**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）**

**基於Zero Learning Polygames深度增強式學習框架**

**開發各類型電腦對局程式**

計畫類別：■個別型計畫　　□整合型計畫  
計畫編號： MOST 110-2221-E-259 -007 -MY3  
執行期間： 110年08月01日至　111年07月31日

執行機構及系所：國立東華大學資訊工程學系

計畫主持人：顏士淨  
計畫參與人員：邱顯棟、邱創業、林駿程、Tatchakorn Saibunjom、Musa Mukhammed

期末報告處理方式：

1. 公開方式：

■非列管計畫亦不具下列情形，立即公開查詢

□涉及專利或其他智慧財產權，□一年□二年後可公開查詢

2.「本研究」是否已有嚴重損及公共利益之發現：■否 □是

3.「本報告」是否建議提供政府單位施政參考 ■否 □是， （請列舉提供之單位；本部不經審議，依勾選逕予轉送）

中 華 民 國 111 年 5 月 31 日

**計畫摘要**

2016年Google DeepMind的AlphaGo擊敗韓國職業棋士，震驚全世界。Google之後不斷研發相關技術，其中Zero Learning系列，包括AlphaGo Zero，Alpha Zero 與MuZero 更是將此技術推廣到其他棋類與遊戲，不再局限於圍棋上面。以上技術主要包括了三種機器學習的技術：蒙地卡羅樹搜尋(Monte Carlo Tree Search)、深度學習(Deep Learning)、和增強式學習(Reinforcement Learning)。相關技術不但可以應用在遊戲產業，也可增加一般資訊產業的競爭力，這也是Google不斷發展這些技術的主要原因。

本計畫研究團隊在2019年開始與Facebook合作開發Polygames，Polygames是一個通用型的Zero Learning增強式學習的框架，可用於各類型對局遊戲，初步成果已經在2020年初公開於GitHub，頗受好評，我們也在2020年開始與Facebook共同發表論文和參加比賽。本計畫將延續這些研究，繼續與Facebook研究改進Polygames。使用Zero增強式學習需要大量運算資源，因此本計劃目標將放在如何以較少資源快速訓練出特定類型遊戲，概念是針對特定類型遊戲的特性改良相關演算法，以及利用遷移學習加快訓練。第一年將研究連線型遊戲的改良，第二年則是針對圍地型的遊戲，第三年則是對可以死而復活的日本將棋遊戲的研究。同時我們也將研究這些類型遊戲的隨機版本。本計劃的成果將可幫助快速開發上述三種類型的遊戲，也可幫助台灣產業界或教育界設計並驗證新的遊戲。期待未來三年持續與國際一流的Facebook AI團隊繼續合作發表論文與參加比賽，在這兩個項目維持目前的好成績，繼續在國際上這個領域佔有一席之地。

在第一年的計畫中，我們發表了兩篇期刊論文，三篇研討會論文，並且在ICGA國際電腦對局競賽獲得了一金二銀的成績，也通過了對弈機器人的專利，並且也舉辦了TCGA台灣電腦學會的對局競賽，對於電腦對局的發展有很大的貢獻，達到了預期的目標。

**關鍵字：蒙地卡羅樹搜尋、深度學習、增強式學習、Zero Learning、Polygames、對局遊戲、遊戲**

**英文摘要**

In 2016, Google DeepMind's AlphaGo defeated Korean professional chess players, shocking the world. Google continued to develop related technologies afterwards. Among them, the Zero Learning series, including AlphaGo Zero, Alpha Zero and MuZero, promoted this technology to other chess and games, not limited to Go.

The above technologies mainly include three types of machine learning technologies: Monte Carlo Tree Search, Deep Learning, and Reinforcement Learning. Related technologies can not only be applied to the game industry, but also increase the competitiveness of the general information industry. This is the main reason why Google continues to develop these technologies. The research team of this project began to cooperate with Facebook to develop Polygames in 2019. Polygames is a general-purpose Zero Learning enhanced learning framework that can be used in various types of game games. The preliminary results have been published on GitHub in early 2020 and are well received. , We will also publish papers and participate in competitions with Facebook in 2020. This project will continue these studies and continue to study and improve Polygames with Facebook.

