ISSN 2186-5647

P-15

追跡問題における知覚情報の分散と粗視化

日大生産工　○鳥邊　大輝　日大生産工　山内　ゆかり

1. まえがき

強化学習(1)とは、試行錯誤して繰り返し学習を行っていくことである。本論文では強化学習の中でもProfit Sharing学習を行う。マルチエージェントシステムとは、複数のエージェントと呼ばれる自律的な行動主体から構成されるシステムである。各エージェントは環境から与えられる情報を知覚し、それぞれが目標を達成することを目指して行動する。個々のエージェントの比較的単純な知覚行動によって、全体として複雑で大規模な問題を解決しようとするシステムである。

大規模な問題を効率的に解くためには、各エージェントが協調して動作することが必要である。しかし、エージェントの行動を初めに設計することは、問題が複雑である程難しくなる。そこで、各エージェントが周囲の情報を得て知覚し、行動する事で試行錯誤的に環境への適応を目指すマルチエージェント強化学習に関する研究が行われている。

マルチエージェント強化学習において、各エージェントは他のエージェントを環境の一部とみなして学習を行い、その得られた結果の報酬によって各状態に適した行動を学習していく。そのためエージェント数が増える毎に知覚する状態数と学習時間が大幅に増えることが問題となっている。学習時間を抑える最も単純な手法の一つとして、知覚情報の粗視化がある(2)。知覚できる情報を敢えて制限し、内部情報のいくつかを同じと扱うことで、探索しなければならない状態数を減らし、時間の増加も抑える方法である。この方法は、環境や問題に適合しやすいという利点がある。

しかし、単純に粗視化を行ってしまうと、同時に知識を減らすことになり、学習によって得られる行動判断の精度が下がってしまう問題がある。そこで、これらの問題を防ぐ手法として、金重らによる詳細知覚情報の拡大手法がある(3)。

伊藤僚らは、伊藤昭らによる平均学習残エントロピー、平行学習の手法(2)を用いることで、学習の進行に応じて適応的に粗視化の段階を切り換えることを考えた(4)。これらを組み合わせたマルチエージェント強化学習手法を実装し、追跡問題に対する実験を通して、指標とする平均学習残エントロピーの閾値に関する学習速度、性能への傾向を調べた。その結果として詳細知覚範囲の拡大段階それぞれの学習器を並行して学習させることで、学習の高速化、性能の劣化を大きく抑えることが出来た。しかし、問題点として切り換える際に指標として用いることのできるぶれの少ない評価量が必要であると考えた。

そこで、本研究では、各エージェントに個別の学習段階を持たせ、エージェント毎に適応的段階学習を行う手法を提案する。

2. 従来研究

詳細知覚範囲の段階的な拡大を有効に利用するために、拡大の適切なタイミングを得る手段が必要である。伊藤僚らは、図1のように拡大段階の知覚を持つ学習器を並行して学習させ、各段階における切換のタイミングを次の段階の学習器の評価値から得られる平均学習残エントロピーがある閾値を下回った時点で一段階切り換える手法を採用した(4)。

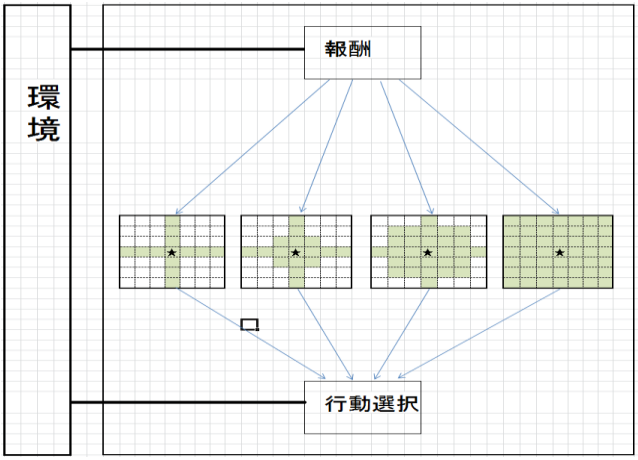

図1.従来手法

具体的には、Profit Sharing学習を行い、状態数が大幅に大きくなることを回避するために知覚情報を粗くし、いくつかの状態を同様とみなすことで探索すべき状態数を削減する方法を行う。

結果として詳細知覚範囲の拡大段階それぞれの学習器を並行して学習させることで、学習の高速化、性能の劣化を大きく抑えることが出来た。しかし、問題点として切り換える際に指標として用いることので

