人互動中應產生的表情與手勢預計研究之方法如下:

(1)影像輸入目標對象的面部微表情、手勢、人體骨架及腳尖朝向訊息(2-b 提及之方法)，透過複雜神經網路分析人的豐富情緒，最終輸出能代表肢體語言的 feature vector

(2)輸入聲音判別目標對象之語意(2-a 提及之方法)，最終輸出能代表語意情緒資訊的 feature vector

(3)結合(1)、(2)之輸出結果[66]，產生一個包含人類情緒資訊的 semantic vector，再經過 fully convolution 及 loss function，最終輸出人類的情緒占比

(4)此時與「目標對象」交流的「學習對象」之表情與肢體行為將作為標籤(label)，用以訓練深度神經網路模型之輸出

(5)結合(3)之結果(人類情緒占比)作為輸入，經由(4)提及之深度神經網路模型，機器人將學習到情緒與對應之表情與肢體語言的關係，最終機器人將會將會擁有類似人類的肢體語言及反應。與如今機器人最大的不同是，本計畫結合了微表情、手勢、腳及骨架分析，同時包含語意、語氣及語速分析，且回應表情並非固定，而是如人類擁有同理心般的豐富表情與手勢。

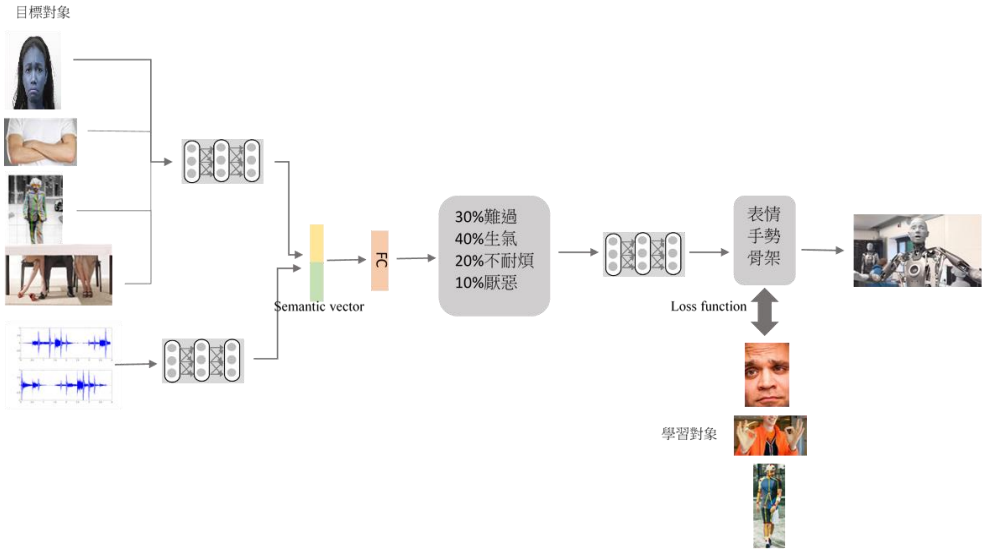


圖 2-9、本計畫預計研究之方法(參考[66])

3-b)機器人同理心對談

同理心指的是能夠理解、感受別人經歷的能力[67]。機器人在與使用者互動的過程中，無論是提供幫助、進行對話，同理心都是一項相當重要的能力。同理心具體而言又可分為兩種，分別是認知同理心和情感同理心[68]。認知同理心是指能夠站在別人的角度、理解別人正在經歷的狀況；情感同理心則是在情感上能夠理解別人、且會對別人的情緒產生對應的共鳴。

為了使機器人能夠具有同理心，本計畫將分別從情感、認知兩點著手。情感部分，使用本計畫前面提到的情緒辨識的技術，透過與使用者進行言語上的互動，能夠即時掌握使用者目前的情緒狀態，進而採取適當的對應動作。而認知同理心的部分，本計畫將使用深度學習技術中的預訓練模型 [63]，透過在下游任務資料集[69]的微調，達到理解使用者意圖的效果，此資料集包含了兩萬五千筆對話資料，並且每筆資料都經過手動標示成提問、同意、建議等五百種不同的意圖。

在實際的應用上，我們可以直接使用模型產生的結果，也就是在做決策時考慮使用者目前的情感、意圖狀態。除此之外，由於模型在產生結果前的運算資料是以向量的形式被儲存起來的，在其他深度學習的模型中，若是使用的維度相同，也可以使用這些向量資料來做運算，例如在用生成式語言模型和使用者進行對話時，可以在輸入的向量中加上對應維度的值[70]，如此便能夠在模型運算階段就直接運用了同理心的資訊。

對機器人而言，能夠得知使用者的情緒和意圖資訊，就如同擁有了人類的情感和認知同理心。在與使用者進行互動時，只要充分利用這些資訊，就能表現出同理心的效果。以對話為例，在使用自然語言生成技術來產生對話內容時，除了使用者輸入內容的語意以外，若能夠另外考慮該內容的情緒、意圖，研究結果[70]表明無論是自動評估或是人為評估的實驗結果都明顯優於單純考慮對話內容的語意。

4. 人類自傳記憶(Autobiography)模型構建

在認知模型理論，人類的情緒與行為意圖會同時受到過往的回憶、當下環境和客觀事實所影響，這種個人化記憶被稱為自傳式記憶(Autobiographical Memory)[71]，包含情節記憶(Episodic Memory)和語意記憶(Semantic Memory)。情節記憶由重要事件所構成，事件可以包含使用者行為、參與人物、發生地點等組成，而語意記憶則是和客觀事實有關的記憶，如家人、出生地、情人和興趣等等。建構具有深度認知 AI 社交機器人的挑戰，除了能理解外在環境和具備常識外，也必須如人類大腦儲存使用者過往記憶與客觀事實進行分析，以進行長期的情緒和行為意圖分析，並提供長者和記憶功能障礙的使用者進行記憶輔助(Memory Assistance)。

根據上述挑戰，過去研究會使用資料庫或基於心理學的計算記憶模型，來建立具備自傳式記憶模型的機器人系統，比如[72]使用資料庫的方式建立自傳式記憶模型，將使用者的語音轉換成文字後，透過自然語言處理的技術來偵測關鍵字和動詞儲存進資料庫，將記憶模型分成主題、時期、事件和情節記憶四層，由眾多的情節記憶來組成事件，事件可以由時期做劃分，最後將不同時期的事件用主題來做分類。然而，[72]的系統在將文字轉化成記憶時，僅使用關鍵字和既定語法規則來分析語句，無法自動進行多種事件抽取、抽取客觀事實和主題分類。而[73]和[74]的機器人系統，透過使用者生活影片為輸入，使用動作分析(activity

recognition)和物件偵測(object detection)來分析使用者行為，並分別透過圖(graph)和資料庫建立記憶模型。受限於動作分析只能偵測使用者基於動作相關的事件，這些記憶系統能分析的記憶非常有限。

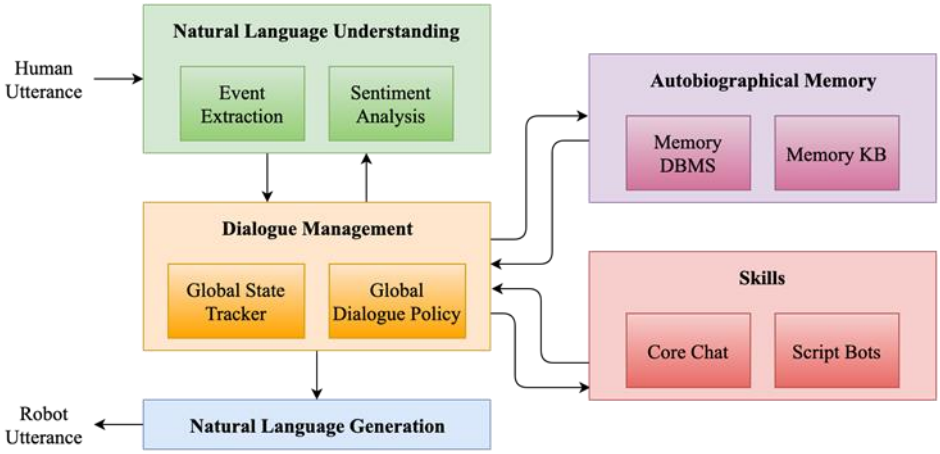


圖 2-10、基於資料庫記憶模型之機器人系統(來源[72])