Using Zero Learning requires a lot of computing resources. Therefore, the goal of this plan is to quickly train specific types of games with fewer resources. The concept is to improve related algorithms for the characteristics of specific types of games and use transfer learning to speed up training. The first year will study the improvement of online games, the second year will focus on land-based games, and the third year will focus on the study of Japanese shogi games that can be resurrected from the dead. We will also study random versions of these types of games. The results of this project will help the rapid development of the above three types of games, and also help Taiwan’s industry or education to design and verify new games. I look forward to continuing to cooperate with the world-class Facebook AI team in the next three years to publish papers and participate in competitions, maintain the current good results in these two projects, and continue to have a place in this field internationally.

In the first year of the program, we published two journal papers, three seminar papers, won one gold and two silvers in the ICGA international computer game competition, passed the patent of the game robot, and also The game competition of TCGA Taiwan Computer Society was held, which made a great contribution to the development of computer game and achieved the expected goal.

**Keywords: Monte Carlo Tree Search, Deep Learning, Reinforcement Learning, Zero Learning, Polygames, Computer Games, Game**

**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）**

**基於Zero Learning Polygames深度增強式學習框架**

**開發各類型電腦對局程式**

**簡介**

本計劃第一年將研究連線型遊戲的改良，包括Othello, Connect6, Hex等連線型遊戲，Othello與Hex又可分為不同棋盤大小的版本。第一年的遊戲勝負判斷較為單純，成果可做為第二年的基礎。在第一年的計畫中，我們發表了兩篇期刊論文，三篇研討會論文，並且在ICGA國際電腦對局競賽獲得了一金二銀的成績，也通過了對弈機器人的專利，並且也舉辦了TCGA台灣電腦學會的對局競賽，對於電腦對局的發展有很大的貢獻，達到了預期的目標。

# 一、前言

本計畫將基於Zero Learning Polygames深度增強式學習框架開發各類型電腦對局程式，將延續本團隊在2019年與Facebook合作開發Zero Learning Polygames的成果，Polygames是一個我們與Facebook合作開發研發的通用型的Zero Learning增強式學習的框架，可用於各類型對局遊戲，初步成果已經在2020年初公開於GitHub，頗受好評，我們也在2020年開始與Facebook共同發表論文和參加比賽，目前已經有相當不錯的合作成果，更獲得了圖靈獎Yann Lecun的肯定。本計畫將延續這些研究，繼續與Facebook研究改進Polygames。使用Zero Learning增強式學習需要大量運算資源，因此本計劃目標將放在如何以較少資源快速訓練出特定類型遊戲，概念是針對特定類型遊戲的特性改良相關演算法，以及利用遷移學習加快訓練。

# 二、研究目的

本計劃第一年將研究連線型遊戲的改良，包括Othello, Connect6, Hex等連線型遊戲，Othello與Hex又可分為不同棋盤大小的版本。第一年的遊戲勝負判斷較為單純，成果可做為第二年的基礎。第二年則是針對圍地型的遊戲，包括九路圍棋、版塊圍棋、幻象圍棋等。圍地的計算較為抽象，幻象圍棋則是隨機性很高的遊戲。第三年則是對可以死而復活的日本將棋遊戲的研究，由於將棋有小盤面的版本，比較適合研究。將棋是一種走子棋類的遊戲，而且棋子可以死而復生，困難度又更高，我們將研究迷你將棋、京都將棋，同時也研究隨機性高的骰子將棋。本計劃的成果可幫助快速開發特定類型的遊戲，也可幫助台灣產業界或教育界設計新遊戲。期待未來三年能夠持續與國際一流的Facebook AI團隊繼續合作研究Polygames，共同發表論文與參加比賽，並培養國際性的AI人才。

# 三、研究方法

Zero Learning的MCTS架構

Zero Learning在訓練與對弈時的MCTS方法流程如圖1，分為以下四個階段:

