強化学習によるAIの個性化を導入した協調型NPCの実装

Wong Chung Bun[†1]　川合康央[†1]

概要：従来の NPC は固定的な行動に基づき，プレイヤーごとの特性に十分適応できないという課題があった．本研究では，Unity と ML-Agents を用いた強化学習により，プレイヤー行動に適応して協力関係を形成する NPC を構築し，さらに行動履歴や協調成功率を保持する AI の個性化を導入した．NPC は PPO により信頼度と連動した行動を選択し，プレイヤーと相互に進化する新しい協調型ゲーム体験を実現する．

1. はじめに

近年, ゲーム分野における人工知能（Artificial Intelligence: AI）技術の発展は著しく，特に強化学習（Reinforcement Learning: RL）は，環境との相互作用を通じて最適な行動方策を自律的に獲得できる手法として注目されている．これまで, RL を用いた NPC（Non-Player Character）の研究では，プレイヤーの行動に応じて難易度や戦略を調整する「適応型 NPC」が数多く提案されてきた．しかし，多くの場合，NPC は平均的なプレイヤー像に基づいて設計されており，プレイヤーごとの行動特性や協調スタイルの違いを長期的に反映する点については十分に検討されていない．

一方，近年のゲームデザインでは，プレイヤー同士の協力やチームワークを重視する Co-op 型体験が重要視されており，NPC に対しても単なる対戦相手ではなく，協調パートナーとしての振る舞いが求められている．このような背景のもと，NPC がプレイヤーと共通の目標を持ち，行動を相互に適応させながら協力関係を形成する仕組みが必要とされている．

本研究では，Unity および ML-Agents を用いて協調型強化学習 NPC を構築し, プレイヤー行動に適応するだけでなく，行動履歴や協調成功率に基づいてプレイヤーごとの特性を反映する「AI の個性化」を導入する．これにより, NPC はプレイヤーとの関係性に応じて振る舞いを変化させ，繰り返しプレイを通じて「自分専用」に成長する協調パートナーとして機能する．本研究の目的は，AI の個性化を組み込んだ協調型 NPC の設計と実装を通じて，プレイヤーと NPC が相互に進化する新しいゲーム体験を提案することである．

2. 関連研究

ゲーム AI 分野において，強化学習は，複雑な環境下で NPC が自律的に行動方策を学習できる手法として広く研究されてきた．代表的な研究として，Atari ゲームにおける Deep Q-Network（DQN）や，囲碁を対象とした AlphaGo などが挙げられ，これらは高次元状態空間においても最適行動を獲得可能であることを示している．これらの成果は，ゲーム AI における学習型 NPC の実用可能性を大きく広げた[1].

NPC の行動適応に関する研究として，Glavin らは Skilled Experience Catalogue（SEC）を提案し，プレイヤーのスキルレベルに応じて NPC の難易度を動的に調整する仕組みを示した．この研究は，プレイヤー特性を考慮した NPC 制御の有効性を示しているが，NPC は主に対戦相手として設計されており，協力関係の形成までは対象としていない．また，堂黒らは人狼ゲームを対象に，ニューラルネットワークを用いてプレイヤーの投票行動から役職を推定する手法を提案し，NPC が人間の意思決定を推測できる可能性を示した．しかし，これらの研究では，NPC がプレイヤーとの長期的な関係性を形成する点については十分に議論されていない [2].

一方, NPC の社会的適応に着目した研究として，Zhou らは Dialogue Shaping を提案し，NPC 間の対話を通じて社会的文脈に基づく行動を学習する手法を報告している．この研究は，NPC が人間的な振る舞いを獲得する可能性を示唆しているが，プレイヤー個人に適応した行動形成や協調関係の個性化については対象外である [3].

協調行動学習の分野では，マルチエージェント強化学習（Multi-Agent Reinforcement Learning: MARL）が注目されており，共通報酬に基づいて複数エージェントが協力行動を学習する研究が進められている．しかし，これらの研究の多くはエージェント間の協調に焦点を当てており，人間プレイヤーとの関係性や個人差を考慮した設計は限定的である．

以上のように, 従来研究では NPC の行動適応や協調学習に関する知見が蓄積されているものの，プレイヤーごとの行動特性や協調履歴を反映した「AI の個性化」を組み込んだ協調型 NPC の研究は十分に行われていない．本研究は，強化学習に信頼度および個性化行動プロファイルを統合することで，プレイヤーとの関係性に基づいて振る舞いを変

†1 文教大学

化させる協調型 NPC を提案し，既存研究との差別化を図るものである．

3. システムの設計と実装

本研究では，プレイヤーと NPC が共通の目標を持って協力行動を行う 2D グリッド型ゲーム環境を対象とし，協調型強化学習と AI の個性化を組み合わせた NPC 制御手法を提案する．NPC は単なる追従型エージェントではなく，プレイヤーとの関係性や行動傾向に基づいて振る舞いを変化させる協調パートナーとして設計されている．（図 1）