近年來，自然語言理解如命名實體識別(Named Entity Recognition)、機器閱讀理解(Machine Reading Comprehension)和關係抽取(Relation Extraction)等任務，因為Transformer[75]和 BERT[76]深度學習模型的出現有重大的突破和應用，本計畫擬基於使用者語音辨識後的文字資訊，利用深度學習模型來將分成多步驟將文字資訊轉化成情節記憶和語意記憶，並透過整合兩種記憶來實現自傳式記憶模型，以提供記憶輔助的功能。

步驟一: 情節記憶

對於情節記憶，本計畫會將文字訊息表示成標示符(token)組成的序列，透過三種深度學習模型，分別是語意角色標註(Semantic Role Labeling)、事件抽取(Event Extraction)和機器閱讀理解(Machine Reading Comprehension)來解析語句。語意角色標註能分析語句的動詞和其餘單詞的關係，適合短文本的記憶分析，事件抽取分析語句中存在的記憶事件種類，機器閱讀理解能透過問答的方式抽取出適當文字片段作為記憶關鍵字。目前對於事件抽取[77]和機器閱讀理解的公開資料集[78]較少關於使用者日常行為，且多為英文資料集，本計畫擬使用中文公開資料集對模型進行訓練，再建立中文日常行為相關資料集，利用遷移學習(Transfer Learning)方式來訓練模型。

步驟二: 語意記憶

對於語意記憶的客觀個人知識，本計畫擬採用命名實體識別和關係抽取模型，首先將語句中的人物和地點抽取出來，並透過關係抽取模型來判斷人物與人物或人物與地點的關係，來作為個人的語意記憶。如圖 2-11[79]所示，根據輸入的語句和提前偵測出的人物 Obama 和地點 Honolulu，[S-PER]代表 Obama 為人物且為主體(subject)，[O-LOC]代表 Honolulu 為地點且為客體(object)，主體和客體之間的連結性，可透過 BERT 作為特徵擷取層，和序列分析常用的神經網路模型 BiLSTM 來預測主體和客體的關係(Relation)，預測結果為出生地(per:city_of_birth)，代表主體為 PER(人物)，和客體的關係為出生地(city_of_birth)。而對於非人物相關的語意記憶如個人興趣，目前沒有中文相關的公開資料集，因此本計畫會透過在情

節記憶部分所訓練的機器閱讀理解模型，先建立中文客觀知識問答的機器閱讀理解模型資料集將模型重新訓練，透過設計好的自定義的問題來抽取出客觀知識。

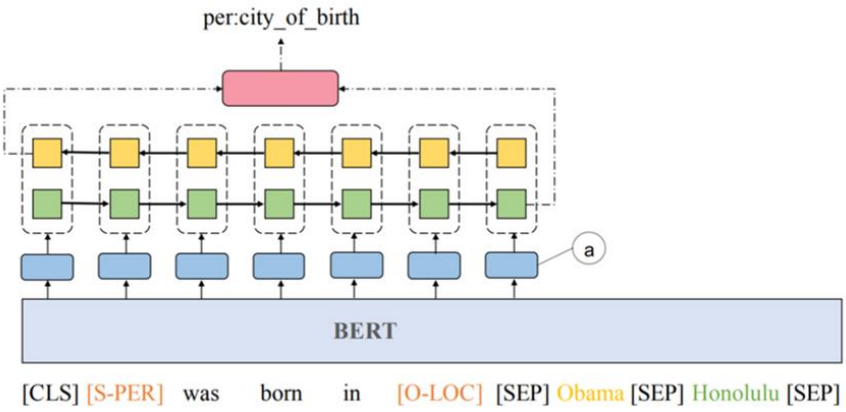


圖 2-11、基於 BERT 模型與實體(Entity)之關係抽取模型(來源[79])

步驟三: 自傳式記憶模型與記憶輔助

基於[72]的記憶模型設計，我們欲設計兩層的記憶模型，分別為主題知識層和記憶事件層，主題知識層主要為語意記憶的知識所構成，而記憶事件層包含情節記憶和語意記憶中的命名實體、情節記憶的時間、情節記憶的使用者行為等。使用者進行記憶輔助時，透過將使用者的問題進行主題分類，配合主題知識層映射到相應主題的記憶群，接著透過檢索模型BM25[80]、ColBERT[81]、DPR[82]或 SBERT[83]找出最相近的記憶作為答案，而若檢索模型找到的記憶相似度低於設定的閾值，則會使用機器閱讀理解模型，以使用者的問題和記憶的文本資訊作為輸入，透過模型輸出來找出適當的片段作為答案。

第三年

在此年度，我們將賦予因果認知能力於機器人自身：先前我們先是讓機器人獲得感知的能力，讓他具有與人類視覺一般能在接收環境資訊之後，進而產生空間認知，如此一來就能在環境與人的互動，抑或是環境與自身的互動上處理相關資訊，也就是空間認知和社交認知兩者，於是機器人就能長期觀察使用者的偏好資料，承擔未來作為陪伴者或照護者的工作一職。

1. 因果推理與預測

基於第一、二年的研究，資訊包含著非結構性資料和結構性資料，比如感測器的原始資料如語音和影像，神經網路模型預測的人類行為、情緒和意圖等，以及自傳式記憶模型儲存的情節記憶與語意記憶等。透過建立長時間搜集下來的資訊之間存在對應關係，如此一來機器人就能夠在這高維度的空間裡歸納出屬於他自己的因果認知，透過習得的因果認知能力機器人就能自己思考自身的行為對使用者的效益之有無，亦或者能夠推斷自身能否在使用者需要時主動提出協助的預測能力。

Positive		Negative		Ambiguous
admiration 👏	joy 😃	anger 😡	grief 😢	confusion 😕
amusement 😂	love ❤️	annoyance 😒	nervousness 😬	curiosity 🤔
approval 👍	optimism 🤞	disappointment	remorse 😔	realization 💡
caring 🤗	pride 😌	disapproval 👎	sadness 😞	surprise 😲
desire 😍	relief 😅	disgust 🤮		
excitement 🤩		embarrassment 😳		
gratitude 🙏		fear 😨		

圖 3-1、情緒辨識種類(來源[84])

在過去研究中，對於非結構性資料和結構性資料，可以透過符號化表示法建立機率模型，透過建立貝氏網路(Bayesian network) [85]來表示資料的因果性，然而對於感測器原始資料影像和語音，無法直接使用符號表示法來建立因果性。相對的，機器學習和深度學習的方法，則是假設資料為獨立同分布(Independent and identically distributed)，透過端對端方式直接學習資料與下游任務的關聯。然而在[86]的研究指出，對於和訓練資料不同的測試資料分布，會造成模型預測不準確，同時這種模型無法應用於多種任務。

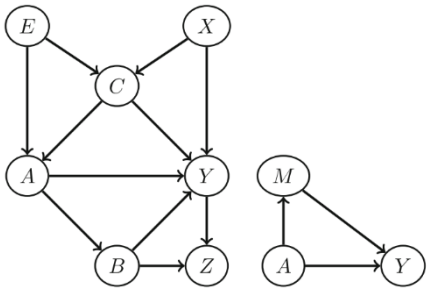


圖 3-2、Causal Bayesian Network(來源[85])