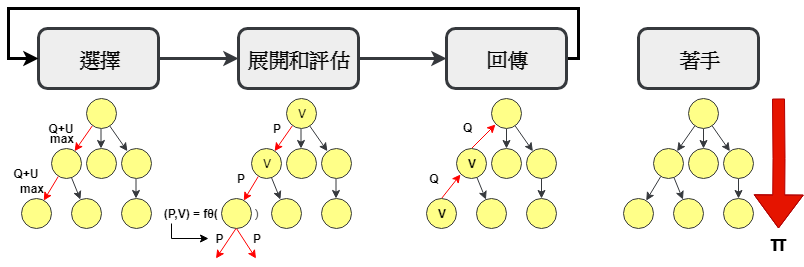


圖 1 Zero Learning MCTS方法流程圖

* 選擇(Selection):

在選擇(Selection)階段，選擇擁有最大的節點做為走子的節點，並透過，Q的計算公式在回傳階段將一起討論，而U的計算公式如下。公式中的為一個決定探索率的常數，可依照需求調整，而P為在環境s狀態下採取動作a的機率，N則為節點被拜訪過的次數。

* 擴展(Expand and evaluate):

當選擇節點選擇到了葉節點後，不知道接下來有什麼步可以走，就需要用到擴展(Expand)，透過將葉節點的所有合法著手展開，並利用類神經網路評估每種著手機率P勝率V如圖2。

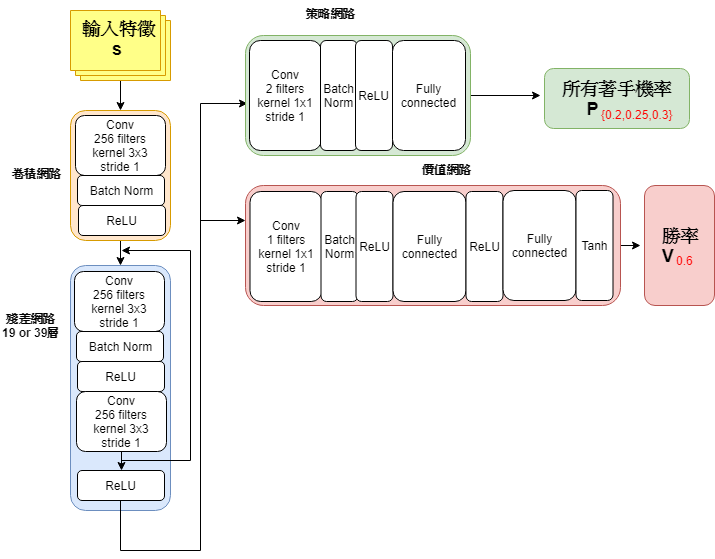


圖 2 Zero Learning神經網路評估示意圖

接著初始化展開的所有節點中的參數如下:

初始化完後將每一個著手t對應的節點中的拜訪次數增加:，並將動作價值更新為平均值如下，至此擴展階段就結束。

* 回傳(Backup)

當擴展結束，將神經網路評估的勝率V回傳至上層節點，若擴展出的下一個局面棋局結束則返回勝負的值。

* 著手(Play)

前面三個階段，通常是是1600次模擬，也就等於一至三步驟重複執行1600次才決定下一手怎麼走，透過為數這麼多的模擬次數後，每個局面該選擇什麼樣的走法、得到什麼樣的局面和勝率，基本上都能在搜尋樹上找到。在狀態選擇動作a的策略可以表示如下:

其中為控制探索水平的溫度常數，如果的值為0，代表著手階段會選擇最多拜訪次數的節點作為著手。當棋局結束，保存棋局中所有狀態s和每一手策略並更新神經網路之參數。

Zero Learning中的特徵提取

在一般應用深度卷積神經網路增強棋類的程式中，會利用人類知識如：棋型、棋子價值…等。而在Zero Learning中通常僅標示棋子的位子而不告訴神經網路。以圍棋為例，可以增加8步歷史步，故總共為17個特徵平面，16個P1、P2擁有的棋子之歷史盤面特徵，1個代表當前局面顏色特徵。應用在不同棋類的時候可以利用以下公式快速換算出棋盤總共需要多少特徵平面，其中M代表有多少種棋子，T代表歷史盤面，L代表特殊規則和顏色的標示，N則代表棋盤大小。

