**成果報告書**

**應用多代理者深度學習演算法於路徑規劃之設計**

組員:徐廣亮 資工四A 411018183

指導教授:劉建興

著於中華民國一一四年

**目錄**

一、簡介 3

1.1系統簡介 3

1.2系統特色 3

二、系統描述 4

2.1系統架構 4

2.2系統實現 7

2.3系統成效 12

三、結論 14

四、參考資料 14

一、簡介

1.1系統簡介

我希望透過改進Informarl模型，再多代理者避免碰撞並達到指定目標物為基礎的架構，增加多代理者模型的判定使其能夠更有效率的執行目標任務。其主要核心為多代理者透過圖形聚合的方式處理代理者網路，結合距離、網路大小等建立優先條件實現類似於代理者自行排隊的效果來增進模型能力。

1.2系統特色

採用PPO為核心的架構，能夠有效限制跌帶速度與梯度變化大小，使其更加穩定且保持有效。控制Actor-Critic架構來實現模型網路的執行與評判，並由GNN作為聚合圖形的運算。

一張含有 文字, 圖表, 數字, 字型 的圖片

自動產生的描述

圖1 : INFORMARL模型概念圖

二、系統描述

2.1系統架構

本專題由結合<Informarl>結合<基於 MADRL 的具有防碰撞機制的雙航跡設計車上網路>的實現為內容，此處為介紹Informarl架構以及如何設計及實現進一步的防碰裝機制(2.2系統實現講解)。

Informarl主要由環境、訓練、評判、資訊聚合四個部分組成，分別由PPO、Policy、Actor-Critic、GNN四種模型實現。

環境與PPO

圖2 : INFORMARL模擬環境概念圖

一張含有 文字, 圖表, 地圖, 行 的圖片

自動產生的描述環境資訊由PPO整理，交錯與傳遞於整個模型間，同時擔當處理、彙整訓練匹伺候的資訊以及權重網路更新。

在此模型終結點網路指考慮相關的覆蓋節點，節省運算以及傳遞資源，再藉由PPO進行整合以及修剪梯度等學習參數，已達到擁有區域(agent本身與相關節點)，和全域(整個多代理者權重網路)兩者結合的主要架構。(配合圖2)

一張含有 文字, 字型, 螢幕擷取畫面, 行 的圖片

自動產生的描述

圖3 : INFORMARL訓練執行概念圖

訓練與評判與Policy

一張含有 文字, 字型, 螢幕擷取畫面, 圖表 的圖片

自動產生的描述Policy根據PPO傳入的數據和規定進行動作執行模擬，藉由切換演員網路(Actor)及評論家網路(Critic)，來進行訓練操作，將數據傳入Actor進行動作的篩選與檢測，並評估相對應的價值(根據動作後相對目標的距離、碰撞等條件)。

將自身涵蓋的節點作為動作參考計算碰撞，相對計算所有節點(agent)之間的距離和狀態更加省時、省資源，再Actor收到來自Policy的資料後，便會針對action\_spac(配合圖4)中記錄的行動比率篩選相對合適的行動，並且計算行動本身的價值後規還給Policy。