Balancing and Coarse-Grained of Sensory Information in Tracking Problem

Daiki TORIBE,Yukari YAMAUCHI

きるぶれの少ない評価量が必要であると考えた。

3. 提案手法

伊藤僚らは、マルチエージェント強化学習において、平均残エントロピーを用いた知覚情報の適応的粗視化を提案した。しかし、ぶれの大きい不安定な評価値であるという問題があった。

本研究では、図2のように各エージェントに個別の学習段階を持たせ、エージェント毎に適応的段階学習を行う手法を提案する。

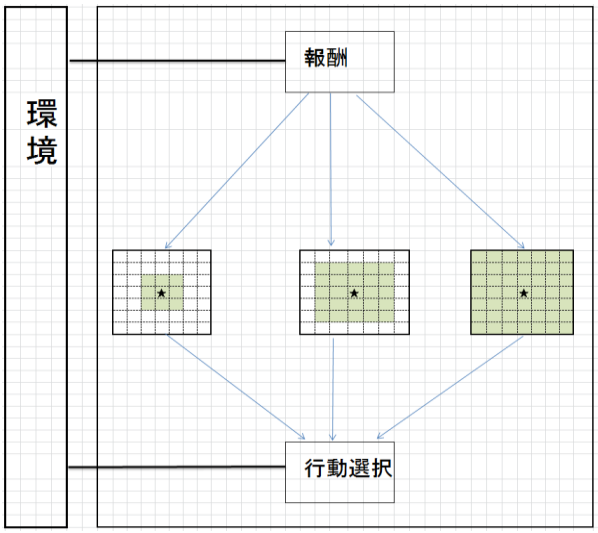

図2.提案手法

4. 実験環境

本研究の実験環境は、図3のように7×7のトーラス状の盤面を使いハンターエージェント 5 体と逃亡エージェント 1 体を置く。各エージェントに Profit Sharing 学習を行う。エージェント同志が干渉する場合は、その場に立ち止まるものとする。ハンターエージェント全てが逃亡エージェントに隣接することを目的とする。

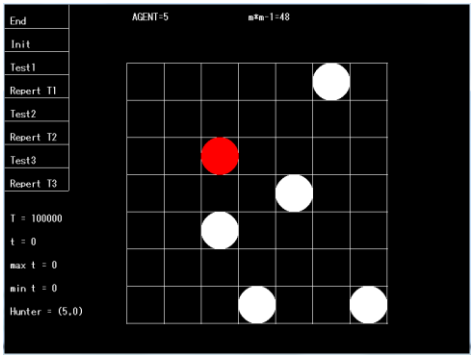

図 3.実験環境

各エージェントは開始時の初期位置はランダムで配置する。評価値を 1 とし割引率はγ=0.2 とする。捕獲した時、捕獲に関わったハンターエージェント全てに 1.0 の報酬を与える。

捕まえるまでの過程として、制限知覚学習と完全知覚学習をハンターエージェントに持たせ学習させていく。各エージェントが個別の学習段階を持ち、エージェント毎に学習器の切換を行うが、その時の切換のタイミングは平均残エントロピーを使って切換を行う。

5. 実験結果

従来研究では、時間のかかる初期の学習を高速化するために制限知覚学習を用い、ある程度学習を行った段階で知覚を完全知覚学習で行っていた。そのとき単純に閾値を下回った時、切換したり、平均学習残エントロピーの過去10000ターンの平均値が閾値を下回った場合に切換を行ったりしていたが、本研究ではエージェント1体ずつに適応的段階学習を持たせるためエージェント1体毎に適切なタイミングで切換ができることで平均捕獲ターン数を縮めることができる。

6. まとめ

本研究では、ハンターエージェントの各エージェントに個別の学習段階を持たせ、エージェント毎に定期王的段階学習を行う手法を提案した。この手法以外にも、平均残エントロピーの扱い方や新たな評価値を探すと、より良い結果が得られる可能性があると考えた。

「参考文献」

1) 三上貞芳，皆川雅章，「強化学習」，森北出版, 2000
2) 伊藤昭，金渕満，知覚情報の粗視化によるマルチエージェント強化学習の高速化：ハンターゲームを例に，電子情報通信学会論文誌. D-1，情報・システム, I-情報処理, Vol. J84-D-1, No. 3, pp.285-293, 2001
3) 金重徹，片山謙吾，南原英生，成久洋之：知覚情報の粗視化に基づくマルチエージェント強化学習の性能比較，自律分散システム・シンポジウム資料, Vol. 19, pp. 267-272, 2007
4) 伊藤僚，吉川毅，野中秀俊，マルチエージェント強化学習における知覚情報の適応的粗視化，26th Fuzzy System Symposium, September 13-15, 2010