以迷你將棋為例，由於迷你將棋的棋盤大小為，我們將盤面轉換為45個二元特徵如圖3，每種特徵的大小皆為，分別代表了雙方棋子的特徵45種(包含升變與被吃掉的)，代表顏色的特徵1種，代表重複盤面的特徵1種。

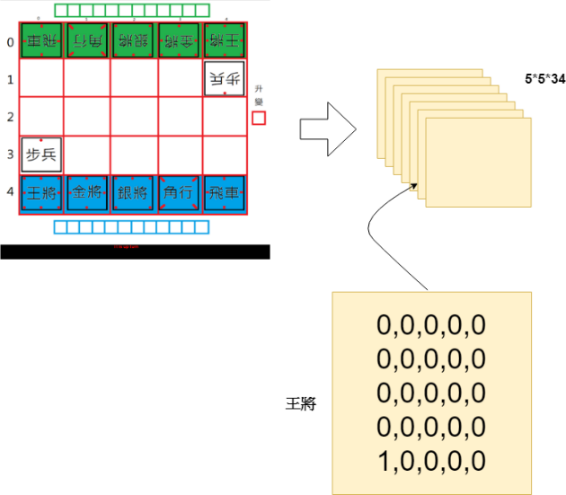


圖 3迷你將棋特徵提取示意圖

Zero Learning神經網路模型的訓練

神經網路模型的訓練流程如圖4，透過自我對弈(Self play)來產生訓練棋譜，並在每512步後進行參數更新，對於某些類型的遊戲來說，要如何判斷遊戲結束是個問題，也會影響遊戲的品質。例如走子型遊戲，必須在自我對弈階段對遊戲規則增加了當連續N步沒有吃子動作產生則判定平手的特殊規則，讓模組更加具有攻擊性而不是選擇沒有意義的動作。

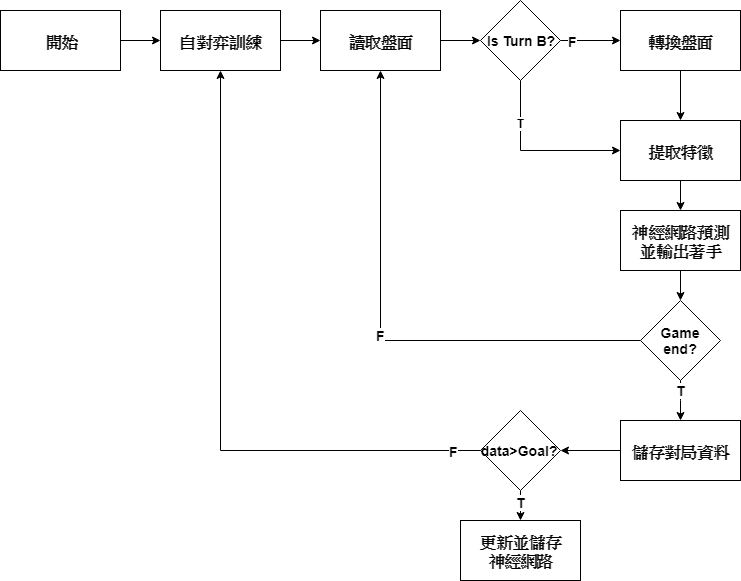


圖 4 神經網路訓練流程圖

Polygames比賽程式的設計

Polygames主要是訓練類神經網路model，網路model訓練好之後，需要一個可以比賽的系統搭配，圖5為主程式之流程，使用者可以選擇遊玩雙人模式或對抗電腦，對抗電腦則可以選擇先手或後手，且可選擇對抗純MCTS方法或本計劃實踐之程式，如此可以測試程式。且為了讓使用者有良好的體驗，必須以圖形化介面呈現。

