人間の副目的を尊重した協力型ゲームの AI に関する研究
（Respecting human sub-objectives in cooperative game AI）

北陸先端科学技術大学院大学　学籍番号 2350007
氏名　林 辰宜
主任研究指導教員氏名　池田 心

1. はじめに

近年，深層強化学習の登場により，ゲーム AI は様々なゲームで人間のトッププレイヤに比肩・超越する強さを獲得しつつある．これは，プレイヤ間の協力を含む「協力型ゲーム」についても同様である．

一方，人間を楽しませるという観点おいては，必ずしも強い AI が優れているとは限らない．協力型ゲームでは，単に主目的を協力して達成するのみならず，仲間の好みや意図に合わせて挙動を調整することも，よいゲーム体験のために重要である．特に，人間は主目的とは関係の薄い副目的を有している場合もあり，そのような副目的に対しても，協力可能な AI を設計することは好ましいと考える．

協力型ゲームの味方 AI に関係する研究領域に，Population-based Training（PBT）と呼ばれる方法の枠組みがある．これらの方法では，初対面の人間との協調可能な AI の作成を目標としており，テスト時に人間とプレイさせるゲーム AI（本番用 AI）を 2 段階の工程で訓練する．1 つ目の工程では，本番用 AI を訓練するためのゲーム AI のプール（訓練用 Population）を本番用 AI と独立させて学習させる．2 つ目の工程では，訓練した Population からゲーム AI をサンプリングし，それを訓練相手として本番用 AI を学習させる．これらの学習は，人間のデータを必要としない強化学習によって行われる．それぞれが異なった振る舞いを見せるゲーム AI から構成される訓練用 Population が，現実の人間と同様の多様性を有している場合，それらとの協力経験を得た本番用 AI は，様々な人間との協力が可能になるとされる．

そのため，既存の PBT 研究は，Population を多様化させる方法が重点的に研究されてきた．中でも，Hidden Utility Self-Play（HSP）[1]では，人間の副目的を内包する表現である「効用」を Population 内の各エージェントに割り当てることで，人間的な特徴を備えた相手への対応が可能な本番用 AI を実現しようとした．一方，HSP を含む多くの PBT 手法では，主目的の達成を目的としており，本番用 AI も訓練相手の効用ではなく，高い主目的スコアを得る振る舞いを学習するような仕組みとなっている．また，Population に人間の副目的を導入しようとする HSP においても，訓練される Population は主目的を軸にした協調を前提としているため，人間でしばしば見られる「主目的とは関係の薄い副目的」に対する協調を学習しづらいという課題があると考えた．冒頭で述べたように，人間を楽しませるには，相手の主目的だけでなく，副目的への協調も想定することが重要である．この観点に関して，既存の手法では，十分に人間を満足させる本番用 AI を作成できないと考えた．

本研究では，HSP（原始 HSP）をもとに，より人間の副目的への協調が可能な 2 つの手法を提案した．

その 1 つは，(1)主目的とは関係の薄い副目的を有する Population の学習，(2)本番用 AI の訓練時に訓練相手 AI の効用への接待を可能とする「接待 HSP」である．原始 HSP の Population の訓練では，主目的を追求するゲーム AI と副目的を含む効用を追求するゲーム AI の組で学習が行われていたため，プレイヤ間での協調を必要とする副目的を試みる振る舞いを獲得しづらいという課題があった．(1)は，主目的，副目的を含みうる同一の効用を追求する AI 同士で学習を行うことで，これに対処しようとするものである．また，原始 HSP では，本番用 AI を訓練する際に，主目的の達成度を最大化させることを目指す学習を行っていた．(2)では，この設定を変更し，訓練相手の効用の満足度を最大化させるような設定とすることで，時には主目的を犠牲にしてでも相手に接待するような振る舞いを本番用 AI に獲得させることを試みた．

2 つ目の提案手法は，接待 HSP を発展させた「不満忖度 HSP」と呼称する手法である．接待 HSP の Population 訓練時には，様々な効用のゲーム AI が発生する過程で，外見上の振る舞いは類似している一方，効用を満足させるための接待戦略が相反するような AI の組み合わせが想定される．それらの AI を見分け，それぞれに適切な接待を行うのは，原理的に困難である．他方，人間同士のプレイでは，ゲーム内の伝達行動，例えば不満を示すような行動を取ることが知られている．不満忖度 HSP では，そのような人間の不満を示す行動を Population に付与し，本番用 AI との訓練時に訓練相手 AI が不満を感じた際に，特定の行動パターンを発生させる．相反する効用の AI では，不満を示す条件が異なるため，これにより，外見上の振る舞いに差異が生まれ，本番用 AI はそれぞれの AI を見分けた上で適切な接待が可能になると考えた．