近年來，對於因果推理和預測的研究，開始探討資料的因果學習，[86]提到因果模型和統計模型的差別如下圖，在於將聯合分布(joint distribution)拆解成多個機率分布，每個機率分布由觀察變量和干預變量(intervention)所組成，透過多種干預方式來建立模型對資料深層理解，使得神經網路模型學習到的表示法應用到多種任務，具備強健性，而實作方法包括半監督學習(semi-supervised learning)、對抗樣本(Adversarial examples)等。本計畫擬設計兩個任務分別為使用者介入與否和使用者期待的互動服務種類的分類問題，透過第一年和第二年模型所構建的環境認知、使用者行為、意圖、情緒等資訊，建立多模態變量，同時和影像與語音資料訓練神經網路模型，使得機器人透過當下的觀測能夠感知在環境中是否要介入使用者並提供使用者適當的服務。

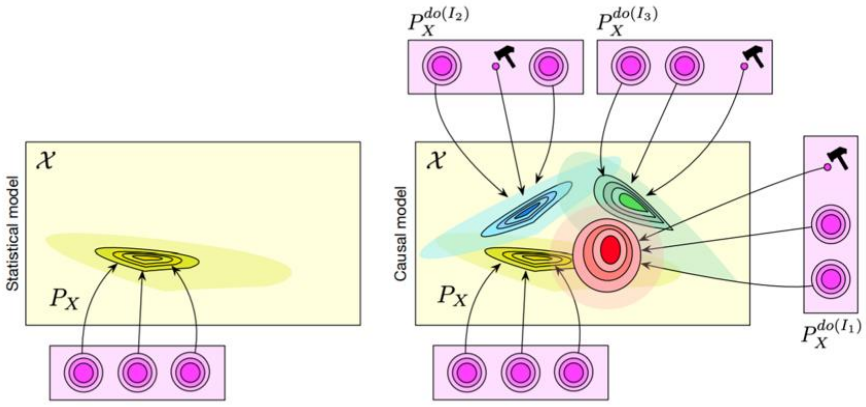


圖 3-3、統計模型和因果模型的差別(來源[85])

本計畫擬首先使用自然語言生成模型如 GPT-2[87]，建立文字資料集以透過自然語言句子描述環境認知、使用者行為、意圖、情緒，使用 TeacherForcing 的方式訓練模型，使得能將環境認知、使用者行為、意圖、情緒等資訊轉化成自然語言句子。再來，本計畫擬參考多模態的 Transformer 架構如 VATT[88]，模型輸入包含來自機器人感測器收到的影片(Video)和語音(Audio)，以及將環境認知、使用者行為、意圖、情緒等使用訓練的自然語言生成模型轉化成文字後輸入，藉由 VATT 作為特徵抽取模型，透過少量的標註資料，使用半監督式學習(semi-supervised)的方法來訓練模型，使得模型能夠預測機器人是否介入使用者以及適合提供的服務種類。

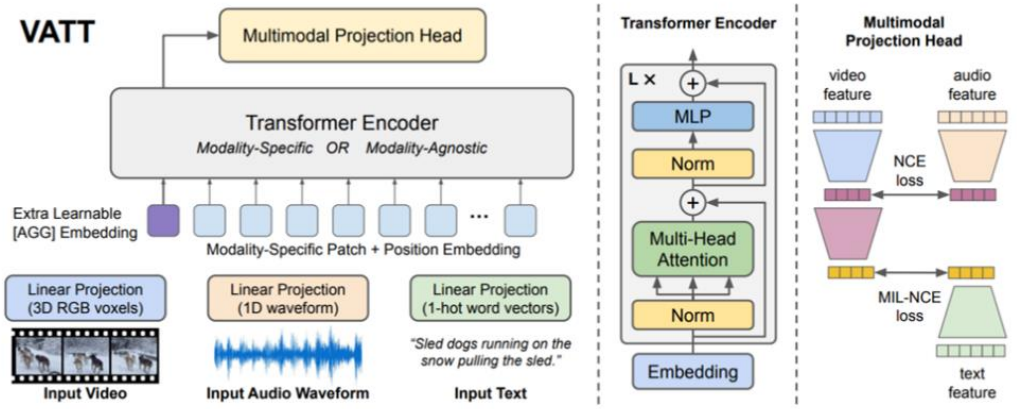


圖 3-4、VATT 模型架構圖(來源[88])

2. 建立信任關係

設計一種能夠贏得人類信任的輔助機器人，使之被人類社會接受、成為人類的陪伴者或顧問，是一項具有挑戰性的工作。人與人之間存在著不同層次的相互依賴關係，彼此間的信任是一個動態建立的過程，可以幫助我們從低級依賴關係(陌生人)轉變為高級(朋友)。而作為一個缺乏移情能力的機器人，如果不知道周遭人類對自己的信任程度，就很難預測互動行為產生的後果。因此，機器人需要尋找策略以在準確的時間推斷和調整人機信任(Human-Robot Trust[89])水平。

思考機器人的社會角色，人們顯然期望機器人能夠引導自己對事件做出正確的決策從而使自己輕易取得最大利益。儘管如此，人類卻無法完全仰賴機器人維繫自身、及社會的穩定發展，因此在此計畫中，我們認同「人們不應過度信任機器人，而應主動自主地做出選擇，並通過積極的思考來評估替代方案，而不是“聽從”機器人的建議」這樣的觀點。具體來說，我們預期機器人在完成既定協作任務的條件下，還應以使人類的認知、心智、積極性等處於良性均衡的狀態為目標，因此，需要研究如何設計策略，使機器人能夠引導人類調整決策、逐步達到人機協作情境下“適當”的信任水平。

2-a) 機器人與人類的信任模型

在過去的近三十年間，人機互動與賽局理論在彼此相對孤立的情況下發展了不同的信任理論：人機互動側重於信任模型的潛在維度、層次、相關性和前因，而賽局理論則側重於單一信任決策背後的心理學和策略，二者都試圖衡量信任的期望及風險。Yosef 在人機信任關係中引入了社交心理學和互賴理論的概念[90]，發現互賴理論的度量比賽局理論提出的理性或規範的心理推理假設能夠更好地捕捉社交中的“過度信任”現象，並進一步探討了互賴理論——其重點是承諾、強制和合作——如何解決人機信任中的許多潛在結構和前因，揭示了機

器人代替人類參與互動產生的關鍵異同主要體現在信任互動中。在此計畫中，我們擬採用互賴理論的原理構建人機信賴關係模型來探索機器人的最佳化策略。

2-b) 相互依賴理論

Harold Kelley 和 John Thibaut 於 1959 年在他們的著作 The Social Psychology of Groups[91] 中首次提出了互賴理論，通常也被稱為相互依賴(依存)理論。在最新的定義中，人際關係是通過人際相互依存來定義的，即人們的互動過程會影響彼此的體驗。相互依存理論有四個基本假設：1）結構原理，2）轉化原理，3）相互作用原理，4）適應原理。

最基本的原理是相互作用原理(也稱為 SABI 模型)，用於評估影響任何給定相互作用的變量，封裝在方程中

$$I = f[A,B,S]$$

即人際互動 (I) 都是給定情況 (S) 的函數 (f)，加上互動中個體 (A & B) 的行為和特徵(動機、特質等)。

在此計畫中，擬利用結構原理中的解構控制，關注互動的一個成員對另一個成員的影響方式。它定義了雙人賽局中的三種控制類型：a) 參與者控制——個體行為對個體結果的影響，b) 夥伴控制——每個個體行為對互動中其他個體結果的影響， c) 聯合控制——每個人的行為對每個人的結果的共同影響，來建立賽局數學模型。以往賽局理論通常對玩家的行為有“理性”假設，賽局均衡策略解法之一是子博弈完美均衡，執行違反信任的行為是被信任一方的弱支配理性策略。但在現實世界中，信任經常被給予和實現。人們選擇信任和值得信賴，而不是僅僅基於收益和風險規避作出選擇。另一種解決方案是混合策略納什均衡，所有參與者都希望最大化收益，這樣沒有參與者可以單方面選擇做得更好的舉動，這將我們的視野從離散決策擴展機率模型的連續解，討論納什均衡點的存在。

