第十九屆旺宏科學獎

創意說明書

參賽編號：

作品名稱：

姓名：吳邦寧

關鍵字：基因算法、強化學習、機器學習

摘要

前人使用Genetic algorithm來訓練神經網路玩UNO，在淺層的神經網路下十分有效，但是在相對深層的神經網路下成效不彰。

細菌使用基因來記憶所學所能，因此，只能完成相對簡單的任務；而人類使用大腦來記憶所學所能，因此，人類能完成相對複雜的任務。受此啟發，何不捨棄基因算法，使用Deep-Q-Learning來設計神經網路呢？

本研究使用四種不同的算法，Deep-Q-Learning algorithm、Genetic algorithm、Naïve offensive/defensive algorithm以及Dummy algorithm做為實驗組以及對照組。

Deep-Q-Learning algorithm對上Dummy algorithm約有的勝率，對上Naïve offensive/defensive algorithm約有/的勝率，對上另外一個Deep-Q-Network的時候，先手約有的勝率，後手約有的勝率。

Genetic algorithm對上Dummy algorithm約有的勝率，對上Naïve offensive/defensive algorithm約有/的勝率，對上另外一個Genetic Network的時候，先手約有的勝率，後手約有的勝率。

Deep-Q-Learning algorithm對上Genetic algorithm，在先手的情況下約有的勝率，在後手的情況下約有的勝率，就算關閉了Q-Learning的學習功能，Genetic algorithm也難以望其項背，可見Q-Learning是較為優秀的演算法。

根據實驗數據得知，使用Deep-Q-Learning能夠有效的增加勝率，印證了「人類比細菌更能處理複雜問題」的事實。

壹、研究動機

我在網路上看到一篇「Notes on machine learning - Playing UNO」，文章中使用Genetic algorithm、Dummy algorithm以及Naïve offensive/defensive algorithm進行實驗。與Dummy Algorithm、Naïve offensive/defensive algorithm相較下，Genetic perceptron的勝率到達左右，筆者在Genetic perceptron內加入Hidden layers，不過成效不彰。

我認為，成效不彰是因為難以用單純的Genetic fluctuation來訓練神經網路，細菌使用基因來記憶所學所能，因此，只能完成相對簡單的任務；而人類使用大腦來記憶所學所能，因此，人類能完成相對複雜的任務。

Q-Learning好比人腦，Genetic algorithm好比細菌，若將Q-Learning應用在玩UNO上，好比細菌演化成人類，

貳、研究目的

一、探討先手與後手的利與弊

二、探討各演算法之間的勝率

三、探討Deep-Q-Network與Genetic-Network的利與弊

參、文獻回顧

# 一、Notes on machine learning - Playing UNO

<https://john-hearn.info/articles/notes-on-machine-learning-playing-uno>

As previous stated, the article shows several ways to conduct UNO games with neural networks.

# 二、RL - Cards

<https://arxiv.org/pdf/1910.04376.pdf>

It’s an open source library recently published in Feb 2020 for conducting researches of card games with Reinforcement learning. Despite its usefulness, the library can’t meet my customized needs, so I decide to use it as a blueprint to build my own.

# 三、Deep-Q-Learning

肆、研究器材

|  |  |  |
| --- | --- | --- |
|  | 器材 | 用途 |
| 研究所需器材 | *Tensorflow* | Lessen the coding complexity for developing. |
| *Tkinter* | Visualize the gaming process. |
| *NVIDIA 1660 Ti* | Hardware for increase training performance. |
| *CUDA* | Utilize GPU to increase training performance. |

伍、研究方法

# 一、遊戲規則

# 二、Dummy algorithm

一言以蔽之，這個演算法等同「找一張能丟的卡或抽牌」。

# 三、Naïve Offensive/Defensive Algorithm

一言以蔽之，激進演算法等同「先丟最強的功能牌，再丟普通牌，沒牌再抽卡」；保守演算法等同「先丟普通牌，再丟功能牌，沒牌再抽卡」。

# 四、Genetic algorithm

演算法如同在一個Cell Petri上模擬許多個Cells，任選兩隻細胞決鬥，根據決鬥結果計算細胞的最終積分，再進行基因擴散(Genetic diffusion)藉此交換基因(Exchange genetic)。

不適應環境的基因會在基因擴散的過程中漸漸消失，適應環境的基因會在基因擴散的過程中壯大，如同達爾文的「適者生存」說，演化過程會淘汰掉不適應的個體，留下「贏家」，也就是適應環境的個體。

# 五、Deep-Q-Learning Algorithm

Q-Learning

Deep-Q-Learning

陸、現階段研究成果

# 一、先手制霸

初始手牌的數量，會影響先手的勝率，下圖展示了不同演算法先手勝率與初始手牌數量的關係。

# 二、各演算法之間的勝率表

可以見到，

# 三、Q-Learning與Genetic Algorithm的勝率

這兩種演算法都會「自我學習」，讓自己越來越強，讓兩種不同的演算法相互對抗，並觀察勝率可以繪製成下圖。

柒、結論

# 一、探討先手與後手的利與弊

在特定條件下，先手能夠「完全制霸」對手，也就是存在一種出牌順序，使得先手必勝，如下圖所示。

先手只需要出「停止」、「迴轉」再出「紅5」，先手就贏了，

# 二、各演算法之間的勝率

Naïve offensive algorithm對上Naïve defensive algorithm時，offensive有較高的機率勝出，這是因為defensive algorithm常常把功能牌留到最後一刻，導致沒機會丟出功能牌。

# 三、Deep-Q-Network與Genetic-Network的利與弊

就算關閉了Deep-Q-Network的學習功能，讓Genetic-Network繼續演化，Genetic-Network也難以望其項背。

捌、參考資料

玖、後記

筆者是一個心智健全，認知功能完備的正常人，為了人類的榮譽，為了人類的尊嚴，筆者與Deep-Q-Network大戰三十回合，以下是筆者的戰績。

筆者慘遭親手創造的神經網路虐殺。