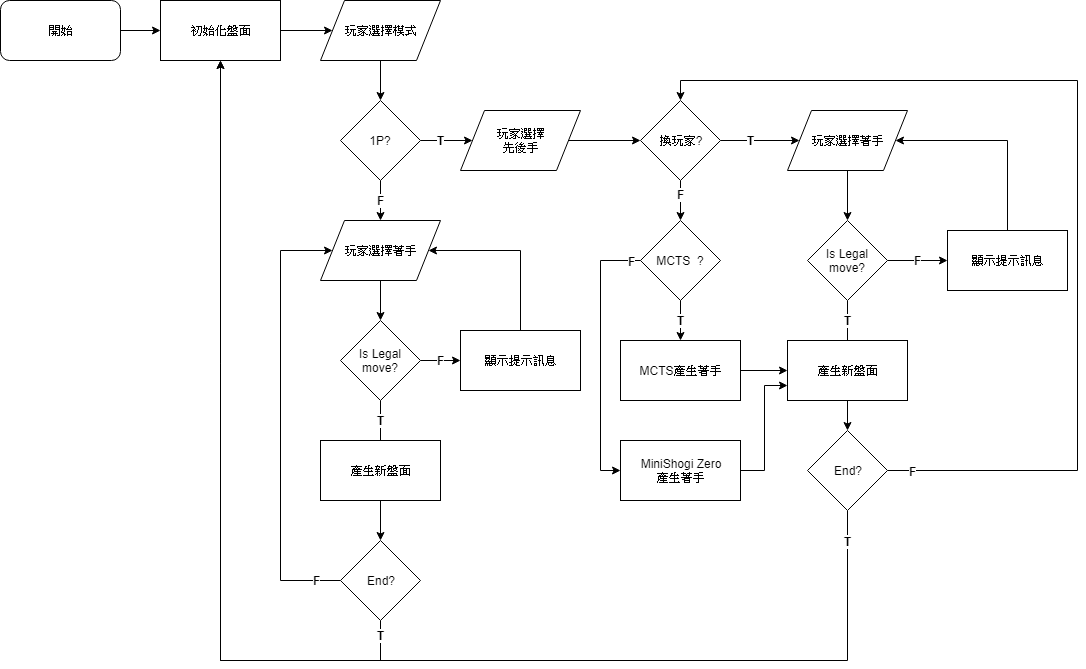


圖 5比賽程式流程圖

Polygames圖形化介面設計

Polygames目前只有簡單的文字介面，本計畫中，也將為各類型遊戲設計圖形化介面。以將棋為例，由於傳統將棋遊戲單看棋子比較難以記得棋子對應的走法。所以必須設計圖形化的的遊戲介面，例如迷你將棋可以設計如圖6，將每種棋子轉換為方型，並在棋身可以走的對應方位標示紅點或線，點代表的是僅能走一格，而線代表的是最遠可移動至棋盤邊界(如果沒有被阻擋)，右邊升變下的方格則是當棋子需要升變時可以核取的方塊，下方的黑色線條中則是負責顯示各種提示訊息，上下皆有的空格區域則負責顯示雙方吃掉的棋子。

