

科学研究費助成事業 研究成果報告書

令和 6 年 6 月 19 日現在

機関番号：13903

研究種目：基盤研究(C)（一般）

研究期間：2019～2023

課題番号：19K12118

研究課題名（和文）連続的意思決定を要する環境における報酬評価システムの進化

研究課題名（英文）Evolution of Reward Appraisal Systems in Environments with Sequential Decision Making

研究代表者

森山 甲一（Moriyama, Koichi）

名古屋工業大学・工学（系）研究科（研究院）・准教授

研究者番号：10361776

交付決定額（研究期間全体）：（直接経費） 3,400,000円

研究成果の概要（和文）：本研究は、環境中で行動する複数の主体（エージェント）が、行動の結果として環境から得られる報酬をもとに独立に行動を学習する環境において、協力的な行動を学習する仕組みについて検討した。特に、現実と同様に過去の決定が未来に影響を及ぼす環境として、複数のエージェントの協力が必要なゲームを対象として研究を行った。主に、利己的な行動では目的が達成されないゲームにおいて、エージェントが内部で報酬を生成して協力的な行動を導く手法を検討し一定の成果を得た。さらに、そのような性質を持つ簡単な環境を考案して調査したところ、内部報酬によらない協力行動の学習の可能性が明らかになった。

研究成果の学術的意義や社会的意義

本研究では、個々が独立に報酬を最大化しようとする利己的な個体間の協力行動の発生について検討した。既存の利他的行動に関する数理的研究では主に進化を議論しているが、個体の学習の影響を考慮したものは少ない。本研究は、主に個体の学習に着目し、それを制御する報酬の設計を進化計算で試みたものである。学習における報酬設計は近年重要な課題となっており、学術的意義は大きい。また、既存の数理的研究では各行動が時間的に独立な場合の議論が多く、本研究のように過去の行動が未来に影響を及ぼす環境を対象としたものは少ない。さらに、今後のこの分野の研究の発展のため、このような環境の簡単な例を示したことは重要である。

研究成果の概要（英文）：This work investigated a mechanism in an agent in an environment with others that allows the agent to independently learn cooperative behavior from given rewards. In particular, this work used computer games requiring cooperation of multiple players as the environment where sequential decisions were necessary. In games where self-interested behavior failed to satisfy objectives, this work obtained positive results in proposals that invoke cooperative behavior through intrinsically generated rewards in the agent. Moreover, by proposing and investigating a simple game with the above properties, this work showed that cooperation may occur without the intrinsically generated rewards in such games.

研究分野：知能情報学

キーワード：マルチエージェントシステム 強化学習 協調 報酬設計 内発的動機付け 進化 人工社会 ゲーム

1. 研究開始当初の背景

我々人間は、個々が独立した存在でありながら、互いに協力することで、対立を軽減したり、個々では不可能なことを成し遂げたりする。ここで重要な点は、人間は基本的に自らの意思や目的を持つものであり、外部から直接的に協力を強要されるわけではない点である。人間の集団である社会において、それぞれの持つ意思や目的が完全に同一になることはなく、場合によっては自己の利益が他者の損失になることもある。人間は、そのような社会においても協力し得る。人間のような知的な人工物の構築を目的とする人工知能研究において、このような、個々が独立した存在であり、場合によってはトレードオフの関係にあるにも関わらず、互いに協力し得る人工物（エージェント）を構築することは、1つの大きな研究テーマである。

ある種の個体が他者を利する利他的行動が発生するメカニズムについては、理論生物学および進化ゲーム理論の分野で古くから議論されている。そのシミュレーション手段としてよく用いられる進化計算手法では、染色体や遺伝子と各個体の戦略が直接対応している。しかし、人間を含む高等動物の戦略は常に染色体や遺伝子に制御されているわけではなく、個体の学習が重要な役割を果たしている。生物の学習プロセスを模したアルゴリズムとして強化学習がある。しかし、強化学習は外部から与えられる報酬と呼ばれる評価値を最大化する戦略を学習するアルゴリズムであるため、一般に利己的な行動を獲得する。