2. 研究方法

本研究では，「Overcooked 環境[2]」を題材として，手法の検討・実装・評価を行った．このゲームでは，2人のプレイヤが協力して料理を作成・提出し，制限時間以内に可能な限り高いスコアを獲得することを主目的としている．本環境は，多くの Population-based Training のテストベッドとして使用されていると同時に，人間的な副目的を表現することが可能である．

本研究は，提案手法が有効性を示すような事例を Overcooked 環境で設計し，訓練用 Population の学習時の振る舞い，本番用 AI の相手 AI に対する接待能力を調査する実験により，提案手法の評価を行った．これらのデータは，学習における経過時間を横軸，評価指標の値を縦軸とする学習曲線の形で表現される．接待 HSP は原始 HSP，不満忖度 HSP は接待 HSP と学習曲線を比較することで，手法の優位性を示す．

3. 結果と考察

本論文では，接待 HSP について 2 つ，不満忖度 HSP について 1 つ，良好な結果を示した事例を述べる．

接待 HSP を対象とした実験では，「外見上の振る舞い，満足させる接待戦略が異なる 2 種類の AI（玉ねぎが好きな O，トマトが好きな T）への接待」および「主目的と関係の薄い副目的を有する AI（皿が置かれている状態が好きな F）への接待」の場合について，原始 HSP と接待 HSP の訓練用 Population，本番用 AI を比較した．前者の例では，2 種類の訓練相手 AI から構成される Population に対して，原始 HSP の本番用 AI は片方の AI のみを最適に近い形で接待した一方，接待 HSP の本番用 AI は両者を見分け，それぞれに適した接待を切り替えることが確認された．後者では，主目的と関係の薄い副目的を有するように学習を試みた AI のみで構成される Population に対して，原始 HSP の本番用 AI は相手を満足させる協調とは正反対の振る舞いを学習した一方，接待 HSP の本番用 AI は最適に近い接待を行うことが確認された．また，Population の訓練においても，原始 HSP では副目的に協調するような振る舞いを学習できなかったのに対し，接待 HSP はそのような振る舞いを学習できることが確認された．

不満忖度 HSP を対象とした実験では，「主目的とは関係が薄く，接待戦略が相反する 2 種類の AI（皿が置かれている状態が好きな F，自分で皿を置くのが好きな P）への接待」の場合について，接待 HSP と不満忖度 HSP の本番用 AI を比較した．その結果，接待 HSP の本番用 AI は両者を見分けることができなかった一方，不満忖度 HSP の本番用 AI は相手を見極め，それぞれ最適に近い接待を行うことが確認された．このことから，実際の人間との協力を想定したとき，不満を示すような行動パターンを有するエージェントを Population に含ませることで，積極的に不満を示すような人間をより満足させることが可能になると考える．

4. まとめ

協力型ゲームで人間を楽しませるような AI を設計するとき，単に主目的を協力して達成する以外にも，人間プレイヤの副目的を尊重した戦略を AI が取ることが，人間プレイヤのゲーム体験にとって重要となる．本研究では，Population-based Training の既存手法である Hidden Utility Self-Play（原始 HSP）を，よりこのような需要に応えられるよう発展させた「接待 HSP」および「不満忖度 HSP」を提案した．実験の結果，接待 HSP は原始 HSP に対して，不満忖度 HSP は接待 HSP に対して，特定の場合に協力相手への協調能力が高いことが示された．

本研究では，不満忖度 HSP を簡単のため，単純な条件分岐と機械的な行動パターンによって実装した．一方，この手法は，強化学習の主要な手法で採用されている TD 誤差や Q テーブルの値を用い，本研究で扱った「期待累積報酬が低くなる場合」を検知するというような一般化も可能と考えている．また，不満を示す挙動についても，人間一般に通ずるヒューリスティックや人間のデータをもとにモデリングすることで，実際の人間に類似したものに置き換えるような発展も可能である．これにより，実際の様々な人間との協調が可能な AI の実現に寄与するものと考える．

参考文献

[1] Yu, C., Gao, J., Liu, W., Xu, B., Tang, H., Yang, J., ... & Wu, Y. (2023). Learning zero-shot cooperation with humans, assuming humans are biased. arXiv preprint arXiv:2302.01605.

[2] Carroll, M., Shah, R., Ho, M. K., Griffiths, T., Seshia, S., Abbeel, P., ¥& Dragan, A. (2019). On the utility of learning about humans for human-ai coordination. Advances in neural information processing systems, 32.