2-c) 以信任均衡為目標的行為策略

以往可信賴的機器人技術側重於研究人類對機器人信任的前因和後果，新的側重於開發機器人積極獲得、校準和維持人類用戶信任的策略，在此計畫中，我們將此目標定義為“信任均衡”，即維持人類信任的同時以保持人類的積極性為宗旨，避免發生“過度信任”。此類研究的重點在於賦予機器人認知、推理能力，例如前文中通過概率模型建立因果推理。我們將構建動態模型模擬調節信任的過程，找出機器人的最佳化策略。

第四年

構建自主學習技能：為使社交機器人與人的相處能具有溫度，以便人類社交情感上對該機器人有極高的接受度，則機器人也要能夠敏銳地感知情緒、推導原因，學習具同理心的社交行為，使能做出適切的行動決策以提供及時、貼心的服務。

1. 自主學習

除了基本的交談外，機器人也需要在不同場合來回答使用者的一些疑問，並利用網路上的知識或專家資料來輔助學習新知。本計畫在特定場合問答方面，擬將大量知識引入到資料庫中。若使用者對特定知識有相關的疑慮，可以使用自然的口語對話方式詢問機器人，機器人會根據資料庫中的內容來給予相對應的回覆，以減輕工作人員的負擔，也節省使用者需要查詢的時間。對於資料庫裡缺失的知識而無法回答時，機器人會擷取網路資訊來學習，除了增長資料庫裡現有的知識之外，也會和專家進行再次確認。

本計畫擬使用 Sentence-Transformer[92,93]以及 FAISS[94]這兩項主要的技術，以衛生教育資訊為例，可以將台大醫院健康管理中心提供的健康、衛教知識及現有的大型健康、醫學知識庫的大數據適度轉換後，即可加以訓練出"搜尋健康知識"的深度學習模型，可利用自然語言方式方便地詢問所需的健康、醫學知識。此項目將在第四年進行研究，結合自然語言處理、資料索引與自主學習的技術實現。

近代由於深度學習的發展，使得人們在許多自然語言處理的技術上有很大的突破，但在相似度搜尋(similarity search)的任務中往往因為模型架構而需要耗費大量的運算資源與時間[92]。然而，問答系統需要準確並即時地給予使用者答覆，因此，計畫擬使用深度學習 Sentence-Transformer[92,93]中的 Bi-Encoder 預訓練模型作為語句編碼器(sentence encoder)。Bi-Encoder 使用了孿生網路(Siamese Network)的結構，此訓練方式能讓模型學習到句子之間的關係，得出句向量(sentence embedding)，也可以直接使用餘弦相似度(cosine similarity)來衡量兩個句向量之間的相似度，如圖 4-1 所示。由於資料庫中的文本向量可以預先計算，因此即可大量提升查找的速度。我們可以將使用者的問題映射到高維空間中，和資料庫向量(corpus embeddings)進行相似度比對，並回傳最匹配的答覆給使用者。

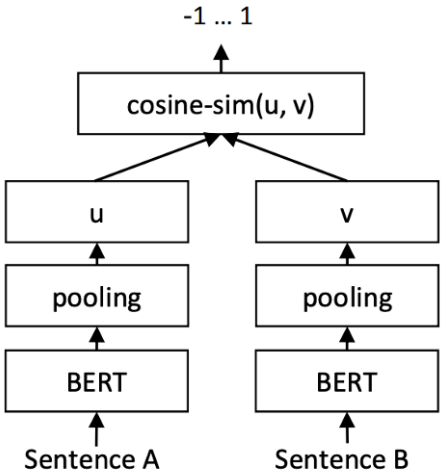


圖 4-1、Bi-Encoder 架構 (來源[92])

擬透過搜集使用者常問的相似問題來訓練模型。訓練時，每一個批次都會隨機選擇 K 個相似問題對為訓練樣本，模型的任務是要將相似問題彼此之間的句向量更為接近，擬使用 Multiple Negative Ranking Loss[94]作為訓練模型時的損失函數。此 loss 期望輸入一批配對過的相似問句 $(a_1, p_1), (a_2, p_2), \ldots, (a_K, p_K)$，其中我們假設 (a_i, p_i) 是兩個相似問句，並假設 (a_i, p_j)，$i \neq j$ 為不相似的問句。Multiple Negative Ranking Loss 的公式如下：

$$L(a, p, \theta) = -\frac{1}{K}\sum_{i=1}^{K}[S(p_i, a_i) - log \sum_{j=1, i\neq j}^{K} e^{S(p_i, a_j)}]$$

其中，θ 為訓練中可調整的模型參數，S 為兩個句子經過模型與餘弦相似度函數得到的數值，由此函數可以看出，在最小化此 Loss 的情況下，會使得相似句對 (a_i, p_i) 的向量愈相近愈好，而不相似問句之間的向量相距越遠。當此函數無再繼續下降時，代表模型已經收斂。

問答系統中的資料庫會有大量資料，為了加速相似度搜尋(similarity search)時所需要的時間，擬使用 FAISS[94]，這是由 Facebook AI Research 所開源的一項技術，為稠密向量

(dense vector)提供高效率的相似度檢索框架，不僅速度快也提供許多檢索方法方便開發者使用。將資料庫內的每一筆數據使用深度學習模型得到固定維度的向量後儲存，並利用 FAISS 來索引，快速查找出最相似的資訊。

然而，經驗和智慧是與時俱進的，對於使用者新穎的問題，機器人應擷取網路資訊或向專家學習最新的知識以回饋給使用者，微調並提升模型的效果。本計畫擬使用基於池的主動學習流程(Pool-based Active Learning Cycle) [95]來最大化模型的效能並同時找出具有代表性的少量樣本，以節省標記資料所需成本，如圖 4-2 所示。若使用者的提問與資料庫內的問題相似度太低，將以網路資訊或專家給予之知識與資料庫答案進行比對，假如資料庫中也無相似答案供參考，則會引入資料庫供機器人學習，否則可利用使用者提供之問句微調語句編碼器模型。

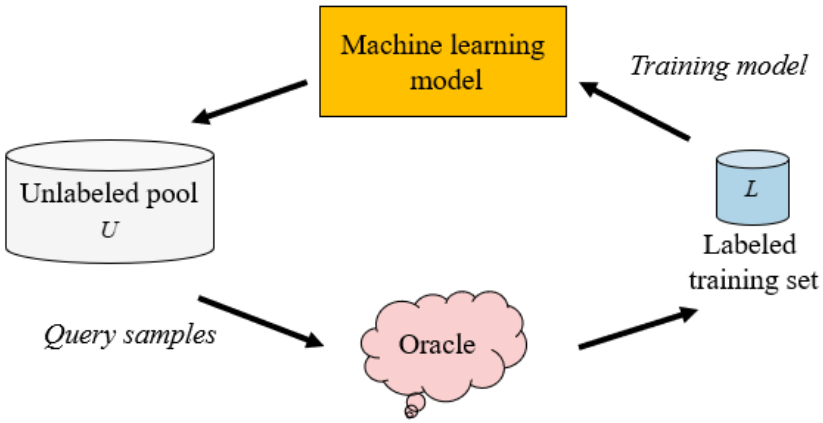


圖 4-2、基於池的主動學習流程(來源[95])

