

# 科技部補助專題研究計畫報告

## 機器學習輔助人機互動以提升服務型機器人之社交智慧

報告類別：成果報告

計畫類別：個別型計畫

計畫編號：MOST 108-2218-E-992-310-

執行期間：108年06月01日至109年05月31日

執行單位：國立高雄科技大學電機資訊學院電腦與通訊工程系

計畫主持人：曾士桓

計畫參與人員：碩士班研究生-兼任助理：孫尉豪

碩士班研究生-兼任助理：尤翔鵬

本研究具有政策應用參考價值：☒否 ☐是，建議提供機關  
(勾選「是」者，請列舉建議可提供施政參考之業務主管機關)  
本研究具影響公共利益之重大發現：☐否 ☐是

中華民國 109 年 08 月 08 日

中文摘要：隨著機電一體化、人工智慧、機器學習、計算機視覺和語音信號處理的進步使機器人系統應用於許多現實世界的環境。由於許多這些環境涉及人類，因此開發具有社交智慧(Social Intelligence)的機器人系統是一個越來越重要的目標。本計畫欲藉由服務型機器人的特色於能盡量與人們親切的互動，因此架構規劃整合AI(人工智慧)、機器學習的自動化系統，開發創新的人機互動服務功能，建構服務型機器人在社交環境下兼顧人性化與個人化。本計畫所研發之功能模組及人機互動方式可運用於目前業界開發之機器人平台，逐步應用於一般環境中。

本計畫已完成項目有三：一是利用Self-play reinforcement learning方法實作機器人在五子棋對弈，其方法包含了Monte-Carlo Tree Search(MCTS)演算法以及深度學習的技術，透過自我訓練進而不斷強化人工智慧在五子棋對弈的能力，以便達到能與人進行對戰的程度；二是開發Zenbo服務型智慧機器人之餐廳服務功能，其功能包含餐廳介紹、帶位和點餐服務；三則是結合了驅力(drives)、動機(motivation)，以及強化式學習，讓機器人在產生出主宰的動機後，利用強化式學習來學習動機。並導入了使用者意圖推論以及基於使用者意圖的決策方式。

中文關鍵詞：社交智慧型機器人，強化式學習，深度學習。

英文摘要：Robotic systems have been applied to many real-world environments with advances in mechatronics, artificial intelligence, machine learning, computer vision, and voice signal processing. Since many of these environments involve humans, the development of robotic systems with social intelligence is an increasingly important goal. This project intends to use the characteristics of service robots to interact with people as kindly as possible. Therefore, the framework plans to integrate AI (artificial intelligence) and machine learning automation systems, develop innovative human-computer interaction service functions, and build service robots in Both humanity and personalization are taken into account in a social environment. The functional modules and human-machine interaction methods developed in this project can be applied to the robot platforms currently developed in the industry and gradually applied in general environments. Three works have been completed in this project: One is to use the Self-play reinforcement learning method to implement a robot playing Gobang game. The method includes the Monte-Carlo Tree Search (MCTS) algorithm and deep learning technology, which is continuously strengthened through self-training. The ability of artificial intelligence to play backgammon in order to reach the level of fighting against people; the second is to develop the restaurant service function of Zenbo service-type intelligent robot, which includes restaurant introduction,

seat and ordering services; the third is a combination of driving Drives, motivation, and reinforcement learning allow robots to use reinforcement learning to learn motivation after they generate dominant motivation. And introduced user intention inference and decision-making methods based on user intention.

英文關鍵詞：Social Intelligent Robot, Reinforce Learning, Deep Learning.

# 科技部補助專題研究計畫成果報告

(☐期中進度報告/☒期末報告)

機器學習輔助人機互動以提升服務型機器人之社交智慧

計畫類別：☒個別型計畫 ☐整合型計畫

計畫編號：MOST 108-2218-E-992-310-

執行期間：108 年 6 月 01 日至 109 年 5 月 31 日

執行機構及系所：國立高雄科技大學 電腦與通訊工程系

計畫主持人：曾士桓

共同主持人：

計畫參與人員：尤翔鵬、孫尉豪

本計畫除繳交成果報告外，另含下列出國報告，共 0 份：

☐執行國際合作與移地研究心得報告

☐出席國際學術會議心得報告

☐出國參訪及考察心得報告

本研究具有政策應用參考價值：☒否 ☐是，建議提供機關\_\_\_\_\_

(勾選「是」者，請列舉建議可提供施政參考之業務主管機關)

本研究具影響公共利益之重大發現：☒否 ☐是

中 華 民 國 109 年 8 月 10 日

# 目錄

摘要.....	II
Abstract.....	II
一、 前言.....	1
二、 研究目的.....	1
三、 文獻探討.....	1
四、 研究方法暨成果.....	3
五、 結果與討論.....	17

## 摘要

隨著機電一體化、人工智慧、機器學習、計算機視覺和語音信號處理的進步使機器人系統應用於許多現實世界的環境。由於許多這些環境涉及人類，因此開發具有社交智慧(Social Intelligence)的機器人系統是一個越來越重要的目標。本計畫欲藉由服務型機器人的特色於能盡量與人們親切的互動，因此架構規劃整合 AI(人工智慧)、機器學習的自動化系統，開發創新的人機互動服務功能，建構服務型機器人在社交環境下兼顧人性化與個人化。本計畫所研發之功能模組及人機互動方式可運用於目前業界開發之機器人平台，逐步應用於一般環境中。

本計畫已完成項目有三：一是利用 Self-play reinforcement learning 方法實作機器人在五子棋對弈，其方法包含了 Monte-Carlo Tree Search(MCTS)演算法以及深度學習的技術，透過自我訓練進而不斷強化人工智慧在五子棋對弈的能力，以便達到能與人進行對戰的程度；二是開發 Zenbo 服務型智慧機器人之餐廳服務功能，其功能包含餐廳介紹、帶位和點餐服務；三則是結合了驅力(drives)、動機(motivation)，以及強化式學習，讓機器人在產生出主宰的動機後，利用強化式學習來學習動機。並導入了使用者意圖推論以及基於使用者意圖的決策方式。

**關鍵詞：** 社交智慧型機器人，強化式學習，深度學習。

## Abstract

Robotic systems have been applied to many real-world environments with advances in mechatronics, artificial intelligence, machine learning, computer vision, and voice signal processing. Since many of these environments involve humans, the development of robotic systems with social intelligence is an increasingly important goal. This project intends to use the characteristics of service robots to interact with people as kindly as possible. Therefore, the framework plans to integrate AI (artificial intelligence) and machine learning automation systems, develop innovative human-computer interaction service functions, and build service robots in Both humanity and personalization are taken into account in a social environment. The functional modules and human-machine interaction methods developed in this project can be applied to the robot platforms currently developed in the industry and gradually applied in general environments.

Three works have been completed in this project: One is to use the Self-play reinforcement learning method to implement a robot playing Gobang game. The method includes the Monte-Carlo Tree Search (MCTS) algorithm and deep learning technology, which is continuously strengthened through self-training. The ability of artificial intelligence to play backgammon in order to reach the level of fighting against people; the second is to develop the restaurant service function of Zenbo service-type intelligent robot, which includes restaurant introduction, seat and ordering services; the third is a combination of driving Drives, motivation, and reinforcement learning allow robots to use reinforcement learning to learn motivation after they generate dominant motivation. And introduced user intention inference and decision-making methods based on user intention.

**Keywords：** Social Intelligent Robot, Reinforce Learning, Deep Learning.

## 一、 前言

近年來，機器人已經從傳統的，受限制的工業環境遷移到更多社會和結構較差的環境，如博物館[1]，機場[2]，餐館[3]和護理中心[4]。隨著機電一體化、人工智慧、機器學習、計算機視覺和語音信號處理的進步使機器人系統應用於許多現實世界的環境。由於許多這些環境涉及人類，因此開發具有社交智慧(Social Intelligence)的機器人系統是一個越來越重要的目標。

這種轉變給機器人技術帶來了無數的挑戰，從硬體到高水準行為設計。一個特別棘手的挑戰是為機器人灌輸足夠的社交智慧，即確保他們在與人交往或在人類面前完成任務時尊重社會規範(Social Norm)。例如，面向一群人的機器人必須保持適當的距離，方向，目光接觸水平等。

## 二、 研究目的

本計畫欲藉由服務型機器人的特色於能盡量與人們親切的互動，因此架構規劃整合 AI(人工智慧)、機器學習的自動化系統，開發創新的人機互動服務功能，建構服務型機器人在社交環境下兼顧人性化與個人化。本計畫所研發之功能模組及人機互動方式可運用於目前業界開發之機器人平台，逐步應用於一般環境中。