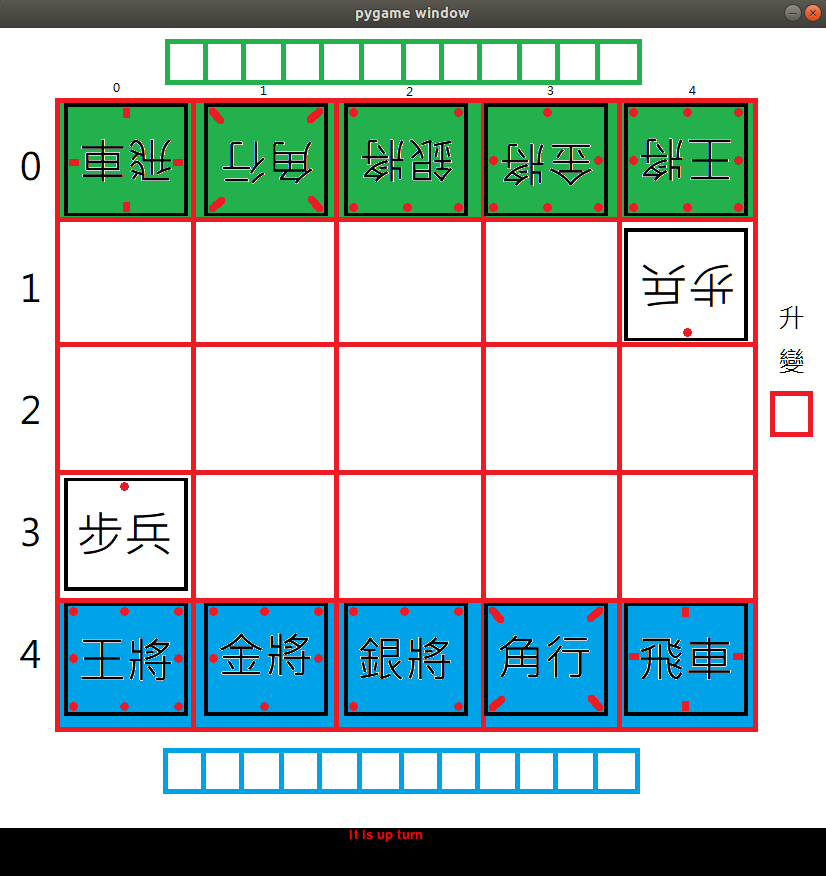


圖 6迷你將棋圖形化介面

# 四、研究結果

第一年目標為Polygames連線型類型遊戲的改良，第一年的各個項目內容如下：

* Othello, Connect6, Hex圖形化介面與測試工具的改良
* 三種連線型遊戲程式基本功能，包括開局殘局特殊規則判斷的效率的改進
* 基於Polygames的Neuroplasticity設計包括Policy與Value的神經網路與測試
* 研究適用於連線型遊戲的比賽模式(Tournament mode)
* 研究不同棋盤size的Network之遷移學習(Transfer learning)
* 研究三種遊戲間的遷移學習
* 研究修改連線型遊戲規則，使用類似Network與遷移學習的可行性
* 參加各種電腦對局比賽，測試成果

# 五、計畫期中成果自評

在第一年的計畫中，我們發表了兩篇期刊論文，三篇研討會論文，並且在ICGA國際電腦對局競賽獲得了一金二銀的成績，也通過了對弈機器人的專利，並且也舉辦了TCGA台灣電腦學會的對局競賽，對於電腦對局的發展有很大的貢獻，達到了預期的目標。

本計畫發表的期刊論文及會議論文

已發表期刊論文

1. S. Kavak, X. -D. Chiu, S. -J. Yen and M. Y. -C. Chen, "Application of CNN for Detection and Localization of STEMI Using 12-Lead ECG Images," in IEEE Access, vol. 10, pp. 38923-38930, 2022, doi: 10.1109/ACCESS.2022.3165966. (SCI, IF=3.745, 主持人為通訊作者)

2. Cheng-Chan Yang, \*Shi-Jim Yen, Xing-Dong Chiu, Kuo-Chu Wu, Shih-Cheng Ye, San-Hua Su, and Hsiao-Yi Huang, “Decision Tree Based Body Constitution Diagnosis System for Traditional Chinese Medicine,” Evidence-Based Complementary and Alternative Medicine, vol. 2022(9654):1-10, Article ID 5560087, 10 pages, 2022. https://doi.org/10.1155/2022/5560087. (SCI, IF=2.629, 主持人為通訊作者)