2. 強化學習

強化學習(Reinforcement Learning，簡稱 RL)強調如何基於環境而行動，以取得最大化的預期利益。其靈感源自於心理學中的行為主義理論，即有機體如何在環境給予的獎勵或懲罰的刺激下，逐步形成對刺激的預期，產生能獲得最大利益的習慣性行為。機器人通過對多模態資料的融合學習，可獲得類人的智慧，理解人類社會的社交行為並能夠遵循規則進行自主決策與行動。當前主流的 RL 演算法皆使用深度神經網路(Deep Neural Network)，有基於價值函數的 DQN[96]、及基於策略梯度的 DDPG[97]。但是，深度強化學習有個很大的缺點，它在數據上的採樣效率非常低，即需要大量的訓練資料才能完成學習。

圖 4-3、一般的強化學習

2-a) 記憶輔助神經網路 (Memory-augmented Neural Network)

近年來有不少機器學習的研究在探討神經網路架構的記憶性，如早期的遞歸神經網路(RNN)和長短期記憶模型(LSTM)讓網路可以處理有序列性的資料。後來還有衍生把記憶模組和計算模組分開的架構如 Neural Turing Machine [98] 及其進階版 [99] 等，這些架構在元學習(Meta-learning，含義為學會學習)領域的研究中 [100] 也證明了可以提高模型的學習能力及採樣效率[101]。

2-b) 補償因果推論的強化學習演算法

本計劃擬使用強化式學習結合記憶輔助神經網路及認知機器人領域的相關研究[102,103]，來輔助機器人原本的因果推論系統，提高系統在不同環境的適應性，使機器人可以在更複雜的情況採取相應的動作。

用特定因果推論的方式決定要執行的動作雖然會比起資料導向的方法來得更可靠且更有解釋性，但是可以預期此方法並不夠泛化，可能無法解決每一種情況。所以，本計劃擬訓練一個 RL 模型，當發生因果推論方式無法解決的情形時，此 RL 模型會取而代之。

由於我們已經有一個因果推論的系統，故我們可以利用此系統輸出正確的成對數據，用於訓練、初始化 RL 模型。雖然說此 RL 模型不會處理因果推論系統可以處理的情況，但初始化的意義在於使模型可以有基本的能力。此模型的特點是它可以長期的在線主動學習(online training)，當因果推論系統正常運作時，它會用因果推論系統采取的決策直接更新模型的權重，而當遇到其餘狀況時，RL 會輸出一個決策，再透過人類給予的反饋來更新權重。隨著用戶的使用，我們可以預期此 RL 模型會產出越來越好的策略。

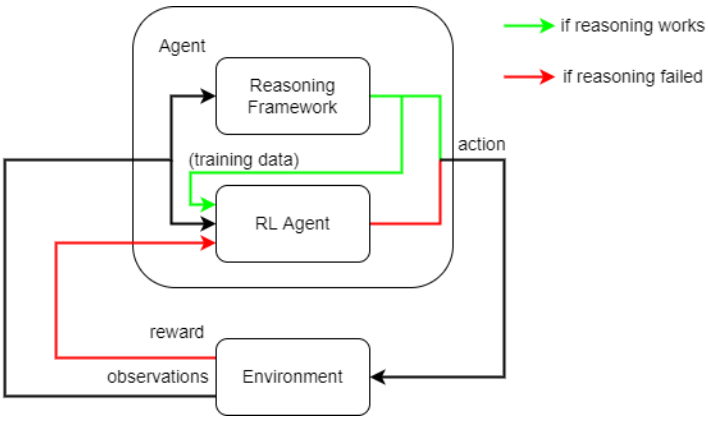


圖 4-4、輔助因果推論系統的強化學習

3. 場域應用

由於本計劃設計之機器人的特殊性質在醫療護理領域具有極高的適配性，我們擬將機器人置於醫院環境或照護機構並完成各項功能之完美應用。

3-a) 機器人於未建圖的陌生環境執行空間認知與導航

由於醫院環境的特殊性質，空間複雜度高且人員密集，預先建圖並獲得準確的地圖需要耗費大量的時間成本。基於對環境的快速認知能力，我們的機器人對於陌生的環境具有強適應力。相比於傳統的移動平台機器人，我們的機器人可以通過目前所處環境的語義地圖結合自身存儲之記憶認知自身的位置並實現在陌生環境中的移動導航，無需預先建圖。

3-b) 社交認知關注周遭人類

醫護人員與病患之人數一直處於一種失衡的狀態，一個醫護人員往往要照顧 5 個甚至以上的病患，對於醫護人員和病患都有巨大的壓力。我們的機器人可以作為勞動力一定程度緩解雙方的壓力。

機器人辨識人類意圖及行為後提供協助，例如，機器人可以代替人類來監管那些需要24小時照料的病患，所有的諸如端茶送水之類的雜事都可以完成。面對一些特殊狀況，機器人也能及時做出反應，如：通過姿勢識別來判斷病患的動作是否有異常或是否有摔倒並做出預警。

機器人亦可通過同理心對談輔助心理治療、與人類建立互賴關係。隨著當前社會工作壓力愈發繁重，出現心理問題之病患數量也率創新高，而心理治療往往費時費力，並且專業性很高，需要有專業的心理治療師才能完成。我們的機器人就可以解決心理治療師人力資源不足的情況。針對擁有心理疾病之病患，機器人可以從患者的動作、語氣、用詞的變化敏銳地判斷其當下的心理狀態，並使用對應的最適交流模式，增進雙方的關係，達到舒緩壓力的作用。

(三) 預期完成之工作項目及成果。

本計畫之研究，擬以四年的進度，完成一個「具有深度(層)認知能力之 AI 社交型機器人系統」，分四個階段逐步建立該社交型機器人不同面向的認知能力。其預期完成之工作項目及成果分別敘述如下:

第一年：建立空間認知及自主移動行為

1. 預期完成之工作項目
   - ➢ 建構社交型移動機器人系統
     - ◆ 硬體架構設計與傳感器單元整合
     - ◆ ROS 軟體系統架構與底層控制模組
     - ◆ 多傳感器系統校正與配準
   - ➢ 開發空間認知與導航演算法
     - ◆ 三維視覺認知(全景分割)演算法開發
     - ◆ 全景分割數據集構建
     - ◆ 物件語義理解演算法開發
     - ◆ 符號化拓撲圖與導航演算法開發
     - ◆ 語義導航演算法開發
     - ◆ 實際場域測試演算法成效
   - ➢ 空間中人群行為/移動模式認知

- 多目標追蹤演算法開發
- 人群行為/移動模式類別定義
- 人群行為/移動式辨識演算法開發
- 改進的社交導航演算法開發
- 實際場域測試演算法成效

2. 對於參與之工作人員，預期可獲之訓練
   - ➢ 機器人軟硬體系統構建
   - ➢ 傳感器資料如 RGB-D 點雲、Lidar 點雲單元測試與資料可視化
   - ➢ 基於認知心理學的社交行為分析(人群移動模式)
   - ➢ 電腦視覺、自然語言雙模態認知語義融合演算法設計、實現與應用
   - ➢ 社交導航演算法設計、實現與應用
3. 預期完成之研究成果
   - ➢ 全景分割模型
   - ➢ 基於多層級空間認知的語義導航與建圖演算法
   - ➢ 多目標追蹤模型
   - ➢ 社交導航演算法
4. 學術研究、國家發展及其他應用方面預期之貢獻
   - ➢ 機器人於非約束環境執行基於認知的輕量化語義導航之應用
   - ➢ 不同語義環境中的人群行為/移動資料