この背景の下、研究代表者は進化計算と強化学習を組み合わせ、エージェントの戦略を間接的に制御する方法として、強化学習に用いる報酬を変更する仕組み、すなわち評価メカニズムを導入し、強化学習で得た行動の結果に基づいてそれを進化させる方法を提案してきた。しかし、対象とする環境として、ゲーム理論における繰り返しゲームという、毎回の意思決定が独立している特殊な「社会」を用いていた。一方で、我々人間の社会では、個体は連続的な意思決定を行う。過去の決定が未来に影響を与えるという意味で、毎回の意思決定は独立ではない。そのため、対象とした「社会」が現実のものと乖離しているという問題があった。

2. 研究の目的

上記の背景の下、本研究は対象とする「社会」をより現実に近い、連続的な意思決定を必要な環境とした。そして、その環境において、進化計算と強化学習を組み合わせ、強化学習の結果に基づく進化により、その環境における協力行動を間接的に導く評価メカニズムについて調査・検証することを目的とした。

3. 研究の方法

対象とする現実的な「社会」としては、連続的な意思決定が必要なことに加え、各エージェントにとっては開いた環境、すなわち他者について事前知識がないことを想定しつつ、実験者にとっては制御可能な環境であることが望ましい。そのような環境として、本研究では主にコンピュータゲームを扱った。コンピュータゲームは、意思決定に必要な全ての情報が得られるとは限らない状況で、与えられる評価（例：得点）について目的（例：最大化）を達成するためにプレイヤーが行動する環境であると考えることが出来る。従って、複数のプレイヤーが存在するゲームでは、プレイヤーをエージェントと考えると、ここで想定する「社会」となり得る。また、コンピュータゲームはそれ自体が目的を持っているため、その上で動くエージェントについて、どのような仕組みがその目的を達成しやすいか、という視点で評価できる利点がある。

コンピュータゲームにおけるエージェント、すなわち人工知能プレイヤーの作成の研究は、古くから行われている囲碁、将棋などのボードゲームに加え、アクションゲーム、格闘ゲーム、シミュレーションゲームなど様々なジャンルのゲームで行われている。さらに、様々なゲームにおける適切な戦略を求めようとする汎用ビデオゲーム人工知能 (GVGAI: General Video Game Artificial Intelligence) の研究も盛んに行われており、競技会が開催されている。複数のプレイヤーが存在するゲームの多くは、プレイヤー同士が対戦して勝敗を決めるものであるが、プレイヤーの協力が目的を達成するための要件であるゲームも存在する。本研究では、このような連続意思決定を要する「協力すべき」ゲームに注目し、これらのゲームにおいて、エージェント間の協力をもたらす評価メカニズムについて調査した。

4. 研究成果

本研究ではいくつかの「協力すべき」ゲームなどについて検討を行った。

- (1) 「追跡問題」はマルチエージェント強化学習で広く用いられているテストベッドで、複数の追跡者が（単一または複数の）獲物を2次元環境で捕獲するというゲームである。追跡者の

移動速度よりも獲物の移動速度を大きくすることで、追跡者が獲物を単純に追いかけるだけでは捕獲できず、待ち伏せなどの追跡者間の協力が必要となる。

この環境において、本研究と同様に評価メカニズムについて言及する内発的動機付け強化学習 (IMRL: Intrinsically Motivated Reinforcement Learning) の一種である好奇心探索手法[1,2]により、見慣れない状態に追加の報酬を与えることで、捕獲回数を増やすことができるかを調査した[3,4,5]。その結果として、好奇心探索手法が捕獲回数の増加に有効であることが示された。実際に追跡の状況を観察すると、1体の追跡者が回り道をすることで、結果として待ち伏せする形になる動きが見られた (図1)。

さらに、追跡問題において捕獲回数を増やす評価メカニズム (関数) を遺伝的プログラミングで獲得する手法について研究を行った[6]。その結果の一部として、状態の訪問回数が少なく、予測される累積報酬の誤差が大きい場合に大きくなる評価関数が得られた。これは前者の結果と符合するものである。