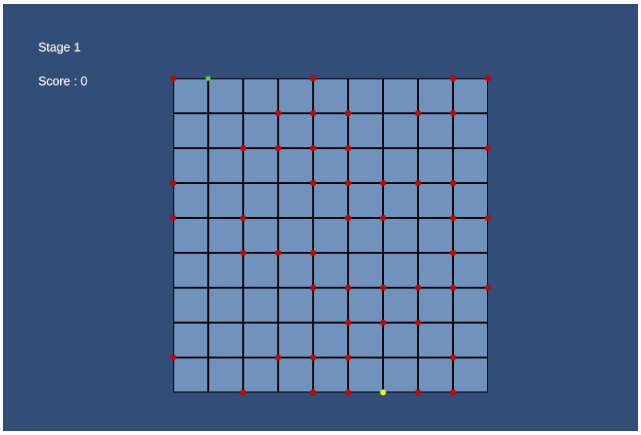

図 1　ゲーム内のマップレイアウト

3.1 システム構成

ゲーム環境は Unity 上に構築した 10×10 の 2D タイルベースのグリッド空間であり，プレイヤーおよび NPC はターン制で 1 ステップごとに 1 マス移動する．プレイヤーは手動操作により行動を選択し，NPC は学習済みの方策に基づいて自律的に行動する．両者は同一マップ上でタスクを共有し，協調行動によってステージクリアを目指す．

NPC は Unity ML-Agents の Agent として実装され，Proximal Policy Optimization（PPO）を用いて行動方策を学習する．観測情報には，プレイヤーとの相対位置，周囲のグリッド状態，タスク進行状況を含め，環境およびプレイヤー行動を統合的に把握できる構成とした．行動空間は「上」「下」「左」「右」「待機」の 5 種類とし，各行動は 1 マス移動または静止に対応する．

3.2 AI の個性化と信頼度モデル

本研究の特徴として，NPC の内部状態に AI の個性化（Personalized AI）を導入する．NPC はプレイヤーの行動履歴から，移動傾向，協調成功率，リスク選好などを簡易プロファイルとして保持し，これを観測情報の一部として利用する．これにより，同一環境下であっても，プレイヤーごとに異なる協調スタイルに適応した行動選択が可能となる．

さらに，NPC の行動制御には信頼度（Trust Value）を導入する．信頼度は，過去の協調成功・失敗履歴や行動の一貫性に基づいて更新され，NPC がどの程度プレイヤー行動を信用して先回り支援を行うかを制御する．信頼度が高い場合，NPC は自律的かつ積極的な支援行動を選択し，低い場合にはプレイヤーの動きを観察しながら慎重に行動する．この信頼度と個性化情報を統合することで，NPC はプレイヤーとの関係性に応じた柔軟な協調行動を形成する．

3.3 マップ生成と環境多様化

提案手法では，ゲーム開始時およびステージ遷移時にマップ構造を自動生成する仕組みを導入する．マップは複数のプリセット配置，または乱数に基づく配置規則から生成され，プレイヤーおよび NPC の初期位置，タスク関連ポイント，通行可能領域の連結性を考慮した制約条件を設けている．

この動的マップ生成により，プレイヤーは毎回異なる空間構造に直面し，NPC は特定のマップに依存しない協調行動を学習する必要がある．その結果，提案手法に基づく NPC は，環境変化に対しても汎化性能を持つ協調方策を獲得できる．

4. 学習過程

提案手法に基づく NPC の学習過程を評価するため，学習中に得られた各種指標を図 2〜図 5 に示す．本研究では，Policy Loss，累積報酬（Cumulative Reward），エピソード長（Episode Length），および複数試行間の挙動を用いて，学習の安定性と協調行動の形成過程を分析した．

図 2 は，PPO による学習中の Policy Loss の推移を示している．複数の学習試行において，Policy Loss はおおむね一定範囲内で推移しており，大きな発散は確認されなかった．この結果から，提案手法における方策更新は安定して行われており，過学習や不安定な更新が生じていないことが示唆される．

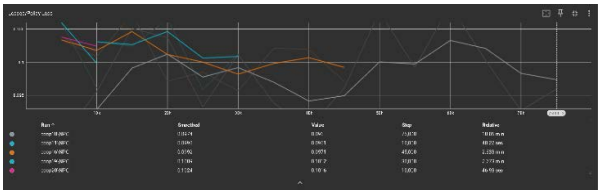

図 2　PPO による学習中の Policy Loss の推移

図 3 は，学習ステップに対する累積報酬の変化を示している．学習初期では負の報酬を示す試行も見られるが，学習の進行に伴い報酬は徐々に増加し，正の値へと収束する傾向が確認された．これは，NPC が環境構造およびプレイヤー行動を学習し，協調タスクをより効率的に達成できる方策を獲得したことを示している．