## 三、 文獻探討

近年來國內外有許多種類的商業用服務型機器人的研究及產品，如美國 MIT Jibo[5] 和華碩 Zenbo Junior<sup>1</sup>，最近陸續問世，如圖 1 所示，Jibo 跟 Zenbo 皆有可擔任家中行政助理的功能，例如，Jibo 與 Zenbo 都有及時提醒的功能，提醒使用者代辦事項的時間，並都能提供視訊和語音對話的功能，其中 Jibo 較為特殊之處在於可以透過本身的一些轉動機構產生不同的姿勢便於讓與他人互動更加有趣，而 Zenbo 機器人則是透過螢幕上的表情來達成與人有趣的互動，此外 Zenbo 還可以利用語音來操作家電的功能，讓使用者的家居更有智慧。



(a)



(b)

圖 1. (a) MIT 的 Jibo，(b) 華碩生產之 Zenbo

---

<sup>1</sup> <https://zenbo.asus.com/tw/>

鑒於機器人在人類生活中快速竄起的趨勢，國內 3C 知名企業：「鴻海股份有限公司」近來亦投入機器人研發，並代為生產製造一款名為 Pepper 機器人(如圖 2(a) 所示)，目前已在國內銷售上市。Pepper 的高度為 121cm、重量 28kg，是一具在結構上仿效人的臉部、手臂與身體的人型機器人。另外，鴻海買下日本夏普公司，旗下的 RoBoHoN 手機機器人，有頭、有四肢，大約 19.5 公分高的小機器人(如圖 2(b) 所示)，支援 3G、4G 和 Wifi 網路。近來，鴻海內部的 B 次集團亦生產一款小樂居家機器人(圖 2(c) 所示)，橢球造型，可置放於桌上、遊走地面，也可自我充電，結合阿里巴巴雲端功能，供作家中小管家之用。由上述發展可看出鴻海欲發展智慧服務型機器人之企圖，結合未來物聯網，鎖定一般消費族群。



(a)



(b)



(c)

圖 2. (a) Pepper, (b) RoBoHoN, (c) 小樂

然而，以上的機器人所展現的多半是單一功能，其利用的人工智慧技術層次與社交互動的變化性也是相當有限，因此在與人的交流、對話和行為上，難以滿足人們在互動多樣性的要求，也因此無法在複雜度高的應用情境下適時地轉換自己的角色提供相對應的服務。本計畫擬藉市場上現有之商業服務型機器人，逐步建立服務型機器人各式之軟體關鍵智慧模組，包括「導航」、「避障」、「感知」、「人員追蹤」、「語音對話」等等，並有效率地整合各項關鍵模組並導入人工智慧的技術以提升服務型機器人與人互動的多樣性，打造一個可替使用者提供創新協同合作的社交環境。

### 重要參考文獻

- [1] S. Thrun, M. Bennewitz, W. Burgard, A. B. Cremers, Dellaert, et al., "Minerva: A second-generation museum tour-guide robot", in Robotics and automation, 1999., vol. 3. IEEE, 1999.
- [2] R. Triebel, K. Arras, R. Alami, L. Beyer, S. Breuers, R. Chatila, M. Chetouani, Cremers, et al., "Spencer: A socially aware service robot for passenger guidance and help in busy airports", 2015.
- [3] Y. Qing-xiao, Y. Can, F. Zhuang, and Z. Yan-zheng, "Research of the localization of restaurant service robot", International Journal of Advanced Robotic Systems, vol. 7, no. 3, p. 18, 2010.
- [4] K. Shiarlis, J. Messias, et al., "TERESA: a socially intelligent semi-autonomous telepresence system", 2015.
- [5] Rane, Pranav, Varun Mhatre, and Lakshmi Kurup. "Study of a home robot: Jibo." International journal of engineering research and technology. Vol. 3. No. 10 (October-2014). IJERT, 2014.



## 四、 研究方法暨成果

現有市面上陪伴型機器人(如 Pepper、Zenbo) 所表現之能力均是透過事前人為程式設定，一旦“淺顯”的環境條件滿足後即啟動預設功能，否則即停留於“待命”(standby)狀態，但面對陪伴/照護長者之長時、多變、不確定、個人化環境，現有機器人技術顯然不足以應付此等挑戰，故本計畫提出以下三種研究尋求改善：

### ● 服務型機器人之深度強化學習

作者 D. Silver 等人[1]於 2016 年提出關於 AlphaGo 整個系統的架構，從策略模擬至自我對下的過程皆有詳細描述，主要架構是由深卷積神經網路加上對局資料庫進行訓練，透過 Monte-Carlo Tree Search 強化整個 Policy Network，再利用 Policy Network 以迴歸方式計算出 Value Network，這樣在實際對弈時不僅可以使用 Policy Network 搭配 Monte-Carlo Tree Search 尋選擇落子位置，並使用 Value Network 縮短 Monte-Carlo Tree Search 的時間及搜尋深度且預測輸贏的結果，而在實際對弈中也確實地打敗了世界棋王。

作者 D. Silver 等人[2]於 2017 年又提出 AlphaGo Zero，主要描述 AlphaGo Zero 在學習過程中，不需要依靠人對它進行指導，也就是不必藉由吸收我們提供的知識及經驗，透過此篇描述的新的強化學習演算法，使用一個無經驗的神經網路搭配搜尋演算法，靠著自我對弈，從零逐漸提升為超越 AlphaGo 的水準。

以上兩者最大的差異在於 AlphaGo 在訓練時需要依靠棋譜，人為處理過的數據，也就是過往的經驗並進行反饋來預測或決策，而 AlphaGo Zero 卻不用，只需要給他圍棋的規則，讓其從自我對弈中不斷的摸索，憑藉著自我訓練來建立出自己的模式，時間上，AlphaGo Zero 也遠遠少於 AlphaGo，所需用的硬體資源也較少，這也能看出，利用監督學習加上強化學習與單純的強化學習這兩者所展現出的結果。

本計畫將 Self-play reinforcement learning 方法實作機器人在五子棋對弈，其方法包含了 Monte-Carlo Tree Search(MCTS)演算法以及深度學習的技術，透過自我訓練進而不斷強化人工智慧在五子棋對弈的能力，以便達到能與人進行對戰的程度。訓練完成後，與人進行對戰方面，玩家一開始可以選擇先手以及後手，接著選擇要遊玩的難易度，總共分為 Easy、Normal 和 Hard，讓玩家可以與不同強度的人工智慧進行對弈，並體驗其中不同強度人工智慧的差異。

#### 1. Monte-Carlo Tree Search(MCTS)結合深度學習

MCTS 是對弈遊戲常用的搜尋演算法，其流程如圖 3 所示：

- Select 始於樹之根節點，當遇到葉節點則終止，選擇落子是透過具有最大值的  $Q(s, a)$  加上 upper confidence bound(UCB)  $U(s, a)$  來決定，而  $U(s, a)$  取決於先前儲存的概率  $P$  以及該節點的訪問次數  $N$ ， $N$  隨著訪問的次數遞增。
- Expend and evaluate 如遇到葉節點則把當前  $s$  輸入 Deep Neural network 得出該節點之  $(p,$

v) 訊息，接著進行 Backup。Backup 從葉節點開始回傳該路徑之 Q 值與 V 值。Play 一旦搜尋完成，會在節點 s 處選擇 a 進行落子。

對於整體來說，MCTS 是外層的大框架，Deep Neural network 則是輔助 MCTS 進行搜索。因此對於一個盤面 s，我們使用 MCTS 建造一棵樹，而在建造的過程中若碰到葉節點則返回其  $(p, v)$ 。建立完成後則會取得  $\pi$ ，並根據  $\pi$  來決定下一步棋的位置，接著取得下一個 s，重複此過程，一局棋結束後則會取得結果 z。

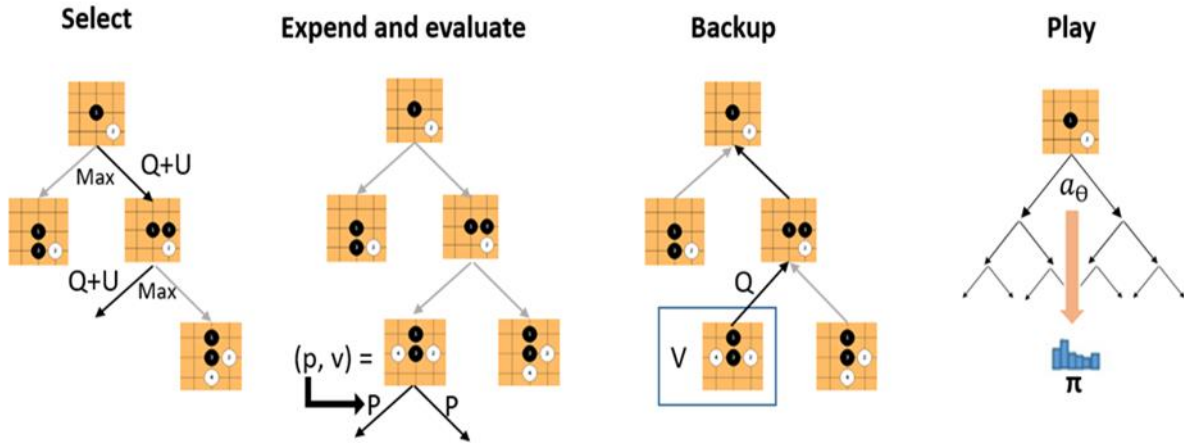


圖 3. Monte-Carlo tree search 演算法步驟[2]