図1 追跡問題における好奇心探索の結果の例 (左から右へ時間経過を示す) [4]. 赤い追跡者3体が緑の獲物を追いかけるが、1体だけ下から回り込むことで、最終的に捕獲できている。

- (2) 「Akka Arrh」は2016年GVGAI競技会の2人ゲーム部門における「協力すべき」2次元環境ゲームである。脱出のための宇宙船を敵の攻撃から守りながら、2次元環境中にある宇宙船の鍵を取得しなければならない。

従来の繰り返しゲームにおける評価メカニズム導出手法 (効用導出関数) を適用するには、位置などを表す状態などの変数を新たに入力とする必要がある。そこでまず、入力の自由度を高めるため、効用導出関数を人工ニューラルネットワークで表現した[7] (図2)。それからこの手法をこのゲームに適用したが、残念ながら満足できる結果はまだ得られていない。

そこで、2次元環境における敵や味方、壁や鍵など各オブジェクトの相対位置が重要であるとの仮定を置き、それら他者からの距離に応じた内部報酬を与えることとし、その適切な距離を遺伝的アルゴリズムで求める手法を提案した[8] (図3)。この手法は (外部) 報酬の獲得という観点では一定の成果を得たが、実際には鍵を取ることよりも敵を倒すことを主に学習してしまった。これは、鍵を取得するよりも敵を倒し続けた方が (外部) 報酬が大きいというゲームの設定によるものであり、エージェントに鍵を取得する動機がないためである。

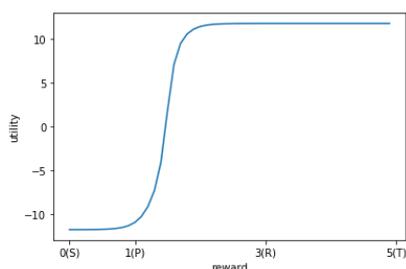


図2 人工ニューラルネットワークで得られた4人のジレンマゲームにおける協調行動を導く効用導出関数[7]。

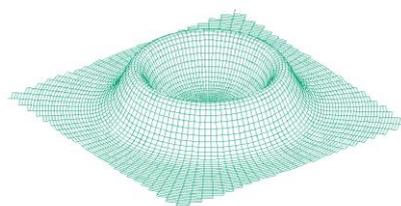


図3 距離に基づく内部報酬のイメージ。各オブジェクト中心からの距離に応じて報酬を追加で与えるものとし、その距離 (図における円の半径) を進化で獲得。

- (3) 「Geometry Friends」は円形の Circle と長方形の Rectangle が協力して空間中のダイヤモンドを取得するゲームである。2次元空間であるが、「追跡問題」「Akka Arrh」とは異なり、縦方向に下向きの重力がかかっており、3次元空間を横から見た構造になっている。Circle は上方向にジャンプできるが、Rectangle はジャンプできないというように、異なる行動集合を持つ。Circle 単体でのジャンプよりも高い位置にあるダイヤモンドを取得する場合、Rectangle が台になる必要がある (図4)。既存手法のようにこの協力的行動をあらかじめルールとして与えるのではなく、協力的行動そのものを学習することが可能かどうかを調査した。このゲームはダイヤモンドを取るという共通の目的があるため、特別な評価メカニズムがなくても強化学習できることが期待される。実際、他者の位置情報の一部を与えることで、シンプルな環境では協力的行動を学習することができた[9]。しかし、環境が物理法則に従うことから、ゲームの操作が難しく結果が安定的ではない。簡単なタスクから徐々に難しくするカリキュラム学習を用いた学習を試しているが、まだ満足する結果が得られていない[10]。

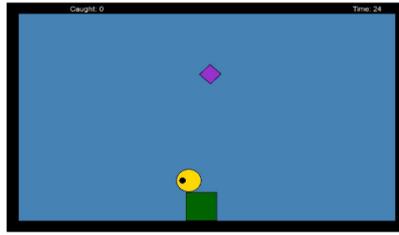


図4 Geometry Friends におけるシンプルな環境およびその環境での協力[10]. Rectangle の上に Circle が乗り、そこからジャンプすることで、初めて高い位置のダイヤモンドを取得可能となる。