第二年：建立人類情緒與行為認知及適度互動

1. 預期完成之工作項目
   - ➢ 人類意圖及行為辨識
     - 人類活動、目標、計畫之類別定義
     - 活動辨識演算法開發
     - 目標辨識演算法開發
     - 計畫辨識演算法開發
     - 實際場域測試演算法成效
   - ➢ 人類情緒辨識與安撫
     - 情緒辨識模型開發與訓練
     - 同理心之對談語料模型或策略學習
     - 實際場域測試方案成效
   - ➢ 展現同理心、情緒之設計
     - 機器人以表情、動作展現同理心或情緒之設計
     - 心理學實驗以評估設計方案之效用
   - ➢ 人類自傳記憶模型構建
     - 情節記憶構建演算法
     - 語義記憶構建演算法
     - 自傳式記憶模型構建
     - 記憶輔助系統研究與開發
     - 心理學實驗以評估輔助系統之效用
2. 對於參與之工作人員，預期可獲之訓練
   - ➢ 認知心理學對人類行為、情緒、記憶觀點之整理回顧
   - ➢ 人類意圖及行為辨識演算法設計、實驗與應用
   - ➢ 體現情感或同理心之回應設計

- ➢ 設計心理學實驗之經驗

3. 預期完成之研究成果
   - ➢ 人類意圖及行為辨識演算法
   - ➢ 機器人情感及同理心體現之回應方案(對談、表情、動作)
   - ➢ 自傳式記憶模型構建與輔助系統
4. 學術研究、國家發展及其他應用方面預期之貢獻
   - ➢ 長照機構中人類行為、意圖的定義
   - ➢ 機器人社交接受度的資料彙整
   - ➢ 人類自主行為、機器人與人類互動之行為資料

第三年：建立因果認知

1. 預期完成之工作項目
   - ➢ 因果推理
     - 因果表示學習相關方法歸納
     - 基於數據的因果關係模型構建
     - 基於記憶與情緒資料的因果推理演算法開發
   - ➢ 干預問題的預測
     - 獨立同分布設置下的預測演算法開發
     - 分布偏移下的預測演算法開發
     - 回答反事實問題的演算法開發
     - 實際場域測試演算法成效
   - ➢ 建立相互依賴關係
     - 相互依賴理論的數學模型
     - 機器人與人類的信任模型
     - 以信任均衡為目標的行為策略學習演算法開發
     - 實際場域測試演算法成效
2. 對於參與之工作人員，預期可獲之訓練
   - ➢ 記憶、情緒資料因果關係定義
   - ➢ 因果推論演算法設計、實現與應用
   - ➢ 機器人基於數據最佳化干預的心理學實驗方法設計
3. 預期完成之研究成果
   - ➢ 因果模型與推論演算法
   - ➢ 機器人干預人類的最佳化理論
   - ➢ 於自然語言理解的輔助科別分類系統
4. 學術研究、國家發展及其他應用方面預期之貢獻
   - ➢ 因果表示學習之綜述
   - ➢ 應用心理學、相互依賴理論於機器人與人類信任關係構建

第四年：知識自主更新與拓展

1. 預期完成之工作項目
   - ➢ 自主學習
     - 建構問答資料庫
     - 語句相似度模型開發與訓練
     - 主動學習演算法開發

- 實際場域測試演算法成效

➢ 強化學習
- 補償因果推論的強化學習演算法開發
- 基於比較的偏好學習演算法開發
- 實際場域測試演算法成效

➢ 場域應用
- 機器人於未建圖的陌生環境執行空間認知與導航
- 機器人辨識長者意圖及行為後提供協助收集反饋
- 機器人與不同親疏關係之長者進行多回合同理心互動對談
- 機器人與長者建立互賴關係

2. 對於參與之工作人員，預期可獲之訓練
   ➢ 自然語言處理相關演算法的設計、實現與應用
   ➢ 主動學習流程開發與應用
   ➢ 機器人學習之效能實驗設計
   ➢ 強化學習之應用範例
   ➢ 場域應用相關實驗流程設計與執行
3. 預期完成之研究成果
   ➢ 基於人機互動之自主學習演算法
   ➢ 機器人行為優化之偏好學習演算法
4. 學術研究、國家發展及其他應用方面預期之貢獻
   ➢ 自主學習機器人於社會之可行性探討
   ➢ 場域應用資料用於心理學實驗分析

參考文獻

[1] Licklider, Joseph CR. "Man-computer symbiosis." *IRE transactions on human factors in electronics 1* (1960): 4-11.

[2] High, Rob. "The era of cognitive systems: An inside look at IBM Watson and how it works." *IBM Corporation, Redbooks* 1 (2012): 16.

[3] Zhang, Yanyu, et al. "Intelligent hotel ROS-based service robot." 2019 IEEE International Conference on Electro Information Technology (EIT). IEEE, 2019.

[4] Hwang, Jinsoo, Seulgi Park, and Insin Kim. "Understanding motivated consumer innovativeness in the context of a robotic restaurant: The moderating role of product knowledge." Journal of Hospitality and Tourism Management 44 (2020): 272-282.

[5] Lewandowski, Benjamin *et al*. "Socially Compliant Human-Robot Interaction for Autonomous Scanning Tasks in Supermarket Environments." *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2020.

[6] Gubenko, Alla, et al. "Educational Robotics and Robot Creativity: An Interdisciplinary Dialogue." Frontiers in Robotics and AI 8 (2021): 178.

[7] Agrigoroaie, Roxana M., and Adriana Tapus. "Developing a healthcare robot with personalized behaviors and social skills for the elderly." *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2016.

[8] Sharma, Aparna, et al. "A systematic review of assistance robots for elderly care." 2021 International Conference on Communication information and Computing Technology (ICCICT). IEEE, 2021.

[9] Saleh, Mohammed A., Fazah Akhtar Hanapiah, and Habibah Hashim. "Robot applications for autism: a comprehensive review." Disability and Rehabilitation: Assistive Technology 16.6 (2021): 580-602.

[10] Broadbent, Elizabeth. "Interactions with robots: The truths we reveal about ourselves." Annual review of psychology 68 (2017): 627-652.

[11] R. C. Luo, Y. -T. Hsu and H. -J. Ye, "Multi-Modal Human-Aware Image Caption System for Intelligent Service Robotics Applications," 2019 IEEE 28th International Symposium on Industrial Electronics (ISIE), 2019, pp. 1180-1185, doi: 10.1109/ISIE.2019.8781144

[12] Zinchenko, Kateryna, and Kai-Tai Song. "Autonomous Endoscope Robot Positioning Using Instrument Segmentation With Virtual Reality Visualization." IEEE Access 9 (2021): 72614-72623

[13] Y. -H. Lee and K. -T. Song, "Real-time Obstacle Avoidance with a Virtual Torque Approach for a Robotic Tool in the End Effector," 2021 IEEE International Conference on Robotics and Automation (ICRA), 2021, pp. 8436-8442, doi: 10.1109/ICRA48506.2021.9561912

[14] K. -T. Song, C. -W. Chiu, L. -R. Kang, Y. -X. Sun and C. -H. Meng, "Autonomous Docking in a Human-Robot Collaborative Environment of Automated Guided Vehicles," 2020 International Automatic Control Conference (CACS), 2020, pp. 1-6, doi: 10.1109/CACS50047.2020.9289713

[15] Hsieh, Shih-Ho, et al. "A Novel Magnetic Dipoles Localization and Mapping Algorithm using Magnetometer Array." 2019 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM). IEEE, 2019

[16] Lai, Yu-Ting K., et al. "Industrial anomaly detection and one-class classification using generative adversarial networks." 2018 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM). IEEE, 2018

[17] Hsieh, I., et al. "A CNN-Based Wearable Assistive System for Visually Impaired People Walking Outdoors." Applied Sciences 11.21 (2021): 10026

[18] Wong, Ching-Chang, et al. "Motion planning for dual-arm robot based on soft actor-critic." IEEE Access 9 (2021): 26871-26885

[19] Wong, Ching-Chang, et al. "Manipulation Planning for Object Re-Orientation Based on Semantic Segmentation Keypoint Detection." Sensors 21.7 (2021): 2280

[20] C. Chan and C. Tsai, "Collision-Free Speed Alteration Strategy for Human Safety in Human-Robot Coexistence Environments," in IEEE Access, vol. 8, pp. 80120-80133, 2020, doi: 10.1109/ACCESS.2020.2988654

[21] Lane, Geoffrey W., et al. "Effectiveness of a social robot,“Paro,” in a VA long-term care setting." Psychological services 13.3 (2016): 292.