## 2. 策略價值網路(Policy-Value Network) 訓練

在最一開始，由於 model 未經過訓練，因此在 self-play 的過程中，就如同兩個皆不會下棋的人在對弈，也因此要用很多時間去訓練和優化，而在訓練過程中，要得知 model 是否有優化與更強，就必須每隔一段時間去評估它。

從圖 4 的自我訓練流程圖中，我們與 AlphaGo Zero 最大差別是在評估神經網路這步驟，AlphaGo Zero 用的評估方式是每隔一段時間，將新的 model 和舊的 model 進行對戰，有如自己每訓練一段時間去跟一個和自己實力相當的人對戰，不過這方法的缺點較難以判定自己是否真的變強，而且 model 收斂速度較慢；而為了要在短時間訓練成靠譜的 model，所以，我們改變了評估方式，就是每隔一段時間，將新的 model 和加入 heuristic function 的 Pure MCTS 去對戰，這就像自己每訓練一段時間，去跟一個會下棋的人對戰，看自己的實力是否有提升，以及提升了多少。綜合上述，跟 AlphaGo Zero 相較之下，我們所用的評估方式可讓 model 收斂速度有較快的優點。

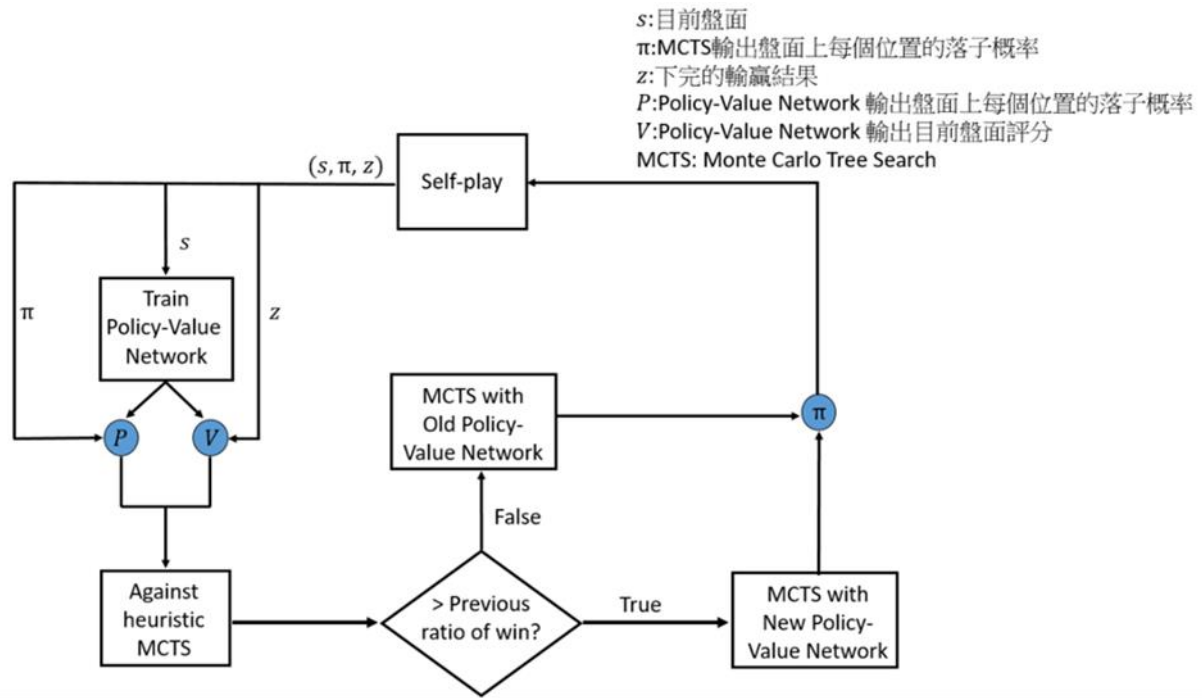


圖 4. 五子棋對弈之自我訓練流程圖

我們 Training 的方法是利用 Self-play 搭配強化式學習的演算法，去訓練策略價值網路 (Policy Value Network) [1]。實作參考作者宋俊瀟提供的程式碼<sup>2</sup>，詳細說明如下：

#### (1) 訓練的 data 從哪裡來？

我們訓練的 data 是透過 Self-play，如圖 5 所示，Self-play 的 data 直接從當前最佳的 model 產生，並用來優化自己。

#### (2) Self-play 生成 Data 的多樣性

為了訓練出有效的策略價值網路，就需覆蓋各式各樣的局面，所以 data 的多樣性是相當重要的，因此我們在 self-play 中加入了 2 種 exploration 的手段，每一次 self-play 前 30 步，使用 MCTS 搭配策略價值網路去 exploration，之後的 exploration 是用 Dirichlet noise 去實現。

#### (3) Data 的保存和擴充

在 self-play 過程中，我們會以當前 player 去儲存一些數據  $(s, \pi, z)$  到 data buffer 中，其中  $s$  代表盤面，會用兩個二值矩陣分別表示自己和對手棋子的位置，另外，五子棋也

<sup>2</sup> 宋俊瀟，AlphaZero 實戰：從零學下五子棋

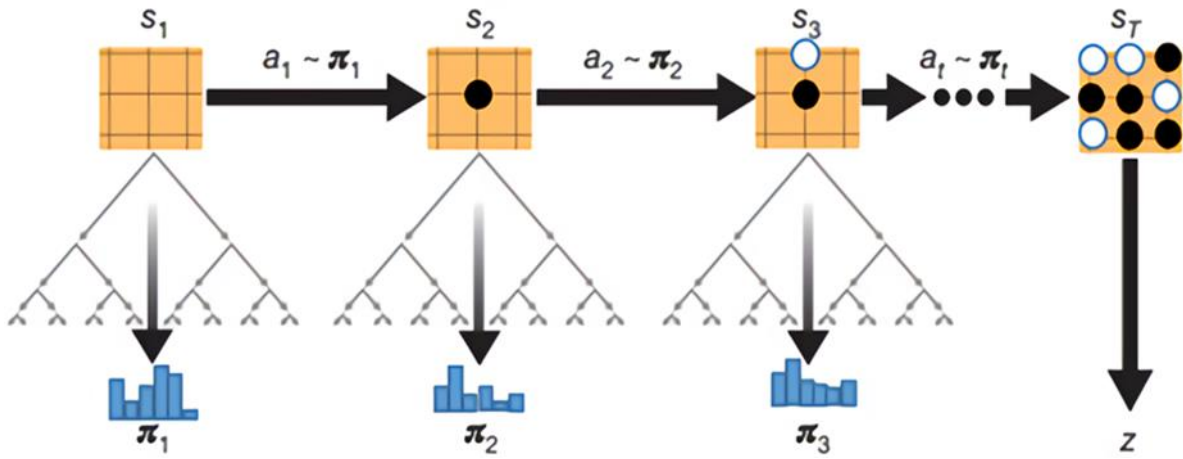


圖 5. Self-play 流程圖[2]

和圍棋一樣有對稱性的等價性質，所以我們會把  $s$  進行旋轉和鏡像翻轉，擴充成 8 種等價狀況； $\pi$  是 MCTS 輸出的概率， $z$  是目前這局下完的結果，贏等於 1，輸等於 -1，平手等於 0。

#### (4) 何謂策略價值網路

所謂的策略價值網路  $f_\theta$  [2]，就是給定它當前的盤面  $s$  作為輸入，而輸出當前盤面每個位置下棋的機率  $p$  和一個對於整個盤面評分值  $v$ ， $(p, v) = f_\theta(s)$

#### (5) 策略價值網路架構

圖 6 顯示當一局棋結束後則會取得結果  $z$  之後，針對每一步中的策略價值網路預測出的  $v_t$  進行最佳化。單一的策略價值網路如圖 7 所示，第一層到第三層使用卷積層，分別使用 32、64、128 個 3x3 的 filter，使用 ReLu 激勵函數。然後分成 policy 和 value 這兩個輸出，policy 這端使用 4 個 1x1 的 filter 來降維，再連接一個全連接層，最後透過 Softmax 非線性函數直接輸出棋盤上每個位置的落子概率；value 這端先使用 2 個 1x1 的 filter 來降維，然後連接一個 64 個神經元的全連接層，再連接一個全連接層，最後使用 tanh 非線性函數直接輸出  $[-1, 1]$  之間的局面評分。

#### (6) 策略價值網路訓練

根據策略價值網路訓練過程圖，我們的目標就是要讓策略價值網路輸出的概率  $p$  更接近 MCTS 輸出概率  $\pi$ ；讓輸出盤面的評分  $v$  更接近預測實際結果  $z$ 。優化方面，我們參考 AlphaGo Zero 的論文使用的損失函數[2]：

$$l = (z - v)^2 - \pi^T \log p + c \|\theta\|^2$$

其中第三項是用於防止過擬合的正規項。

#### (7) 策略價值網路的評估方式

每隔 50 次的 self-play，我們會對當前的 model 進行評估，評估方式是將新的 model 搭配 MCTS 和加入 heuristic function 的 Pure MCTS AI 去對戰 10 局，其中 Pure MCTS