- (4) 街中を歩いている人々は、ほとんど互いにぶつかることなく、目的地に向かって進む。互いにぶつからない、つまり避けることは一種の協力的行動と考えられる。そこで、この動きを強化学習で学習できるかどうかを確認した。多くのエージェントが通路を双方向に進むケース（対向流）では、各エージェントが進行方向の他エージェントの位置と動き（同方向か否か）や障害物を認識できる場合、できるだけ何もない方向を選択しつつ、それが不可能ならば同方向の他エージェントの後をついていくことを学習し、実際の人間の動きに現れる層状の流れを形成することを確認できた[11]（図5）。これは、各エージェントの持つ目的は共通ではないが、通常の強化学習で学習できる例である。

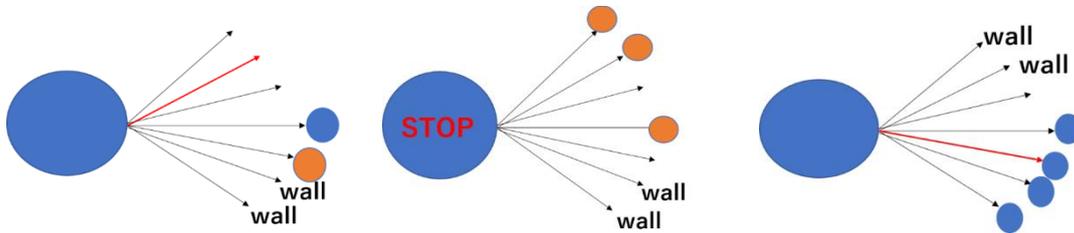


図5 通路を通るエージェントが学習した方策[11]. 青色は右方向に進み、橙色は左方向に進む。赤い矢印または文字が選択された行動。何も検知しない方向を選択する傾向が高いが、逆方向のエージェントが多い場合には停止し、同方向のエージェントが多くなる場合には後を追う。

- (5) 以上からも、「協力すべき」ゲームには様々な種類のものがあり、大きく分けても、利己的な行動を学習して構わないものと、そうでないもの、つまり利他的な行動を学習すべきものがあることが分かる。後者は、個人の利益の最大化が集団の利益の最大化につながらない、すなわち、これらの中にジレンマがあり、その解決のために評価メカニズムが必要と考えられる。そこで、具体的な環境から少し離れて、このようなジレンマのある連続意思決定環境について、より一般的な検討を試みた。

ジレンマ状況にある連続意思決定環境として、連続社会的ジレンマ（Sequential Social Dilemma: SSD）と呼ばれる環境が提案されている。これは、連続意思決定環境をゲーム理論における標準形ゲームに落とし込み、その利得構造でジレンマの度合いを表すものである。SSD は数学的には定義されているが、主に既存の複雑なゲーム環境の分析のために用いられており、その簡潔で具体的な事例が不明であった。SSD 自体の性質を明らかにするためには、2人ゲームにおける囚人のジレンマのような簡単で興味深い SSD 環境が必要と考え、SSD の条件を満たす簡単な「連続ゲーム」を考案し、その上で強化学習エージェントがどのように振舞うかを調査した[12]（図6）。その結果、囚人のジレンマの状況でも、未来の報酬の価値を割り引いて考える場合には、協力的行動が起きやすいことが判明した。

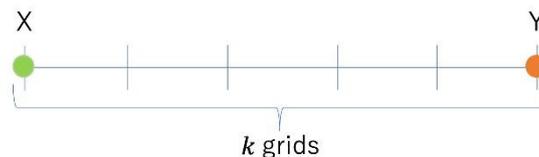


図6 SSD の条件を満たす簡単な連続ゲームの環境[12]. 2体のプレイヤーがそれぞれ相手に近づくかとどまるかを選択する。両者が隣り合うとエピソードが終了し、それぞれ決められた報酬を得る。移動に負の報酬を与えることで、できるだけ移動せずに隣り合いたいという囚人のジレンマの状況が現れる。