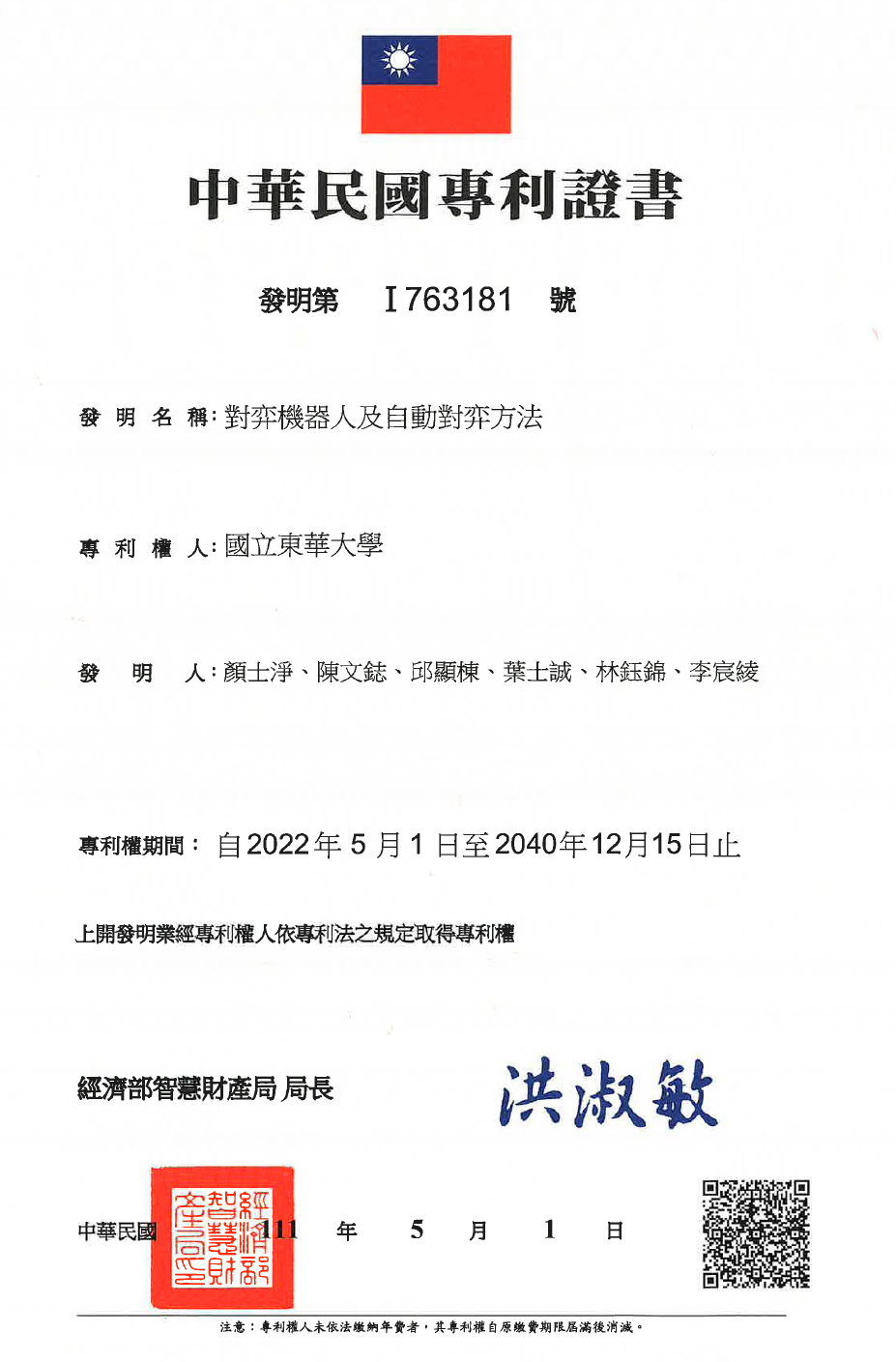
已發表會議論文

1. 王泰翔、蕭子渝、林駿丞、顏士淨、陳志昌、"基於深度學習開發十路黑白棋"，第十九屆離島資訊技術與應用研討會，會議日期：2022年5月27日～29日，會議地點：國立澎湖大學。

2. 邱顯棟、陳冠銘、黃議承、江承澔、洪三峰、顏士淨、"電子棒球好球帶應用"，第十九屆離島資訊技術與應用研討會，會議日期：2022年5月27日～29日，會議地點：國立澎湖大學。