AI 最開始每下一步棋使用 1000 次模擬，當我們訓練的 AI 以 10:0 結束時，就把 Pure MCTS AI 每下一步棋模擬的次數加 1000 次，增加 Pure MCTS AI 的強度，就這樣依此類推，不斷增強。

### (8) 整體訓練流程

根據前面幾點所提的，self-play 所蒐集的 data 就是為了訓練策略價值網路，而評估完後，最優的 model 也會馬上被用在 MCTS 來進行 self-play，生成更好的 self-play data，就這樣兩者互相促進，構成訓練的循環。

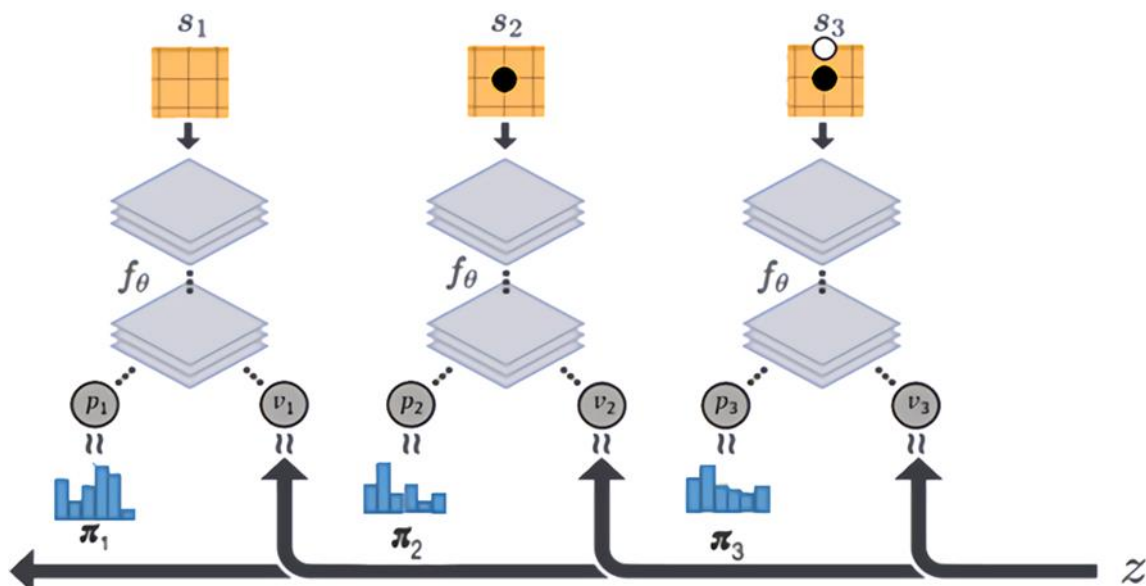


圖 6. 策略價值網路訓練架構圖[2]

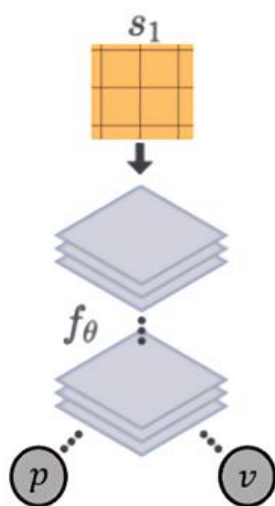


圖 7. 策略價值網路[2]

## 3. 實驗結果

### (1) 模型收斂討論

圖 8 展示的是一次在 9\*9 棋盤上進行五子棋訓練的過程中損失函數隨著 self-play 局數變化的情況，這次實驗一共進行了 3000 局自我對局，損失函數從最開始的 4 點多慢慢減小到了 2 點多左右。

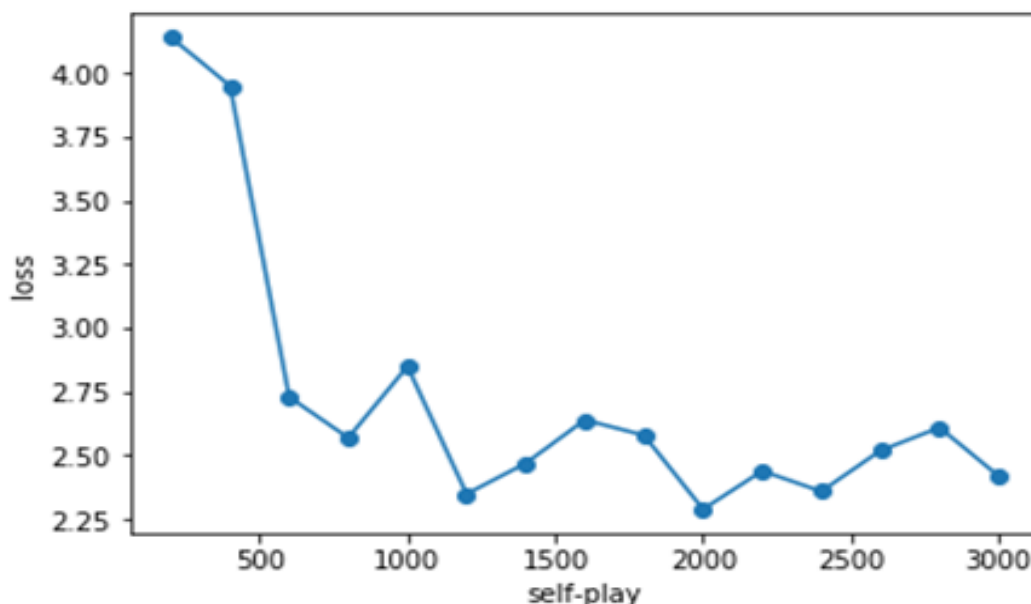


圖 8. 每 250 局的 loss 函數變化值

除了觀察到損失函數在慢慢減小，我們還觀察策略價值網絡輸出的策略（輸出的落子概率分佈）的 entropy 的變化情況，如圖 7。正常來講，最開始的時候，我們的策略價值網絡基本上是均勻的隨機輸出落子的概率，所以 entropy 會比較大。隨著訓練過程的慢慢推進，策略價值網絡會慢慢學會在不同的局面下哪些位置應該有更大的落子概率，也就是說落子概率的分佈不再均勻，會有比較強的偏向，這樣 entropy 就會變小。

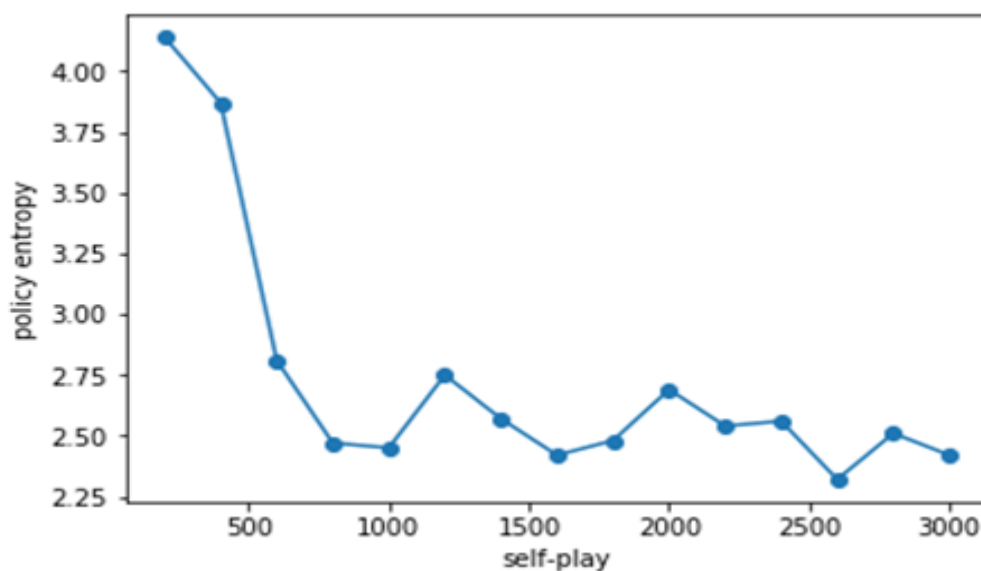


圖 9. 每 250 局 policy 的 entropy 變化

## (2) 不同大小棋盤的訓練結果探討

由圖 10 我們可以觀察到不同棋盤訓練的時間差異，在 13\*13 和 15\*15 的棋盤雖然訓練花的較長的時間，但所訓練出來的 model 卻沒有像 9\*9 棋盤所訓練出來的 model 完整，在我