<本文中で参照した文献>

- [1] 村上知優, 森山甲一, 松井藤五郎, 武藤敦子, 犬塚信博: 深層強化学習における時系列的内部報酬生成器による探索の改善. 情報処理学会論文誌「数理モデル化と応用」, **14(1)**:1–11, 2021.
- [2] Zefeng Xu, Koichi Moriyama, Tohgoroh Matsui, Atsuko Mutoh, and Nobuhiro Inuzuka: Generating Intrinsic Rewards by Random Recurrent Network Distillation. 情報処理学会研究報告, **2021-MPS-132(15)**, 2021.
- [3] 岩科亨, 森山甲一, 松井藤五郎, 武藤敦子, 犬塚信博: マルチエージェント強化学習問題への好奇心探索の適用. 人工知能学会全国大会 (第 35 回) 論文集, 4G1-GS-2j-02, 2021.
- [4] 岩科亨, 森山甲一, 松井藤五郎, 武藤敦子, 犬塚信博: マルチエージェント深層強化学習における好奇心探索の影響. 情報処理学会論文誌「数理モデル化と応用」, **16(2)**:80–90, 2023.
- [5] 稲垣佑竜, 森山甲一, 武藤敦子, 島孔介, 松井藤五郎, 犬塚信博: 獲物が学習を続ける追跡問題への好奇心探索と貢献度の導入. 情報処理学会研究報告, **2024-ICS-212(2)**, 2024.
- [6] 磯部良隆, 森山甲一, 武藤敦子, 島孔介, 松井藤五郎, 犬塚信博: マルチエージェント環境下における強化学習エージェントの評価関数の特定. 人工知能学会全国大会 (第 38 回) 論文集, 1E5-GS-5-03, 2024.
- [7] 高塚遼市, 森山甲一, 武藤敦子, 松井藤五郎, 犬塚信博: 囚人のジレンマ環境における協調行動を導く主観的効用の最適化. 人工知能学会全国大会 (第 36 回) 論文集, 1N1-GS-5-01, 2022.
- [8] 中田瑛, 森山甲一, 武藤敦子, 松井藤五郎, 犬塚信博: 距離に基づく内発的報酬のためのポテンシャル場生成. 情報処理学会研究報告, **2023-ICS-208(1)**, 2023.
- [9] 薄佑太, 森山甲一, 武藤敦子, 松井藤五郎, 犬塚信博: 異種エージェント環境における協調行動の汎化のための状態空間の一般化. 人工知能学会全国大会 (第 37 回) 論文集, 1F4-GS-5-02, 2023.
- [10] 梶野嘉人, 森山甲一, 島孔介, 松井藤五郎, 武藤敦子, 犬塚信博: Geometry Friends におけるカリキュラム学習を用いた協調行動の獲得. 情報処理学会研究報告, **2024-GI-51(21)**, 2024.
- [11] 小林姫華, 森山甲一, 松井藤五郎, 武藤敦子, 犬塚信博: 強化学習による歩行者シミュレーションにおける行動規則の自動生成. 計測自動制御学会 第 48 回知能システムシンポジウム予稿集, B4-3, 2021.
- [12] Ryoichi Takatsuka, Koichi Moriyama, Tohgoroh Matsui, Atsuko Mutoh, Kosuke Shima, and Nobuhiro Inuzuka: A Simple Sequential Social Dilemma Game. *Proc. 7th IEEE International Conference on Agents (ICA)*, 2023.

