

マルチエージェント追跡問題の段階相対視野における モジュラー強化学習

日大生産工 ○沖田 大地 日大生産工 山内 ゆかり

1 まえがき

近年、自律型ロボットの実用に向けて、マルチエージェントシステムでの協調行動の学習アルゴリズムの研究が注目されている。マルチエージェント強化学習における協調課題としては、一体の獲物を複数のハンターで捕獲する追跡問題が多用されている。渡邊らは、エージェントの持つ学習知識をモジュール化することで、学習時の次元の呪いを回避する階層型モジュラー強化学習を提唱した[1]。状態を部分視野で認識することで、状態数の削減を行い、モジュラー化により学習の高速化を実現したが、遠方のエージェントの位置が不明確なため、迅速に獲物に近づく最適行動が行えないなどの問題があった。

本研究では、渡辺らと同様に部分視野による状態数削減を行うが、近傍は正確に、遠方は大雑把に把握する段階的相対視野を取り入れることで、さらなる状態数の削減による学習の高速化と、遠方に位置した獲物に対しても、迅速に接近することが可能になり、捕獲までのステップ数の削減を実現の両立を目指す。

2 強化学習による追跡問題

強化学習はエージェントが試行錯誤に行動し目的を達成できたら報酬を与えエージェントを学習させるものである。追跡問題では強化学習によってハンターが獲物を捕獲するまでのステップを削減する。図1に示す8×8のグリッドワールド環境において、絶対座標系は原点を(1, 1)と設定しそこからの距離を位置情報とする(図2)。相対座標系はハンター自身の位置を(0, 0)とし、そこから獲物及び相棒エージェントの位置を+-7で認識する(図3)。

エージェントの認識する状態が絶対座標で構成される場合、8×8の64通りであるが、エージェントの位置を基準として+-7の広がりでの状態空間を相対座標で構成した場合は15×15の225通りとなる。

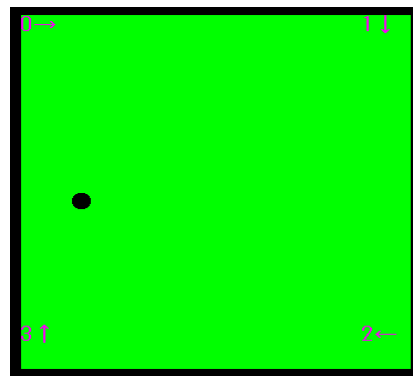


図1. グリッドワールド8×8

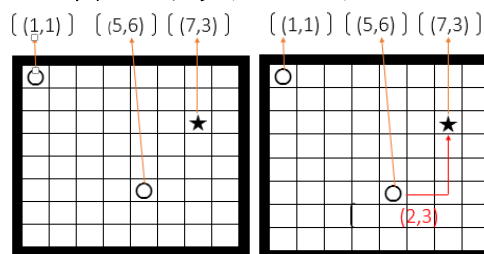


図2. 絶対座標

図3. 相対座標

渡辺らのハンターエージェントは、相対座標系での状態数の削減として、視野の範囲をエージェントを中心に5×5の25グリッドに制限している(図4)。

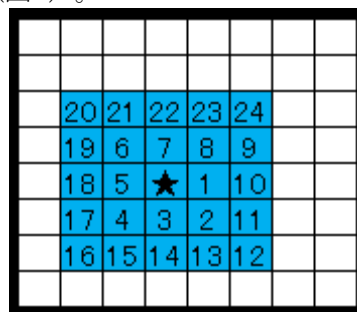


図4. 従来のエージェントから見える視野

ハンターが複数いる環境下では、状態空間はエージェント数分の次元が増加しグリッドサイズや仲間エージェント数が増えると爆発的に増大し学習性能が低下してしまう。その状態を圧縮するために、モジュラー強化学習では、各エージェントが認識する状態を部分空間に分割したモジュラー構造を適用する(図5)