們這次的訓練中我們發現，在相同條件下不同大小的棋盤所要花費的訓練次數是不一樣的，在越大的棋盤需要調整給它的訓練次數，這樣才能訓練出一個完整的 model。

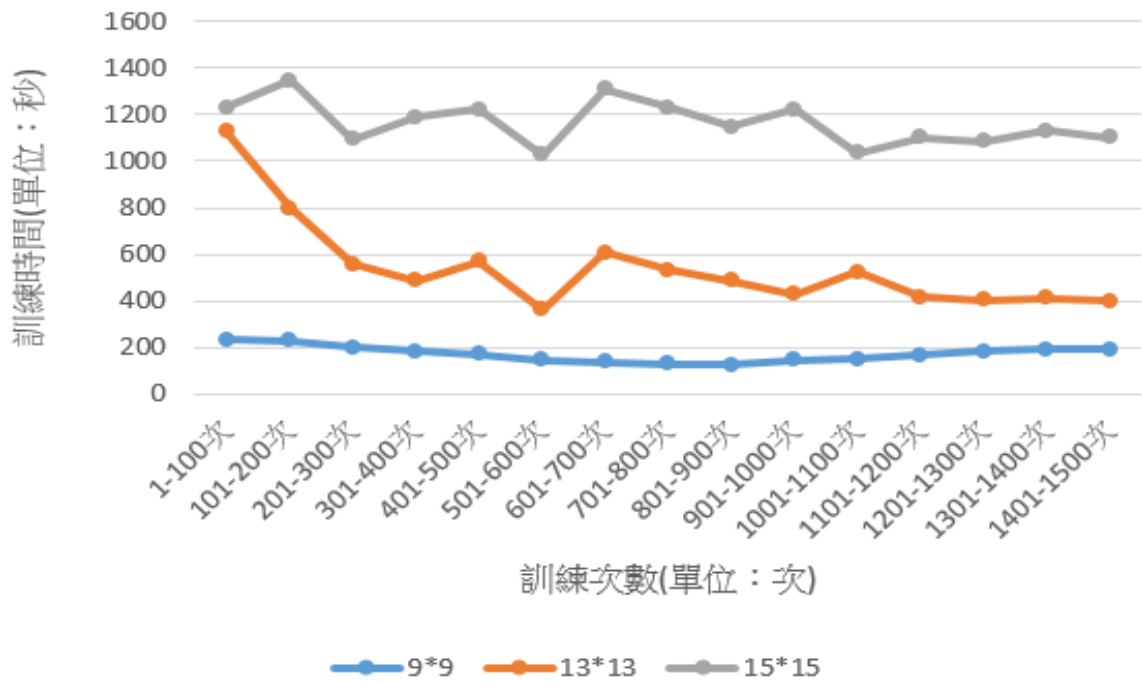


圖 10. GPU 自我訓練的時間

### (3) 策略價值網路訓練時間的差異

由於 GPU 比 CPU 擁有更多的 ALU(算術邏輯單元)，且在策略價值網路運用到大量的向量跟矩陣計算，正是 GPU 的強項，因此我們比較訓練策略價值網路時在 GPU 與 CPU 的耗費的時間，如圖 11 所示，實驗出來的 GPU 在訓練策略價值網路這部分程式的時間確實比 CPU 快上許多倍。

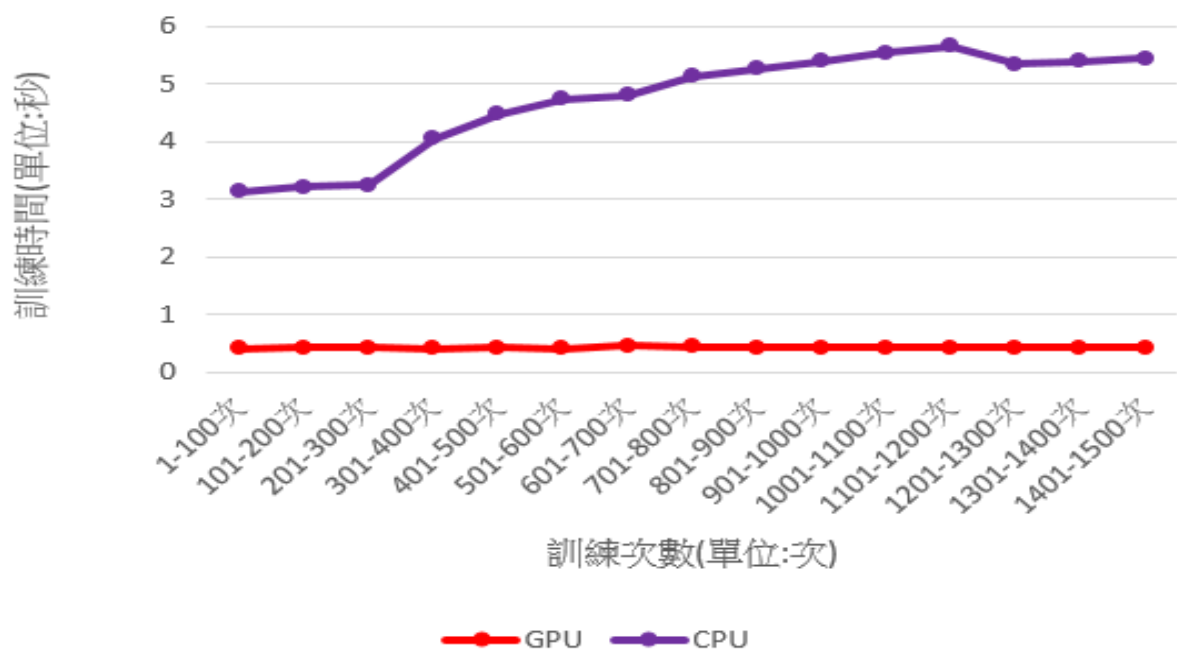


圖 11. 9\*9 棋盤策略價值網路執行時間

由圖 12 所示我們使用兩張不同等級的顯示卡下去做策略價值網路的訓練時間統計，從圖上可以很明顯的看出來 RTX-2060 版本的顯示卡速度快於 GTX-1060 版本的顯示卡，在執行策略價值網路所得到的結果中可以印證在做深度學習這方面需要有較好的硬體設備可以幫助我們省下很多的時間來做規模更大的訓練。

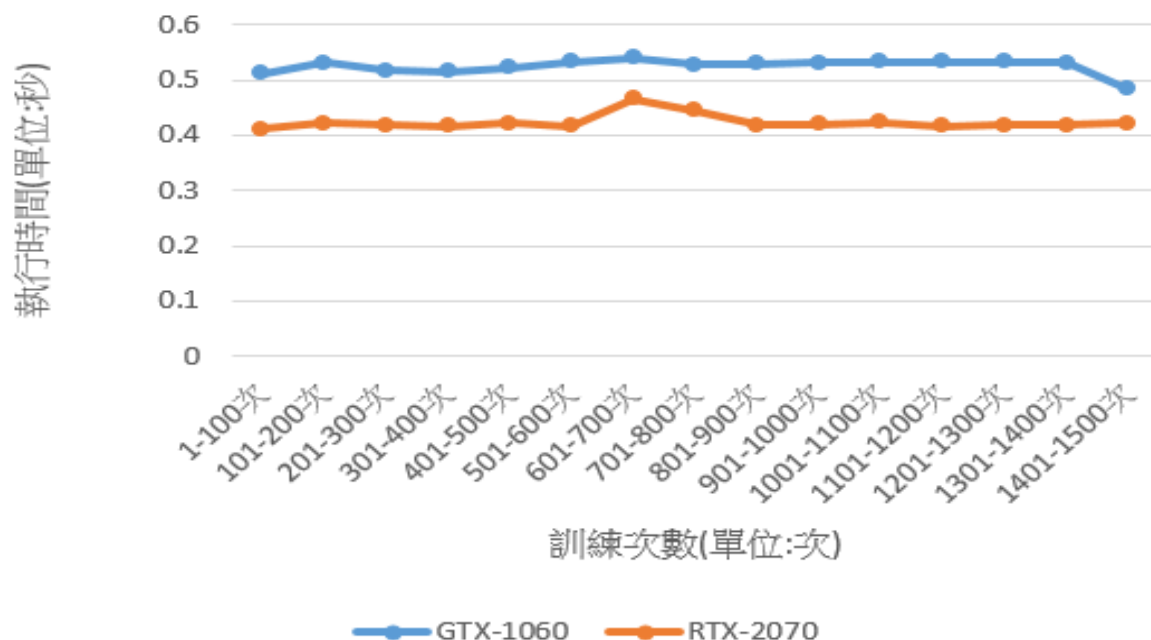


圖 12. GTX-1060/RTX-2070 執行策略價值網路時間

## ● Zenbo 機器人餐廳帶位點餐系統

圖 13 為餐廳帶位點餐與介紹系統的系統架構圖，帶位點餐這一端首先由機器人內建的麥克風感測顧客聲音，接著透過口語理解模組(SLU)將語音辨識資訊轉化為結構化的語意資料，並提供給適當的 Domain 處理為了處理語音辨識結果必須使用 DDE 來解決這些需求，進而提供開發者建立有用的語意資料。當語音辨識完成後 Appbuilder 來決定要進行點餐或者是帶位，點餐會依據開發者設定的菜單資料來呈現在 Zenbo face，帶位則依據桌號的路線資料進行帶位。餐廳介紹這個部分由開發者，進行設計，語音介紹的同時機器人將會有一系列的動作與表情與之配合。

