使用強化式學習(DDQN)來決策多回合制的模擬類型遊戲

Reinforcement Learning with DDQN for Decision Making in Turn-Based Simulation Games

學生:施其均　　指導教授: 蘇銓清

一、研究動機與目的

在多回合制的模擬養成遊戲中，玩家需要在有限的回合數與資源中，進行嚴謹的決策思考，追求做出最佳的選擇以達成最好的成果。在影響各回合的決策中，除了當下所持有的資源與環境狀態外，也要為了未來後續的回合做規劃，而不是只看當下的狀況做決定。傳統的Agent大多是Rule-based algorithm，但是此方法在狀態空間變化複雜的環境下較難彈性應對。為了克服此困難，本專題使用DDQN，使代理人能夠與環境不斷互動，自主學習並優化其長期的決策策略。

二、實作技術與環境

開發環境

作業系統：Windows 10

處理器：Intel Core i5-10300H，RAM 24 GB

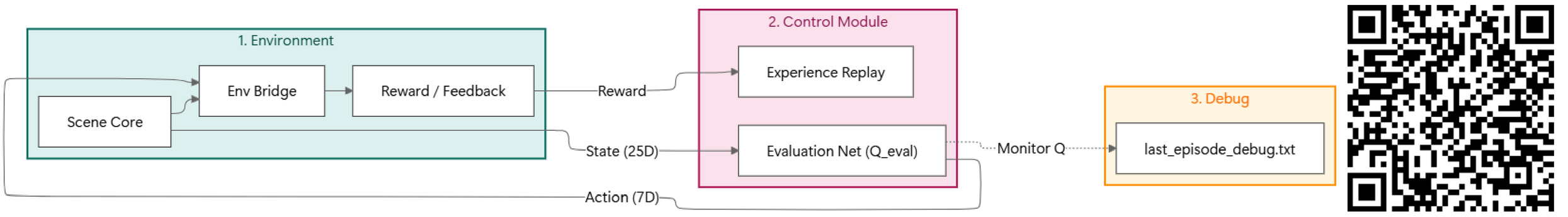
GPU：NVIDIA GTX 1650 Laptop 8GB VRAM

核心框架與工具

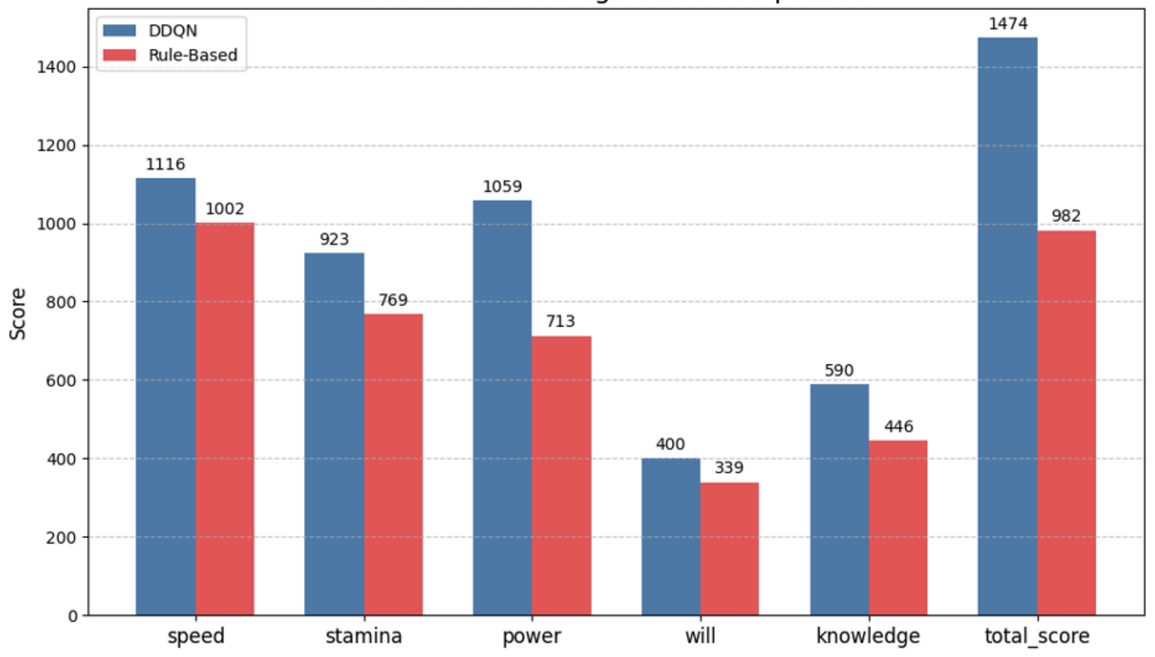
類別	工具
DRL	PyTorch 2.6 + CUDA
資料處理	NumPy、Pandas
視覺化	Matplotlib、Seaborn

三、系統架構圖

本專題採用Environment-Brain Decoupling，將馬娘動態培育環境與深度強化學習推論完全分離。模擬培育環境為系統基底，負責 60 回合時序養成限制與卡片隨機分佈模擬，並建立自適應獎勵函數進行生理約束與非理性操作的強力懲罰機制；中央控制主程式作為通訊調度核心，進行即時閉迴路數據同步。推論決策模組支援DDQN的動態布署，並整合



四、成果展示



Type	DDQN model	Rule-based model
成果上限	高	普通
訓練失敗率	稍高	低
前期拉羈絆表現	較好	普通
彩圈利用效率	高	普通

根據結果與實際操作可以發現，使用DDQN來做決策的表現比單純將規則寫死能夠有更高的上限，在訓練較差時能夠更動態的選擇其他動作(如休息)，在訓練失敗率低且有好的彩圈時，更能在訓練與休息做取捨，藉由做出更高效率的決策，以獲得更好的最終成績。

五、結論與未來展望

結論

傳統Rule-based模型易僵化，而 DDQN 能夠根據當下狀態動態評估未來收益，展現出承擔合理風險換取最高回報的靈活決策能力。資源利用與時序規劃最大化，模型在前期羈絆累積與後期高收益彩圈的取捨上具備顯著優勢，能權衡訓練與休息使整體培育效率最大化。

未來展望

實做更複雜劇本:藉由本專題建構之高擴充性閉環架構，引入更複雜的時序約束與動態事件，挑戰更高維度的連續控制任務。

優化模型穩健性:持續精進獎勵機制與神經網路結構，降低極端狀態下失敗率，提升代理人決策的穩定性與容錯率。

Demo Video