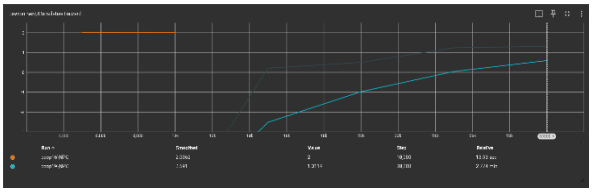

図 3　学習ステップに対する累積報酬の変化

図 4 は，エピソード長の推移を示している．学習初期ではエピソード長が長く，無駄な移動や失敗行動が多く見られたが，学習が進むにつれてエピソード長は短縮する傾向を示した．この結果は，NPC が不要な行動を減らし，タスク達成までの行動効率を向上させたことを意味する．

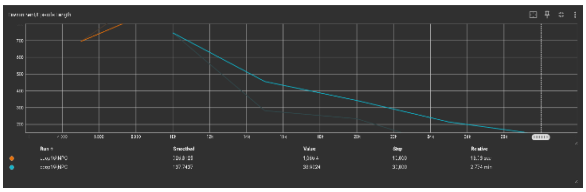

図 4　エピソード長の推移

図 5 は，本研究において実行した複数の学習試行（Run）の一覧を示している．各試行(Run)は，同一の環境設定および学習アルゴリズム（PPO）を用いながら，初期状態や乱数シードの違いによって独立に試行した結果である．これにより，提案手法の学習挙動が特定条件に依存せず，再現性を持つことを確認できる構成としている．

複数 Run を並行して比較することで，学習速度や収束過程にばらつきが存在することが確認される一方，いずれの試行においても学習の進行に伴い協調行動が形成される傾向が見られた．この結果は，提案手法が初期条件の違いに対して一定の頑健性を有していることを示唆している．

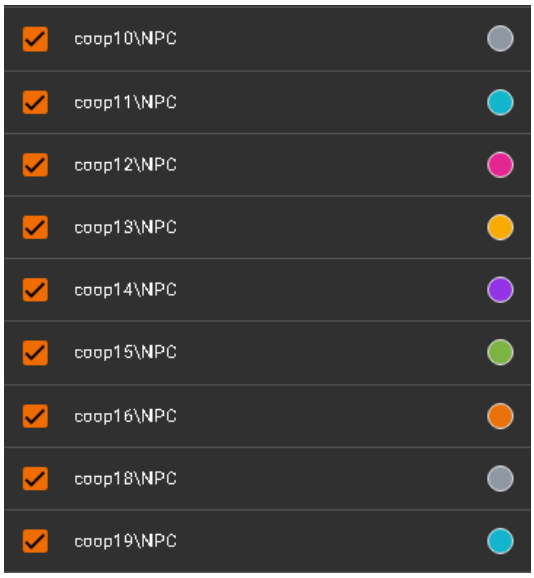

図 5　学習試行の一覧

以上の結果より，提案手法に基づく協調型 NPC は，安定した学習過程を通じて協調行動を獲得し，プレイヤーとの関係性に基づいた行動形成を段階的に実現していることが示された．

5.　まとめと今後の課題

本研究では，従来の固定的または対戦型 NPC を拡張し，プレイヤーと協力関係を形成する協調型強化学習 NPC の設計と実装を行った．Unity および ML-Agents を用いた 2D グリッド型ゲーム環境において，PPO による強化学習を適用し，NPC がプレイヤー行動や環境構造に適応しながら協調行動を獲得できることを示した．

さらに，本研究の特徴として，NPC の内部状態に AI の個性化および信頼度モデルを導入し，プレイヤーごとの行動履歴や協調関係を反映した行動制御を実現した．これにより，NPC は単なる追従や難易度調整を行う存在ではなく，プレイヤーとの関係性に応じて振る舞いを変化させる協調パートナーとして機能することが確認された．学習過程の分析からも，累積報酬の向上やエピソード長の短縮が観測され，提案手法が安定した学習過程を通じて協調行動を形成できることが示唆された．

一方で，本研究にはいくつかの課題も残されている．本手法は報酬設計に依存する部分が大きく，環境やタスク設定が変化した場合の学習効率については今後の検討が必要である．また，現段階ではエピソード単位で方策を固定しており，プレイヤーに対するリアルタイムなオンライン適応は行っていない．

今後の課題としては，LSTM などの時系列モデルを導入することで，プレイヤーの長期的な行動傾向をより精度高く推定することや，マルチエージェント強化学習を用いて複数 NPC が協調・分業する環境への拡張が挙げられる．最終的には，プレイヤーと NPC が相互に学習しながら関係性を深化させる，共進化的なゲーム AI の実現を目指す．

謝辞　本研究は JSPS 科研費 JP23K11728 の助成を受けたものです．

参考文献

[1] Frank G. Glavin, Michael G. Madden: Skilled Experience Catalogue: A Skill-Balancing Mechanism for Non-Player Characters using Reinforcement Learning, IEEE conference on computational intelligence and games, 2018, p.1-8.
[2] 堂黒浩明，松原仁: ニューラルネットワークを用いた人狼推定における投票先情報の有効性評価， GAT2018 論文集，2018，pp.1-4.
[3] Zhou, W., Peng, X. and Riedl, M.. Dialogue shaping: Empowering agents through npc interaction, 2023. arXiv preprint arXiv:2307.15833.