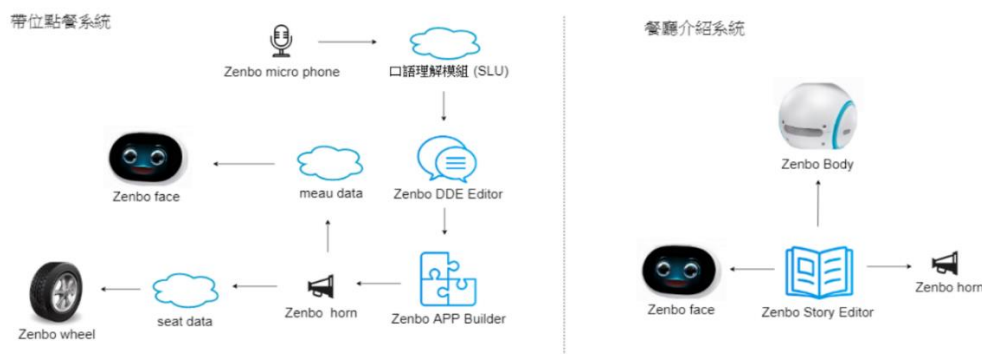


圖 13. 系統架構圖



## 1. 系統運作流程

圖 14 為帶位與點餐系統的流程圖從流程圖中我們可以看到顧客與機器人的對話主要有三個：1.詢問是否有位置 2.告知用餐人數 3.進行點餐帶位的功能首先需要由顧客先行詢問是否有位置，接著告知機器人用餐人數，機器人將依照開發者的設定進行判斷顧客所需要的人數是否還有空缺，有空缺的話將會引導顧客帶到特定人數的桌號，到定點後機器人會在原地待命等待顧客對機器人下點餐的命令，相反地沒有空缺的話機器人則會請顧客在外等候，過程中機器人所有的動作包含回應顧客的語句皆由動作編輯(Appbuilder)完成。

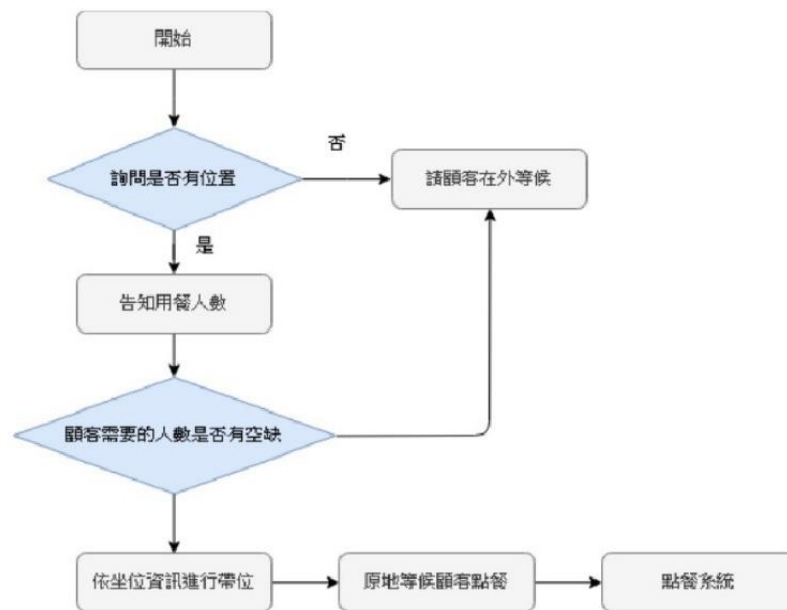


圖 14. 帶位點餐流程圖

## 2. 對話編輯器(DDE)設定

DDE 是一套線上編輯工具該平台又分為兩個組件，其中一個用於構建應用程序，另一個用於測試網站上測試，當使用者在網站上完成所期望的效果時便可以將它 publish 到 DDE 的線上資料庫以便與其他工具進行合作，主要用來設定機器人可以偵測到哪些語句。為了處理語音辨識結果，必須使用 DDE 來解決這些需求，進而提供開發者建立有用的語意資料。使用 DDE 之前需要先清楚一個概念人類與 Zenbo 的對話是透過一連串 Intent(意圖)語句組成。Concept(概念)則是語句的詞彙，而 Plan 則是負責串起語句之間的流程來建立一段完整對話。所以如何使用 DDE 來編輯 Plan / Intent / Concept 是重要的課題[5]。

從圖 15 來看，User 在問答中的語句為 Intent，Intent 中的詞彙 Concept。例如 I want to play the zoo game 的 zoo game 可替換成隸屬同個 Concept "Game" 的 card game 或者 board game，然後 Plan 將三段對話串連起來讓 Zenbo 跟使用者交談[8]。

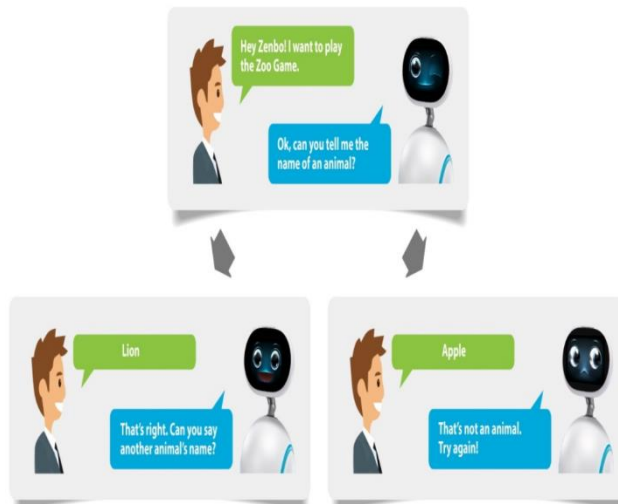


圖 15. Zenbo 對話示意圖

## (1) Plan

設定一情境，情境內只含有一 Intent，Plan 與 Plan 間透過 context 串接形成一個完整的對話，圖 16 以詢問是否有位置、告知用餐人數為例子。顧客會再詢問完位置後告知用餐人數，所以需要先執行 people\_ask\_seat 才會接著執行 Seat\_one，要注意的是後者的 input context 需要與前者的 output context 一致才是正確的執行流程。

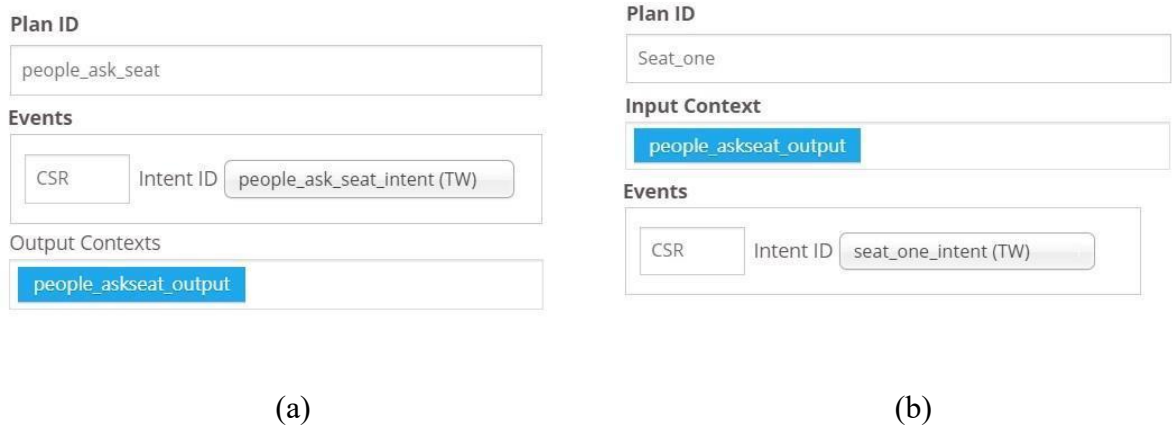


圖 16. Plan ID 設定(a) 顧客詢問座位(b) 顧客告知人數

## (2) Intent

包含情境內所有感測的語句(Sentence)。圖 17 中的 Sentence 皆是機器人可以辨識到的語句。

<p>Intent ID</p> <p>people_ask_seat_intent</p>	<p>Intent ID</p> <p>seat_one_intent</p>																		
<p>Sentences</p> <table border="1"> <tr><td>1</td><td>還有位置嗎</td></tr> <tr><td>2</td><td>請問還有位子嗎</td></tr> <tr><td>3</td><td>裡面有位置嗎</td></tr> <tr><td>4</td><td>裡面還有空位嗎</td></tr> <tr><td>5</td><td>請問裡面還有空位嗎</td></tr> </table>	1	還有位置嗎	2	請問還有位子嗎	3	裡面有位置嗎	4	裡面還有空位嗎	5	請問裡面還有空位嗎	<p>Sentences</p> <table border="1"> <tr><td>1</td><td>我一個人</td></tr> <tr><td>2</td><td>一位</td></tr> <tr><td>3</td><td>只有我一個</td></tr> <tr><td>4</td><td>只有一位</td></tr> </table>	1	我一個人	2	一位	3	只有我一個	4	只有一位
1	還有位置嗎																		
2	請問還有位子嗎																		
3	裡面有位置嗎																		
4	裡面還有空位嗎																		
5	請問裡面還有空位嗎																		
1	我一個人																		
2	一位																		
3	只有我一個																		
4	只有一位																		
(a)	(b)																		