3. 莊惟程，周信宏，顏士淨，電腦黑白棋對奕平台設計與開發，TCGA2022，5月13日，暨南大學，南投。

已獲得專利



期中報告

2022/5/13 於國立暨南大學TCGA電腦對局會議發表期中報告



重要參考文獻

1. Ashton Anderson, Jon Kleinberg, and Sendhil Mullainathan. Assessing human error against a benchmark of perfection. ACM Transactions on Knowledge Discovery from Data (TKDD), 11(4):45, 2017.
2. Tamal Biswas and Kenneth W Regan. Measuring level-k reasoning, satisficing, and human error in game-play data. In 2015 IEEE 14th International Conference on Machine Learning and Applications, pages 941–947, Miami, FL, 2015. IEEE.
3. Neil Charness. The impact of chess research on cognitive science. Psychological research, 54(1):4–9, 1992.
4. Reid McIlroy-Young, Siddhartha Sen, Jon Kleinberg, and Ashton Anderson. Aligning superhuman ai with human behavior: Chess as a model system. In Proceedings of the 25th ACM SIGKDD international conference on Knowledge discovery and data mining, 2020.
5. Hanan Rosemarin and Ariel Rosenfeld. Playing chess at a human desired level and style. In Proceedings of the 7th International Conference on Human-Agent Interaction, pages 76–80, 2019.
6. Reid McIlroy-Young, Russell Wang, Siddhartha Sen, Jon Kleinberg, and Ashton Anderson. Learning personalized models of human behavior in chess. arXiv preprint arXiv:2008.10086, 2020.
7. Chris J Maddison, Aja Huang, Ilya Sutskever, and David Silver, “Move evaluation in go using deep convolutional neural networks”, In International Conference on Learning Representations, 2015.
8. David Silver, Aja Huang, Christopher J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel and Demis Hassabis, ”Mastering the game of Go with deep neural networks and tree search”, Nature, 2016, Pages 484-503, Vol. 529.
9. Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., et al. Mastering chess and Shogi by self-play with a general reinforcement learning algorithm. arXiv preprint arXiv:1712.01815, 2017
10. Silver, D. et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science 362, 1140–1144 (2018).
11. Olivier Teytaud, Tristan Cazenave, Yen-Chi Chen, Guan-Wei Chen, Shi-Yu Chen, Xian-Dong Chiu, Julien Dehos, Maria Elsa, Qucheng Gong, Hengyuan Hu, Vasil Khalidov, Cheng-Ling Li, Hins-I Lin, Yu-Jin Lin, Xavier martinet, Vegard Mella, Jeremy Rapin, Baptiste Roziere, Gabriel Synnaeve, Fabien Teytaud, Shi-Cheng Ye, Yi-Jun Ye, Shi-Jim Yen, Sergey Zagoruyko. “Polygames: Improved Zero Learning,” ICGA Journal, vol. 43, no.3, pp. 1-13, 2020.
12. Ching-Nung Lin, Jr-Chang Chen and Shi-Jim Yen\*, “ Deep Learning Competition Framework on Othello for Education.,” IEEE Transactions on Games, Vol.11, No.3, pp.300-304, September 2019.
13. Reid McIlroy-Young, et.al., "Detecting Individual Decision-Making Style: Exploring Behavioral Stylometry in Chess,"Advances in Neural Information Processing Systems 34 (NeurIPS 2021).
14. atthew Ciolino, David Noever, Josh Kalin, "The Go Transformer: Natural Language Modeling for Game Play", arXiv:2007.03500, 2020.
15. T. Cazenave, "Residual Networks for Computer Go," in IEEE Transactions on Games, vol. 10, no. 1, pp. 107-110, March 2018, doi: 10.1109/TCIAIG.2017.2681042.

附件二

**科技部補助專題研究計畫成果報告自評表**

|  |
| --- |
| 請就研究內容與原計畫相符程度、達成預期目標情況、研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性）、是否適合在學術期刊發表或申請專利、主要發現（簡要敘述成果是否有嚴重損及公共利益之發現）或其他有關價值等，作一綜合評估。 |
| 1. 請就研究內容與原計畫相符程度、達成預期目標情況作一綜合評估    * + ▓達成目標 |
| 1. 研究成果在學術期刊發表或申請專利等情形：   論文：▓已發表 □未發表之文稿 □撰寫中 □無  專利：▓已獲得 □申請中 □無  技轉：□已技轉 □洽談中 ▓無  其他：（以100字為限）無 |
| 1. 請依學術成就、技術創新、社會影響等方面，評估研究成果之學術或應用價值（簡要敘述成果所代表之意義、價值、影響或進一步發展之可能性），如已有嚴重損及公共利益之發現，請簡述可能損及之相關程度（以500字為限）   本計畫成果方面，目前我們已經發表5篇論文。系統實作比賽方面，在2021 ICGA國際電腦賽局競賽中，共獲得1金1銀，金牌的項目為象棋，七月初也即將參加2022 ICGA電腦對局賽，預計將有好的成績表現。綜觀上述，本子計劃第一年將可達到預期的進度與成果。 |