5. 主な発表論文等

〔雑誌論文〕 計15件（うち査読付論文 4件 / うち国際共著 0件 / うちオープンアクセス 13件）

1. 著者名 Koichi Moriyama, Yoshiya Kurogi, Atsuko Mutoh, Tohgoroh Matsui, and Nobuhiro Inuzuka	4. 巻 -
2. 論文標題 Running Reinforcement Learning Agents on GPU for Many Simulations of Two-Person Simultaneous Games	5. 発行年 2019年
3. 雑誌名 Proceedings of IEEE International Conference on Agents (ICA)	6. 最初と最後の頁 50-55
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/AGENTS.2019.8929206	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 村上知優, 森山甲一, 松井藤五郎, 武藤敦子, 犬塚信博	4. 巻 14
2. 論文標題 深層強化学習における時系列的内部報酬生成器による探索の改善	5. 発行年 2021年
3. 雑誌名 情報処理学会論文誌「数理モデル化と応用」	6. 最初と最後の頁 1-11
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 岩科亨, 森山甲一, 松井藤五郎, 武藤敦子, 犬塚信博	4. 巻 16
2. 論文標題 マルチエージェント深層強化学習における好奇心探索の影響	5. 発行年 2023年
3. 雑誌名 情報処理学会論文誌「数理モデル化と応用」	6. 最初と最後の頁 80-90
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -
1. 著者名 Ryoichi Takatsuka, Koichi Moriyama, Tohgoroh Matsui, Atsuko Mutoh, Kosuke Shima, and Nobuhiro Inuzuka	4. 巻 -
2. 論文標題 A Simple Sequential Social Dilemma Game	5. 発行年 2023年
3. 雑誌名 Proceedings of IEEE International Conference on Agents (ICA)	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1109/ICA58824.2023.00011	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 百武佳輝, 森山甲一, 武藤敦子, 松井藤五郎, 犬塚信博	4. 巻 1
2. 論文標題 Sarsaエージェントによる囚人のジレンマゲームでの相互協調の継続回数	5. 発行年 2020年
3. 雑誌名 情報処理学会 第82回全国大会講演論文集	6. 最初と最後の頁 273-274
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また, その予定である)	国際共著 -

1. 著者名 横川滉太, 森山甲一, 武藤敦子, 松井藤五郎, 犬塚信博	4. 巻 JSAI2020
2. 論文標題 MCTSの木の再利用における減衰係数の自動調整	5. 発行年 2020年
3. 雑誌名 人工知能学会全国大会論文集	6. 最初と最後の頁 4B3-GS-1-04
掲載論文のDOI (デジタルオブジェクト識別子) 10.11517/pjsai.JSAI2020.0_4B3GS104	査読の有無 無
オープンアクセス オープンアクセスとしている (また, その予定である)	国際共著 -

1. 著者名 Zefeng Xu, Koichi Moriyama, Tohgoroh Matsui, Atsuko Mutoh, and Nobuhiro Inuzuka	4. 巻 2021-MPS-132
2. 論文標題 Generating Intrinsic Rewards by Random Recurrent Network Distillation	5. 発行年 2021年
3. 雑誌名 情報処理学会研究報告	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また, その予定である)	国際共著 -

1. 著者名 小林姫華, 森山甲一, 松井藤五郎, 武藤敦子, 犬塚信博	4. 巻 -
2. 論文標題 強化学習による歩行者シミュレーションにおける行動規則の自動生成	5. 発行年 2021年
3. 雑誌名 計測自動制御学会 第48回知能システムシンポジウム予稿集	6. 最初と最後の頁 B4-3
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また, その予定である)	国際共著 -

1. 著者名 岩科亨, 森山甲一, 松井藤五郎, 武藤敦子, 犬塚信博	4. 巻 JSAI2021
2. 論文標題 マルチエージェント強化学習問題への好奇心探索の適用	5. 発行年 2021年
3. 雑誌名 人工知能学会全国大会論文集	6. 最初と最後の頁 4G1-GS-2j-02
掲載論文のDOI (デジタルオブジェクト識別子) 10.11517/pjsai.JSAI2021.0_4G1GS2j02	査読の有無 無
オープンアクセス オープンアクセスとしている (また, その予定である)	国際共著 -

1. 著者名 高塚遼市, 森山甲一, 武藤敦子, 松井藤五郎, 犬塚信博	4. 巻 JSAI2022
2. 論文標題 囚人のジレンマ環境における協調行動を導く主観的効用の最適化	5. 発行年 2022年
3. 雑誌名 人工知能学会全国大会論文集	6. 最初と最後の頁 1N1-GS-5-01
掲載論文のDOI (デジタルオブジェクト識別子) 10.11517/pjsai.JSAI2022.0_1N1GS501	査読の有無 無
オープンアクセス オープンアクセスとしている (また, その予定である)	国際共著 -