圖 17. Sentence 設定 (a) 詢問座位 (b) 告知人數

### 3. 機器人動作(APP Builder)設定

使用積木來設計一套動作的組合讓 Zenbo 執行，我們將結合語音辨識的功能來為機器人設計符合當下情境的動作。

#### (1) when hear 積木

需要與 DDE 語句感測工具進行結合，當感測 DDE 內設定的語句，則執行 when hear 內的積木。圖 18 為感測到顧客詢問是否有位置後欲執行的事和感應到顧客告知機器人用餐人數後需要做的事，當機器人得知用餐人數後將會依照指定路線為顧客進行帶位，此處我們與 Casper 不同的地方在於我們的帶位方式，是先採用聲音感測的方式，使得系統更加人性化，再來我們將會為不同人數的桌號設置變數，如此一來才不會出現機器人帶客人到指定位置卻發現位置已經有人的窘境。

#### (2) Grid view 積木

Grid view 是一個選單開發工具，我們將它應用來製作菜單選單，如圖 19 顯示前菜的菜單，圖 20 為菜單的設計方式同樣以變數設定為基礎，等到選單結束後再以條件判斷式判斷變數值，並介紹所選擇餐點。設計主要分成以下元件：

- Text-to-speech：設定 Grid view 選單第一次出現時 zenbo 欲說的語句。
- Title：Grid View 選單上的標題。
- Event：Grid View 的選項

初始有四個選項每個選項，當某 Event 選項被觸發時則執行對應 when receive 積木。

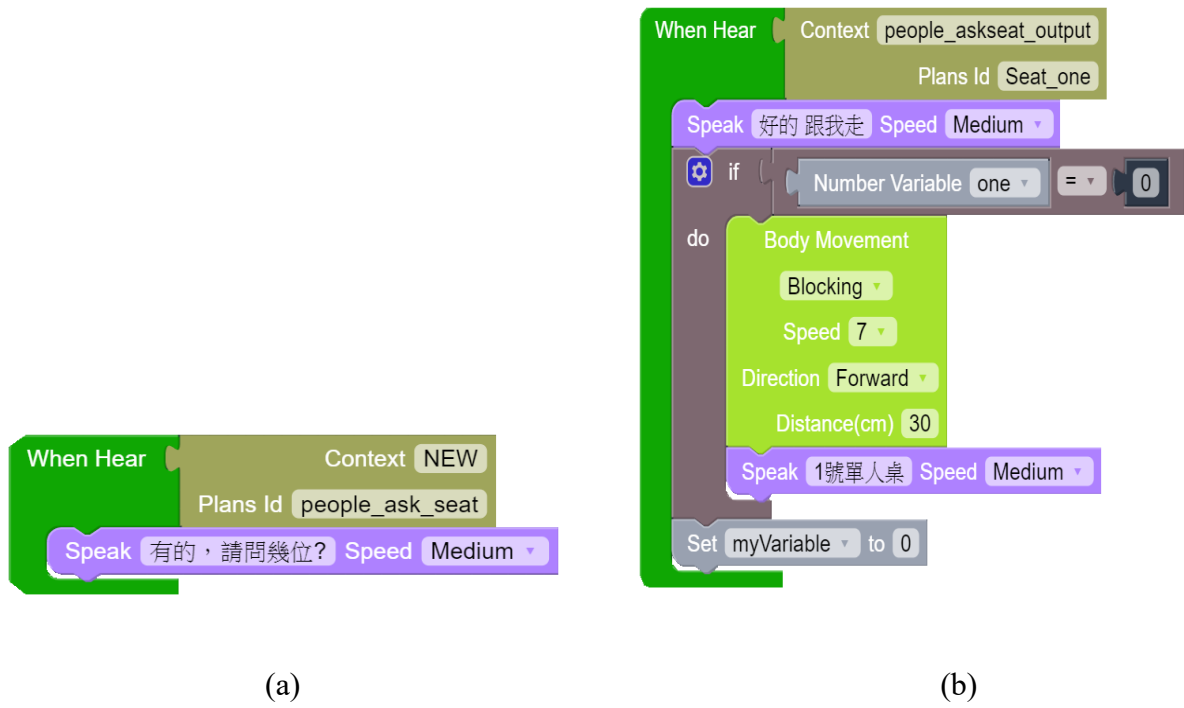


圖 18. When hear 積木設定 (a) 詢問座位 (b) 告知人數



圖 19. 前菜菜單顯示畫面

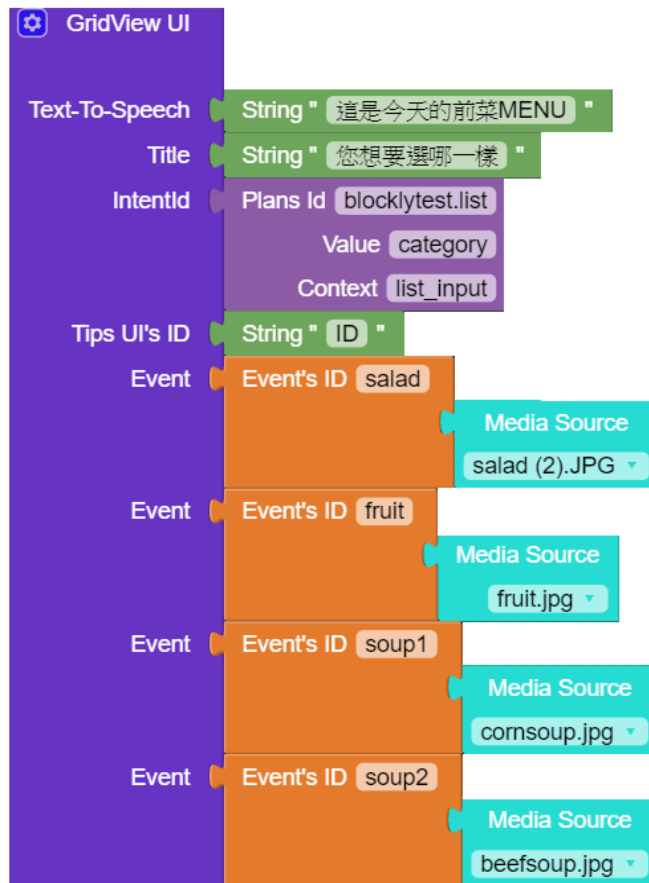


圖 20. Grid view 積木開發前菜菜單

### (3) When receive 積木

Event 被觸發時執行 when receive 積木內的動作，如圖 20 若選擇 event ID 為 salad 的選項，則執行圖 21(a)when receive 中 salad 積木下方的區塊中的變數，變數決定後，則會執行變數對應的下個動作，如圖 21(b)



(a)

(b)

圖 21. When receive 積木 (a)前菜選單變數設定 (b)選單變數選擇後的條件判斷

## ● 服務型機器人之自我決策模型

本計畫參考了 Castro-González [3] 結合了驅力(drives)、動機(motivation)，以及強化式學習，讓機器人在產生出主宰的動機後，利用強化式學習來學習動機。並導入了使用者意圖推論以及基於使用者意圖的決策方式的研究 [4]，同時將該研究成果與平衡驅動理論結合，將服務人類作為機器人內部需求的一環，再根據“強化式學習 (Reinforcement Learning)”在互動中學習滿足人類的意圖。圖 22 說明了自我決策系統整體架構，自我恆定參數包含了驅力、動機，人類參數則是使用者意圖，再加上環境參數，機器人能決策出該執行的動作，透過使用者的回饋，修正這三類參數的重要性，調整出更符合當時情境應該執行的服務。