[22] Sheridan, Thomas B. "Human–robot interaction: status and challenges." Human factors 58.4 (2016): 525-532.

[23] Y. Shih, C. Hsu, W. Wang and Y. Wang, "Feature extracted algorithm for simultaneous localization and mapping (SLAM)," 2015 IEEE International Conference on Consumer Electronics (ICCE), 2015, pp. 497-498, doi: 10.1109/ICCE.2015.7066497

[24] C. -H. Chen, C. -C. Wang and S. -F. Lin, "A Navigation Aid for Blind People Based on Visual Simultaneous Localization and Mapping," 2020 IEEE International Conference on Consumer Electronics - Taiwan (ICCE-Taiwan), 2020, pp. 1-2

[25] Olivares-Alarcos, Alberto, et al. "A review and comparison of ontology-based approaches to robot autonomy." *The Knowledge Engineering Review* 34 (2019).

[26] Beeson, Patrick, et al. "Integrating Multiple Representations of Spatial Knowledge for Mapping, Navigation, and Communication." *Interaction challenges for intelligent assistants*. 2007.

[27] Chen, Kevin, et al. "Topological Planning with Transformers for Vision-and-Language Navigation." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021.

[28] Hu, Yue, et al. "Vision-Based Topological Mapping and Navigation With Self-Organizing Neural Networks." *IEEE Transactions on Neural Networks and Learning Systems* (2021).

[29] J. Li, Z. Li, F. Chen, A. Bicchi, Y. Sun and T. Fukuda, "Combined Sensing, Cognition, Learning, and Control for Developing Future Neuro-Robotics Systems: A Survey," in IEEE Transactions on Cognitive and Developmental Systems, vol. 11, no. 2, pp. 148-161, June 2019, doi: 10.1109/TCDS.2019.2897618.

[30] Li, Shih-An, et al. "Auto-maps-generation through Self-path-generation in ROS-based Robot Navigation." Journal of Applied Science and Engineering 21.3 (2018): 351-360.

[31] Chen, L. Y., et al. "Effects of teleoperated humanoid robot application in older adults with neurocognitive disorders in Taiwan: A report of three cases." Aging Medicine and Healthcare 11 (2020): 67-71.

[32] MLA Yun, Sang-Seok, et al. "A robot-assisted behavioral intervention system for children with autism spectrum disorders." Robotics and Autonomous Systems 76 (2016): 58-67.

[33] Jonell, Patrik, et al. "Machine learning and social robotics for detecting early signs of dementia." arXiv preprint arXiv:1709.01613 (2017).

[34] Chen, Che-Wen, et al. "Outpatient text classification using attention-based bidirectional LSTM for robot-assisted servicing in hospital." Information 11.2 (2020): 106.

[35] Stock-Homburg, Ruth. "Survey of Emotions in Human–Robot Interactions: Perspectives from Robotic Psychology on 20 Years of Research." International Journal of Social Robotics (2021): 1-23.

[36] Wei-Te Chen, Su-Chu Lin, Shu-Ling Huang, You-Shan Chung, Keh-Jiann Chen. “E-HowNet and Automatic Construction of a Lexical Ontology”. COLING, Aug 2010.

[37] Acosta, M., Kang, D., & Choi, H. J. (2008, July). Robot with emotion for triggering mixed-initiative interaction planning. In 2008 IEEE 8th International Conference on Computer and Information Technology Workshops (pp. 98-103). IEEE.

[38] Atkinson, Richard C., and Richard M. Shiffrin. "Human memory: A proposed system and its control processes." Psychology of learning and motivation. Vol. 2. Academic Press, 1968. 89-195.

[39] Baddeley, A.D. and G. Hitch, Working memory, in Psychology of learning and motivation. 1974, Elsevier. p. 47-89

[40] J. H. Lui, H. Samani and K. Tien, "An affective mood booster robot based on emotional processing unit," 2017 International Automatic Control Conference (CACS), 2017, pp. 1-6, doi: 10.1109/CACS.2017.8284239.

[41] Bhargava, Preeti, et al. "The robot baby and massive metacognition: Future vision." *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL)*. IEEE, 2012.

[42] Daglarli, Evren. "Computational Modeling of Prefrontal Cortex for Meta-Cognition of a Humanoid Robot." *IEEE Access* 8 (2020): 98491-98507.

[43] Keren, Guy, and Marina Fridin. "Kindergarten Social Assistive Robot (KindSAR) for children’s geometric thinking and metacognitive development in preschool education: A pilot study." *Computers in Human Behavior* 35 (2014): 400-412.

[44] Ramachandran, Aditi, et al. "Thinking aloud with a tutoring robot to enhance learning." *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*. 2018.

[45] Jamet, Frank, et al. "Learning by teaching with humanoid robot: a new powerful experimental tool to improve children’s learning ability." *Journal of Robotics* 2018 (2018).

[46] Kelly, John E. "Computing, cognition and the future of knowing." *Whitepaper, IBM Reseach* 2 (2015).

[47] Proulx, Michael J., et al. "Where am I? Who am I? The relation between spatial cognition, social cognition and individual differences in the built environment." *Frontiers in psychology* 7 (2016): 64.

[48] Bandura, Albert, and Richard H. Walters. Social learning theory. Vol. 1. Prentice Hall: Englewood cliffs, 1977.

[49] Yule, Peter G., et al. *Modelling high-level cognitive processes*. Psychology Press, 2013.

[50] Qi, C. R., Su, H., Mo, K., & Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 652-660).

[51] Chen, X., Milioto, A., Palazzolo, E., Giguere, P., Behley, J., & Stachniss, C. (2019, November). Suma++: Efficient lidar-based semantic slam. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 4530-4537). IEEE.

[52] Milioto, A., Vizzo, I., Behley, J., & Stachniss, C. (2019, November). Rangenet++: Fast and accurate lidar semantic segmentation. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 4213-4220). IEEE.

[53] K. Mangalam, Y. An, H. Girase, J. Malik, "From Goals, Waypoints & Paths To Long Term Human Trajectory Forecasting," *2021 International Conference on Computer Vision (ICCV)*, 2021, pp 15233-15241

[54] C.H. Wang, Y.C. Wang, M.Z. Xu, David J. Crandall, "Stepwise Goal-Driven Networks for Trajectory Prediction," *2021 arXiv Computer Vision and Pattern Recognition* .

[55] J. Massardi, M. Gravel and É. Beaudry, "PARC: A Plan and Activity Recognition Component for Assistive Robots," *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2020, pp. 3025-3031, doi: 10.1109/ICRA40945.2020.9196856.

[56] Van-Horenbeke, F. A., & Peer, A. (1AD, January 1). Activity, plan, and goal recognition: A Review. *Frontiers*. Retrieved December 18, 2021, from https://www.frontiersin.org/articles/10.3389/frobt.2021.643010/full

[57] Goldman, R. P., Kabanza, F., & Bellefeuille, P. (2019, July 16). Plan libraries for plan recognition: Do we really know what they... *OpenReview*. Retrieved December 18, 2021, from https://openreview.net/forum?id=SyELUyb_-B

[58] Massardi, Jean and Éric Beaudry. “Toward Detecting Anomalies in Activities for Daily Living with a Mobile Robot Using Plan Recognition.” (2020).

[59] Minaee, Shervin, Mehdi Minaei, and Amirali Abdolrashidi. “Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network.” *Sensors* 21.9 (2021): 3046. *Crossref*. Web.