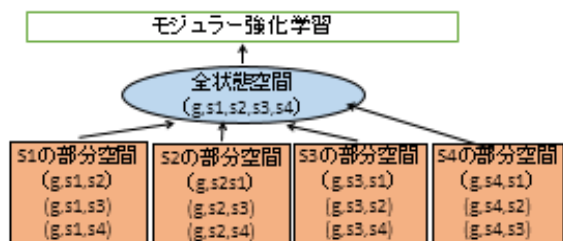


図5. エージェントの状態モジュラーモデル

3 提案手法

本研究ではハンターエージェントの視野を人間に近い段階的相対視野を採用し、獲物を捕獲するまでのステップの高速化を提案する。具体的にはエージェントの近傍は正確に状態を把握するが、遠方を大まかに把握することで少ない状態数で多くのエリアをカバーできるように段階的に視野を設定する。図6に提案手法の視野の範囲を示す。Q学習を行いエージェントが獲物を捕獲するまでのステップを削減できるか検証する。

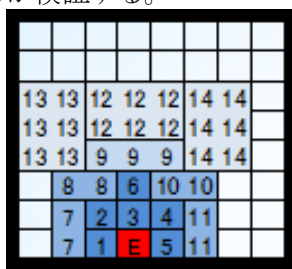


図6. エージェントから見た視野

4 実験環境

本実験で用いたグリッドワールドは従来研究と同様に図1に示される縦8マス横8マスの64マスを採用し、ハンターエージェント4体を角に配置し、獲物エージェント1体をランダムに初期配置した。エージェントの行動は上下左右の4方向に1マス移動するか現状のマスに留まるかの5つの行動ができる。Q学習を用い、ランダムに動く獲物を捕獲したら報酬を与え初期配置に戻る。捕獲条件は獲物を上下左右4方向から囲む、また壁を利用し獲物を囲むことができたなら捕獲成功とする。これを10000エピソード行いステップ数が削減されているか検証する。

5 実験結果および検討

絶対座標系と相対座標系で視野の制限を変え実験をおこなった学習の結果を図7に示す。グラフの横軸はエピソード回数、縦軸は捕獲までのステップ数である。また学習段階における捕獲までの平均ステップ数を表1に示す。

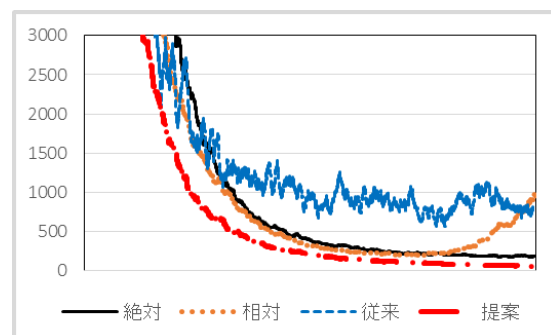


図7. 学習曲線比較

表1. 各学習データ

学習	初期	中期	後期
エピソード	1000付近	5000付近	9000付近
絶対座標	7407.1	381.7	188.9
相対座標	5940.0	319.4	455.5
従来視野	4771.8	948.1	675.0
提案視野	4285.1	206.7	61.2

提案手法による部分視野を持つエージェントは学習の初期、中期、後期のすべてで少ないステップ数での捕獲に成功しており、学習の高速化と最適行動の獲得の実現が示されている。従来の部分視野によるエージェントは、学習初期では他の手法より高速に学習の効果が表れているが、終盤での最適行動は得られておらず、捕獲までのステップ数が削減できていない。提案手法は、視野制限による状態数の削減で、学習の高速化が行えただけでなく、終盤でももっとも良い協調行動を獲得して収束している。

6 まとめ

本研究では、近傍は正確に、遠方は大雑把に把握する段階的相対視野を取り入れた、状態数の削減による学習の高速化と、遠方に位置した獲物に対しても、迅速に接近することが可能な、最適行動を獲得する強化学習手法を提案し、マルチエージェントの追跡問題においてその有効性を検証した。

「参考文献」

- [1] 渡辺、和田、マルチエージェント追跡問題のための相対座標系に基づく階層型モジュラー強化学習、バイオメディカル・フアジィ・システム学会誌 No2, pp. 65-74 (2010)
- [2] 高橋、柿倉、複数台移動のロボットによる協調捕獲に関する基礎研究、日本機械学会論文集 (C編) 71 巻 702 号 (2005)