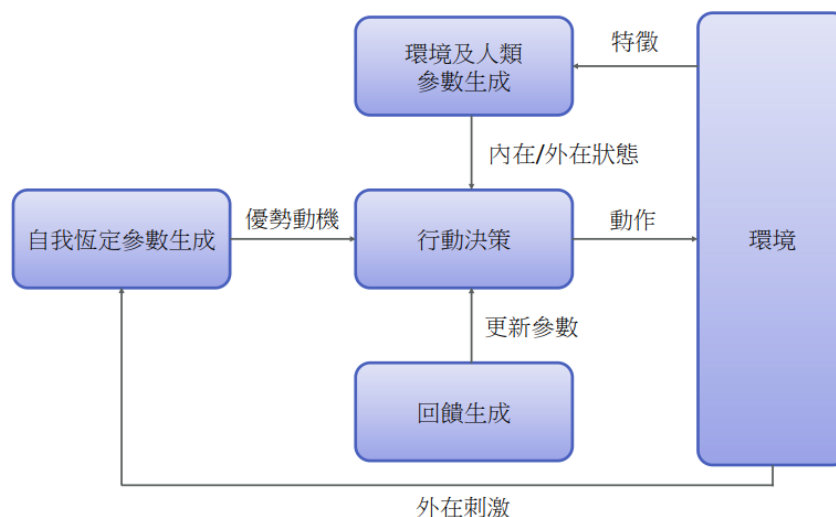


圖 22. 自我決策系統架構

本計畫為機器人設計了四種驅力:成就需求(NAch)、社交需求(NSoc)、休息需求(NRest)、充電需求(NEng)。表格 1 列出機器人可執行的服務與相對應動作。

表格 1 可執行的服務與對應的動作

Target	Actions
Charger	Move to charger, attach, detach, stay.
CD player	Play, stop.
Light	Turn on, turn off.
Other	Chat, rest, move to human, move to resting place

為了評估所提出的系統是否確實滿足了用戶和機器人本身，本計畫定義和使用了四個評量方式: 使用者滿意率(HSR)、累積的驅力平均(CDA)、累積超過的驅力平均(CEDA)、機器人安全使用率(RSR)。表格 2 顯示整體的機器人自主決策的表現，分別觀察他們的 CEDA，可以看出，所有驅力下超過驅動的平均值非常低。這表明機器人的動作成功地消除了內部需求，使其保持在門檻值之下，保持機器人本身的滿足。RSR 的值表明，機器人只花費 10% 的時間感到不安全，其

中大部分來自模擬的早期階段。HSR 的值顯示，當人類用戶表達意圖時，機器人將立即滿足 82 % 的時間。同樣，大多數的不滿意來自於模擬的早期階段，所以隨著迭代次數的增加，RSR 和 HSR 的值預計會增長得更大。

表格 2 預設參數之模擬結果

	<b>NAch</b>	<b>NSoc</b>	<b>NRes</b>	<b>NEng</b>	<b>All</b>
<b>CDA</b>	1.73*	43.5	40.1	44.5	32.5
<b>CEDA</b>	1.73*	2.05	5.72	6.99	4.12
<b>HSR</b>	-	-	-	-	0.82
<b>RSR</b>	-	-	-	-	0.90

## 五、 結果與討論

本計畫第一項完成的研究是用 Monte-Carlo tree search 搭配 Policy-Value Network 實現 Self-play reinforcement learning 方法，並應用在五子棋對弈上。本研究對於棋面邊角的位置較缺乏對抗性，也就是說玩家一直下邊角位置即會獲勝；這部分將於 Training 進行改進與修改，期盼能將此電腦訓練成：在棋面的任何角落，均能達到專業水準。一開始我們是使用 DQN 架構實作，但是其結果皆不理想，我們認為是提供 Training 的 data 太少，加上 Reward 設計無法達到完美，導致最終無法收斂。因此最終改成使用本發表之方法實作，此方法的成果較為理想。

本計畫第二項完成的研究使用語音辨識系統來增加機器人與顧客的互動，自顧客進入餐廳到點餐我們模擬了數個可能出現的情境，使得機器人感測到顧客聲音時得以做出適當的回應與動作，並且搭配餐廳介紹的功能，達到吸引顧客的目的。但是此系統仍有幾點需要改進，比如我們的帶位點餐系統的座位資料與菜單的資料，皆是由開發者直接注入程式區塊，這樣的作法會造成使用者的不便，當每次需要更新座位資料與菜單的資料時，必須直接修改程式碼，而無法透過更新資料庫，來達到同步的效果，另一點是我們使用動作編輯器所製作出來的積木所佔的容量太大，容易導致機器人無法負荷而導致閃退，主要原因是因為我們的積木群模組化的尚未完全，Zenbon 雖然有超音波測距感測器，但是由於技術上的困難，目前還無法使用此感測器進行路徑的校正，未來可以針對以上幾點進行系統的優化。

本計畫最後開發了機器人的自主行為之決策模組，另外亦利用影像及深度學習開發出人所處的情境(例如：可能從事的活動)之推論模組，以便機器人根據情境選出適合的服務方式與人互動(例如：閒聊)，未來可應用在商業服務型機器人平台上。

## 重要參考文獻

- [1] D. Silver *et al.*, "Mastering the game of Go with deep neural networks and tree search," *nature*, vol. 529, no. 7587, pp. 484-489, 2016.
- [2] D. Silver *et al.*, "Mastering the game of go without human knowledge," *nature*, vol. 550, no. 7676, pp. 354-359, 2017.
- [3] Á. Castro-González, M. Malfaz, and M. A. Salichs, "Selection of actions for an autonomous social robot," in *International Conference on Social Robotics*, 2010: Springer, pp. 110-119.
- [4] S. H. Tseng, J.-H. Hua, S.-P. Ma, and L.-e. Fu, "Human awareness based robot performance learning in a social environment," in *2013 IEEE International Conference on Robotics and Automation*, 2013: IEEE, pp. 4291-4296.
- [5] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. "A survey of socially interactive robots". *Robotics and Autonomous Systems*, 42(3-4):143–166, 2003.
- [6] Aaron Steinfeld, Terrence Fong, David Kaber, Michael Lewis, Jean Scholtz, Alan Schultz, and Michael Goodrich. "Common metrics for human-robot interaction". In 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, volume 30, pages 33–40, 2006.
- [7] N. Mitsunaga, Z. Miyashita, K. Shinozawa, T. Miyashita, H. Ishiguro, and N. Hagita. "What makes people accept a robot in a social environment - discussion from six-week study in an office". *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3336–3343, 2008.
- [8] Chao Shi, Michihiro Shimada, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. "Spatial Formation Model for Initiating Conversation". In *Proceedings of Robotics: Science and Systems*, Los Angeles, CA, USA, June 2011.
- [9] S. Satake, T. Kanda, D.F. Glas, M. Imai, H. Ishiguro, and N. Hagita. "How to approach humans?-strategies for social robots to initiate interaction". In *4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 109–116, March 2009.
- [10] M.L. Walters, K. Dautenhahn, K.L. Koay, C. Kaouri, R. Boekhorst, C. Nehaniv, I. Werry, and D. Lee. "Close encounters: spatial distances between people and a robot of mechanistic appearance". In *5th IEEE-RAS International Conference on Humanoid Robots*, pages 450–455, 2005.
- [11] Sang-won Leigh and Pattie Maes. 2016. Body Integrated Programmable Joints Interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 6053-6057.
- [12] Sak, Haşim, Andrew Senior, and Françoise Beaufays. "Long short-term memory recurrent neural network architectures for large scale acoustic modeling," *Fifteenth annual conference of the international speech communication association*, 2014.
- [13] Liu, Wei, et al. "SSD: Single shot multibox detector." *European Conference on Computer Vision*. Springer International Publishing, 2016.
- [14] Lin, Tsung-Yi, et al. "Microsoft coco: Common objects in context." *European Conference on Computer Vision*. Springer International Publishing, 2014.
- [15] Zhou, Bolei, et al. "Places: An image database for deep scene understanding." *arXiv preprint arXiv:1610.02055* (2016).
- [16] Ahmad, Kashif, et al. "USED: a large-scale social event detection dataset." *Proceedings of the 7th International Conference on Multimedia Systems*. ACM, 2016.



108年度專題研究計畫成果彙整表

計畫主持人：曾士桓			計畫編號：108-2218-E-992-310-		
計畫名稱：機器學習輔助人機互動以提升服務型機器人之社交智慧					
成果項目			量化	單位	質化 (說明：各成果項目請附佐證資料或細項說明，如期刊名稱、年份、卷期、起訖頁數、證號...等)
國內	學術性論文	期刊論文	0	篇	李佳陽、王泰淞、張騏岳、江尚紘、曾士桓(2019年06月)。Gomoku Deep。數位生活科技研討會。科技部：108-2218-E-992-310。本人為通訊作者
		研討會論文	1		
		專書	0	本	
		專書論文	0	章	
		技術報告	0	篇	
		其他	0	篇	
		國外	學術性論文	期刊論文	0
研討會論文	1				
專書	0			本	
專書論文	0			章	
技術報告	0			篇	
其他	0			篇	
參與計畫人力	本國籍			大專生	8
		碩士生	2	尤翔鵬、孫尉豪	
		博士生	0		
		博士級研究人員	0		
		專任人員	0		
	非本國籍	大專生	0		
		碩士生	0		
		博士生	0		
		博士級研究人員	0		
		專任人員	0		
	其他成果 (無法以量化表達之成果如辦理學術活動、獲得獎項、重要國際合作、研究成果國				

際影響力及其他協助產業技術發展之具體 效益事項等，請以文字敘述填列。）	
--	--