1. 著者名 中田瑛, 森山甲一, 武藤敦子, 松井藤五郎, 犬塚信博	4. 巻 2023-1GS-208
2. 論文標題 距離に基づく内発的報酬のためのポテンシャル場生成	5. 発行年 2023年
3. 雑誌名 情報処理学会研究報告	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また, その予定である)	国際共著 -

1. 著者名 薄佑太, 森山甲一, 武藤敦子, 松井藤五郎, 犬塚信博	4. 巻 JSAI2023
2. 論文標題 異種エージェント環境における協調行動の汎化のための状態空間の一般化	5. 発行年 2023年
3. 雑誌名 人工知能学会全国大会論文集	6. 最初と最後の頁 1F4-GS-5-02
掲載論文のDOI (デジタルオブジェクト識別子) 10.11517/pjsai.JSAI2023.0_1F4GS502	査読の有無 無
オープンアクセス オープンアクセスとしている (また, その予定である)	国際共著 -

1. 著者名 稲垣佑竜, 森山甲一, 武藤敦子, 島孔介, 松井藤五郎, 犬塚信博	4. 巻 2024-ICS-212
2. 論文標題 獲物が学習を続ける追跡問題への好奇心探索と貢献度の導入	5. 発行年 2024年
3. 雑誌名 情報処理学会研究報告	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 梶野嘉人, 森山甲一, 島孔介, 松井藤五郎, 武藤敦子, 犬塚信博	4. 巻 2024-G1-51
2. 論文標題 Geometry Friendsにおけるカリキュラム学習を用いた協調行動の獲得	5. 発行年 2024年
3. 雑誌名 情報処理学会研究報告	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

1. 著者名 磯部良隆, 森山甲一, 武藤敦子, 島孔介, 松井藤五郎, 犬塚信博	4. 巻 JSAI2024
2. 論文標題 マルチエージェント強化学習環境下における有用な評価関数の特定	5. 発行年 2024年
3. 雑誌名 人工知能学会全国大会論文集	6. 最初と最後の頁 1E5-GS-5-03
掲載論文のDOI (デジタルオブジェクト識別子) 10.11517/pjsai.JSAI2024.0_1E5GS503	査読の有無 無
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

[学会発表] 計6件 (うち招待講演 0件 / うち国際学会 0件)

1. 発表者名 村上知優, 森山甲一, 松井藤五郎, 武藤敦子, 犬塚信博
2. 発表標題 深層強化学習における時系列的内部報酬生成器による探索の改善
3. 学会等名 情報処理学会 第127回数理モデル化と問題解決研究会
4. 発表年 2020年

1. 発表者名 木村哲, 森山甲一, 武藤敦子, 松井藤五郎, 犬塚信博
2. 発表標題 強化学習による衝突回避エージェントモデルの自動生成
3. 学会等名 第18回情報学ワークショップ
4. 発表年 2020年

1. 発表者名 中田瑛, 森山甲一, 松井藤五郎, 武藤敦子, 犬塚信博
2. 発表標題 エージェント間の距離がタスク達成に影響する環境下における報酬の制御
3. 学会等名 SMASH21 Summer Symposium
4. 発表年 2021年

1. 発表者名 木村哲, 森山甲一, 武藤敦子, 松井藤五郎, 犬塚信博
2. 発表標題 強化学習による交差流歩行者エージェントの行動規則の自動生成とその評価
3. 学会等名 SMASH22 Winter Symposium
4. 発表年 2022年

1. 発表者名 横川滉太, 森山甲一, 武藤敦子, 松井藤五郎, 犬塚信博
2. 発表標題 Open-Loop MCTSの木の再利用における適切な初期値設定
3. 学会等名 SMASH22 Winter Symposium
4. 発表年 2022年

1. 発表者名 岩科亨, 森山甲一, 松井藤五郎, 武藤敦子, 犬塚信博
2. 発表標題 マルチエージェント深層強化学習における好奇心探索の影響
3. 学会等名 情報処理学会 第142回数理モデル化と問題解決研究会
4. 発表年 2023年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関