[60] Ung Park, Minsoo Kim, Youngeun Jang, GiJae Lee1, KangGeon Kim, Ig-Jae Kim, Jongsuk Choi. 2021. “Robot Facial Expression Framework for Enhancing Empathy in Human-Robot Interaction.” *2021 30th IEEE*

International Conference on Robot and Human Interactive Communication (RO-MAN) August 8-12, 2021. Vancouver, Canada (Virtual Conference).

[61] Peeraya Sripian , Muhammad Nur Adilin Mohd Anuardi , Jiawei Yu and Midori Sugaya. 2021. "The Implementation and Evaluation of Individual Preference in Robot Facial Expression Based on Emotion Estimation Using Biological Signals."

[62] Saif Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 english words. *In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 174–184.*

[63] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding. *In 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL).*

[64] Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. *In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 4040–4054, Online. Association for Computational Linguistics.*

[65] Jiaqi Shi , Chaoran Liu, Carlos Toshinori Ishi and Hiroshi Ishiguro. 2021. "Skeleton-Based Emotion Recognition Based on Two-Stream Self-Attention Enhanced Spatial-Temporal Graph Convolutional Network

[66] Sicheng Zhao, Yunsheng Ma, Yang Gu, Jufeng Yang, Tengfei Xing, Pengfei Xu, Runbo Hu, Hua Chai, Kurt Keutzer. 2020. "An End-to-End Visual-Audio Attention Network for Emotion Recognition in User-Generated Videos." *In The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20)*

[67] " Babette Rothschild. 2006. Help for the helper: The psychophysiology of compassion fatigue and vicarious trauma. *WW Norton & Company.*

[68] Healey Meghan L., Grossman Murray. 2018. Cognitive and Affective Perspective-Taking: Evidence for Shared and Dissociable Anatomical Substrates. *Frontiers in Neurology.*

[69] Welivita, Anuradha and Pu, Pearl. 2020. A Taxonomy of Empathetic Response Intents in Human Social Conversations. *Proceedings of the 28th International Conference on Computational Linguistics.*

[70] Chujie Zheng and Yong Liu and Wei Chen and Yongcai Leng and Minlie Huang. 2021. CoMAE: A Multi-factor Hierarchical Framework for Empathetic Response Generation. *Association for Computational Linguistics.*

[71] Conway, Martin A., and Christopher W. Pleydell-Pearce. "The construction of autobiographical memories in the self-memory system." *Psychological review 107.2 (2000)*: 261.

[72] Hsiao, Yu-Ting, Edwinn Gamborino, and Li-Chen Fu. "A Hybrid Conversational Agent with Semantic Association of Autobiographic Memories for the Elderly." *International Conference on Human-Computer Interaction. Springer, Cham, 2020*.

[73] C. -Y. Yang, E. Gamborino, L. -C. Fu and Y. -L. Chang, "A Brain-Inspired, Self-Organizing Episodic Memory Model for a Memory Assistance Robot," *IEEE Transactions on Cognitive and Developmental Systems, doi: 10.1109/TCDS.2021.3061659*.

[74] Idrees, Ifrah, Steven P. Reiss, and Stefanie Tellex. "Robomem: Giving long term memory to robots." *arXiv preprint arXiv:2003.10553* (2020).

[75] Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems.* 2017.

[76] Kenton, Jacob Devlin Ming-Wei Chang, and Lee Kristina Toutanova. "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding." *Proceedings of NAACL-HLT. 2019.*

[77] Li, Xinyu, et al. "DuEE: A Large-Scale Dataset for Chinese Event Extraction in Real-World Scenarios." *CCF International Conference on Natural Language Processing and Chinese Computing. Springer, Cham,* 2020.

[78] Dzendzik, Daria, Carl Vogel, and Jennifer Foster. "English Machine Reading Comprehension Datasets: A Survey." *arXiv preprint arXiv:2101.10421* (2021).

[79] Shi, Peng, and Jimmy Lin. "Simple bert models for relation extraction and semantic role labeling." *arXiv preprint arXiv:1904.05255* (2019).

[80] Stephen Robertson and Hugo Zaragoza. 2009. The Probabilistic Relevance Framework: BM25 and Beyond. Found. Trends Inf. Retr. 3, 4 (April 2009), 333–389. https://doi.org/10.1561/1500000019

[81] Khattab, Omar, and Matei Zaharia. "Colbert: Efficient and effective passage search via contextualized late interaction over bert." Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval. 2020.

[82] Karpukhin, Vladimir, et al. "Dense Passage Retrieval for Open-Domain Question Answering." *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020.

[83] Reimers, Nils, et al. "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks." *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics*, 2019.

[84] GoEmotions: A Dataset for Fine-Grained Emotion Classification (https://ai.googleblog.com/2021/10/goemotions-dataset-for-fine-grained.html)

[85] Chiappa, Silvia, and William S. Isaac. "A causal bayesian networks viewpoint on fairness." IFIP International Summer School on Privacy and Identity Management. Springer, Cham, 2018.

[86] Schölkopf, Bernhard, et al. "Toward causal representation learning." *Proceedings of the IEEE 109.5* (2021): 612-634.

[87] Radford, Alec, et al. "Language models are unsupervised multitask learners." *OpenAI blog 1.8* (2019): 9.

[88] [4]Akbari, Hassan, et al. "Vatt: Transformers for multimodal self-supervised learning from raw video, audio and text." arXiv preprint arXiv:2104.11178 (2021).

[89] Lewis, Michael, Katia Sycara, and Phillip Walker. "The role of trust in human-robot interaction." Foundations of trusted autonomy. Springer, Cham, 2018. 135-159.

[90] Razin, Yosef S., and Karen M. Feigh. "Committing to Interdependence: Implications from Game Theory for Human-Robot Trust." arXiv preprint arXiv:2111.06939 (2021).

[91] Thibaut, John W., and Harold H. Kelley. The social psychology of groups. Routledge, 2017.

[92] Reimers, N., & Gurevych, I. (2019, November), "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," *In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Pro-cessing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (pp. 3982-3992).

[93] Reimers, N., & Gurevych, I. (2020, November), "Making Monolingual Sentence Embeddings Multilingual Using Knowledge Distillation," *In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 4512-4525).

[94] Johnson, J., Douze, M., & Jégou, H. (2019), "Billion-scale similarity search with gpus," *IEEE Transactions on Big Data*.

[95] Ren, P., Xiao, Y., Chang, X., Huang, P. Y., Li, Z., Gupta, B. B., ... & Wang, X. (2021). A survey of deep active learning. *ACM Computing Surveys (CSUR), 54(9), 1-40*

[96] Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." *arXiv preprint arXiv:1312.5602* (2013).

[97] Lillicrap, Timothy P., et al. "Continuous control with deep reinforcement learning." *arXiv preprint arXiv:1509.02971* (2015).

[98] Zhu, Guangxiang, et al. "Episodic Reinforcement Learning with Associative Memory." *ICLR 2020 : Eighth International Conference on Learning Representations,* 2020.

[99] Graves, Alex, Greg Wayne, and Ivo Danihelka. "Neural turing machines." *arXiv preprint arXiv:1410.5401* (2014).

[100] Parisotto, Emilio, and Ruslan Salakhutdinov. "Neural Map: Structured Memory for Deep Reinforcement Learning." *International Conference on Learning Representations.* 2018.

[101] Santoro, Adam, et al. "Meta-learning with memory-augmented neural networks." *International conference on machine learning*. PMLR, 2016.

[102] R. Salgado, F. Bellas, P. Caamaño, B. Santos-Díez and R. J. Duro, "A procedural Long Term Memory for cognitive robotics," *2012 IEEE Conference on Evolving and Adaptive Intelligent Systems*, 2012, pp. 57-62.

[103] Vernon, David, Michael Beetz, and Giulio Sandini. "Prospection in cognition: the case for joint episodic-procedural memory in cognitive robotics." *Frontiers in Robotics and AI 2* (2015): 19.