在完成動作後，把完成的資料交由Critic針對這次節點網路的行動做出評價(參考圖5)，並給出分數交還給PPO生成reward以及權重的更新。

圖4 : Action\_space(predictied action)與Critic評估價值概念圖

再處理節點訊息時，已節點圖形的資料形式計算，故使用GNN作為節點網路動作更新的核心算法。

。

圖5 : Critic將圖形資訊聚合後處理概念圖

一張含有 文字, 字型, 行, 螢幕擷取畫面 的圖片

自動產生的描述

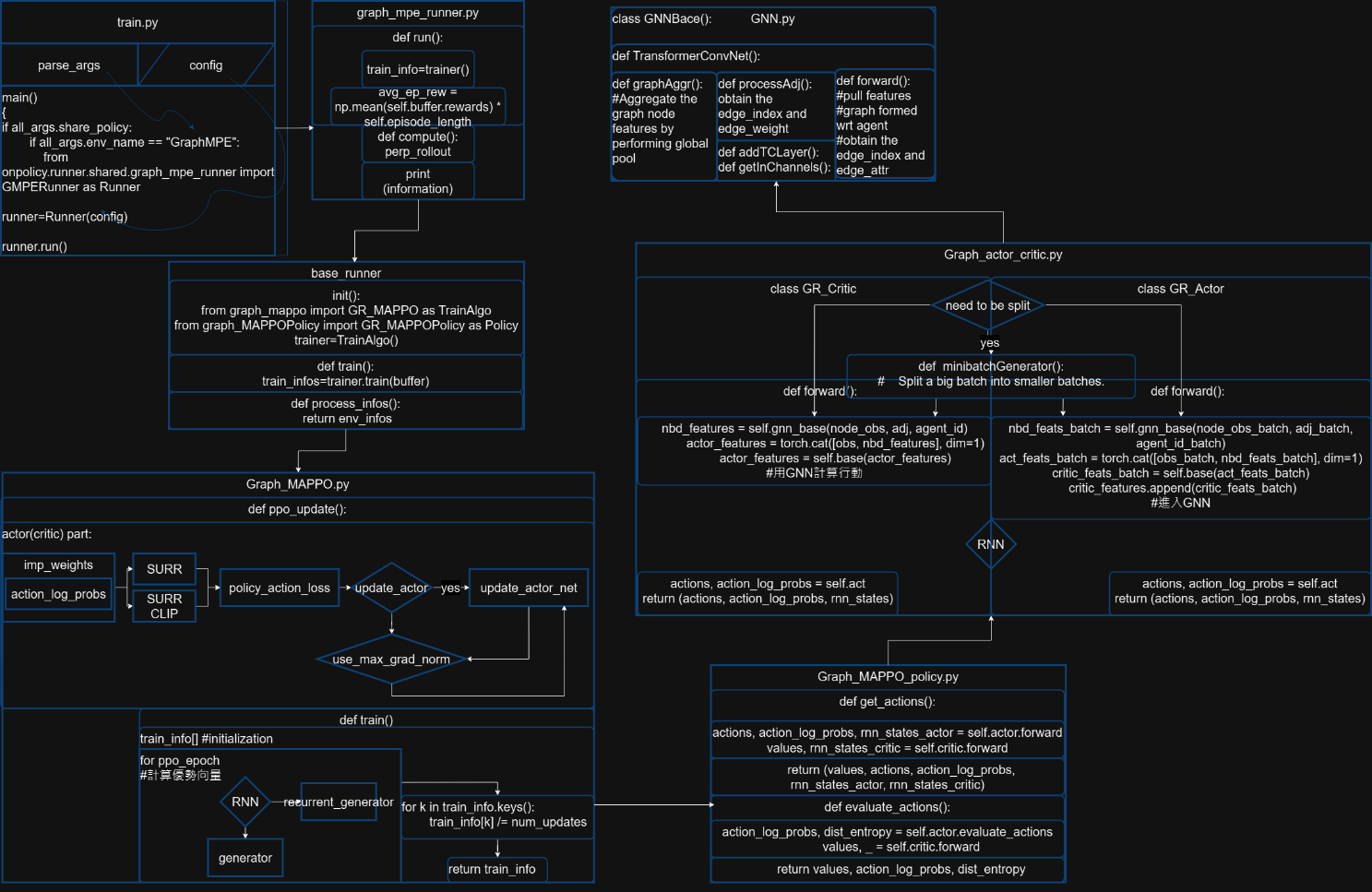
****2.2系統實現

圖6 : INFORMARL主要架構簡易概念圖↑

附圖7為原本的執行流程概念示意圖。由PPO控制訓練批次，Policy控制流程後將資料交給Actor執行、Critic評價。

附圖8是更改新增的實現，再Actor選擇動作時增加篩選的方式，處理動作分布機率，在執行動作選取。

圖7

一張含有 文字, 螢幕擷取畫面, 設計 的圖片

自動產生的描述

首先，按照節點相關度最高(身邊很多節點，會影響很多人)以及距離目標最接近的代理者優先，故[已覆蓋節點\*與目標距離的反比=排名依據(優先度)]。再決定優先度後，根據優先度高低決定此區域中的節點可行動分配，從最高優先度者優先，較低者依照高優先度代理者行動的預測結果屏蔽可能發生碰撞的行為，根據兩者相對距離對行動做出相對強度的屏蔽(叫高優先度者影響較小)。

一張含有 文字, 螢幕擷取畫面, 設計 的圖片

自動產生的描述

圖8

一張含有 文字, 螢幕擷取畫面, 軟體, 字型 的圖片

自動產生的描述

圖9

一張含有 文字, 螢幕擷取畫面, 字型 的圖片

自動產生的描述

圖10

一張含有 文字, 螢幕擷取畫面, 軟體, 陳列 的圖片

自動產生的描述

圖11

2.3系統成效

下四張圖(圖12~15)為訓練結果的reward圖表

12、13為7代理者 14、15為3代理者  
12、14為原始reward

13、15為每10筆reward取平均

圖13

圖12

圖14

圖15

三、結論

我們希望透過修改模型，再多代理者導航和避碰上增加能夠依照優先順序排隊進行來提高效率。

讓多代理者深度學習演算法透過結合不同條件，得以實現更智能、更靈活的路徑規劃，利用深度學習和多代理者系統的分散性，智能體能協作完成複雜任務，如避開障礙物、最小化碰撞風險。

在應用上可廣泛用於自動駕駛、無人機等領域，提高安全性、效率和靈活性，加速技術普及。研究也為智慧城市、工廠自動化等領域提供技術支持，實現城市交通管理、物流配送的智能優化，提升運行效率和資源利用率。

不斷改進與結合，使多代理者深度學習演算法於路徑規劃的應用，既解決現有方法的限制，又推動自主導航技術發展，效率的提升可能不夠明顯，但我相信在不同條件與環境下的結合並繼續改進，能夠有效讓自動導行技術得以更加廣泛的應用在不同領域。

四、參考資料

<https://github.com/nsidn98/InforMARL>

<https://arxiv.org/abs/2211.02127>

<https://arxiv.org/abs/2402.03342>