強化学習エージェントを用いた一般化メタ規範ゲームの解析

○宮田柚月・鳥海不二夫（東京大学）

Analysis of General Meta-Norms Game using reinforcement learning agents

* Yuzuki Miyata and Fujio Toriumi (University of Tokyo)

概要ー　社会的ジレンマを表現する方法として公共財ゲームが広く利用されており，さらに公共財ゲームを拡張したメタ規範ゲームが協調の促進を解析するにあたって用いられている．しかし，これらの研究で用いられる進化型計算でのシミュレーションは実社会で起こりづらいという問題がある．本研究では一般化メタ規範ゲームに対し，強化学習型のエージェントを用いてシミュレーションを行い，既存の進化手法である遺伝的アルゴリズムでの結果と比較し，メタ規範行動を解析する

キーワード:　エージェントベースドモデリング，ゲーム理論，メタ規範ゲーム

1　はじめに

人間社会をモデル化し，本質的な性質を理解する有用な方法としてゲーム理論が利用されてきた．中でも公共財ゲームは社会的ジレンマを含む状況をモデル化したものであり，多くの研究で利用されている．Axelrod[1]は，公共財ゲームに懲罰を導入したメタ規範ゲームを提案し，集団における協調を促進しようとした．また鳥海ら[2]は懲罰に加えて報酬を導入した一般化メタ規範ゲームを用いて，メタ規範行動が協調の促進に与える影響を明らかにしている．

これらの研究では遺伝的アルゴリズム(以下GA)などの進化型計算を用いている．しかし社会的ジレンマゲームの場合，実社会では進化型計算でのシミュレーションのような動きは起こりづらいという問題が指摘されている[3]．　そこで本研究では，強化学習を用いて一般化メタ規範ゲームをモデル化し，エージェントのメタ規範行動を解析する．

2　エージェントの進化方法

エージェントは 1 ステップ当たりゲームフェーズを 4 回繰り返す．その後，4 回で得られた利得の平均をスコアとし，その時の協調率 P_i・反応率 R_i に対する結果として記憶する．その記憶を用いて，強化学習を行う．強化学習の手法としては n 本腕バンディット問題を用いる．取りうる協調率 P_i・反応率 R_i を[0,1]の 0.1 刻みの 11 段階の離散値とし，記憶からそれぞれの値の報酬期待値を次式で計算する．

$$V_t^k = \sum_{j=1}^{n} S_j / n$$

ここで，V_t^kが t ステップ目における P_i および R_i の値がkの時の報酬期待値，nがk値を選択した回数，S_jがn回中j番目のスコアを表す．各エージェントは確率εでランダムに探索を行い、$1-\varepsilon$で$k = 0.0,0.1 \ldots 1.0$の中でV_t^kが最も高い値を次の協調率 P_i・反応率 R_i として選択する．

本モデルでは，初め 50 回までは$\varepsilon = 1$，5000 回まで$\varepsilon = 0.1$,以降は$\varepsilon = 0.05$に設定した．また，エージェントの持てる記憶は過去 200 ステップに限定した．

3　シミュレーション結果と考察

10000 ステップ後の，各μ, δにおけるエージェントの平均協調率の結果を以下の Fig. 1,2 に示す．

Fig. 1,2 より，メタ規範ゲームにおける協調はμ, δが報酬がコストより大きい場合に促進されることがわかった．また，先行研究の結果と比較し，GA を用いた場合と同様の結果が得られることがわかった．GA を用いた場合に比べ，協調率の上昇の仕方が急であるが，これは離散値を用いていることが原因と考えられる．また，強化学習型エージェントの方が協調が早まったと考えられる．

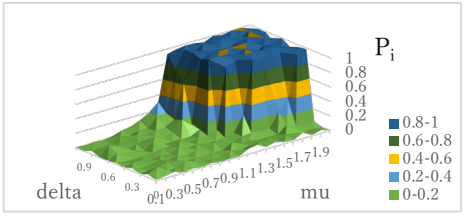

Fig. 1:　シミュレーション結果　協調率

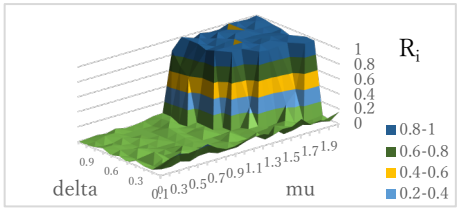

Fig. 2:　シミュレーション結果　反応率

4　まとめ

本研究では，強化学習型エージェントを用いて一般化メタ規範ゲームのモデル化を行った．そしてシミュレーションによって GA を用いた場合と同様の結果が得られることを示した．

今後の課題としては，エージェントの記憶の多様化があげられる．人間により近しいモデル化を目指し，いい/悪い記憶を強く記憶するなどエージェントの記憶にバイアスをかけ，どのよう行動され協調が促進されるか解析する必要がある．

参考文献

1)　R.M. Axelrod. An Evolutionary Approach to Norms. American Political Science Review, **80**-4, 1095/1111, (1986)

2)　Fujio Toriumi, Hitoshi Yamamoto, Isamu Okada,Exploring an Effective Incentive System on a Groupware, Journal of Artificial Societies and Social Simulation, **19**-4, 1/13 (2016)

3)　Takanori Ezaki,Yutaka Horita, Masanori Takezawa, Naoki Masuda,Reinforcement Learning Explains Conditional Cooperation and Its Moody Cousin，PLoS Computational Biology (2016)