平成13年度卒業研究		
入学年度　平成10年度	学籍番号　10117953	氏名　　保知良暢
論文題目　マルチエージェント強化学習における貢献度判別と報酬分配		犬塚研究室

1　はじめに

自由度が大きい複雑なマルチエージェントシステムに強化学習を用いる場合、実装者がシステム全体に対して目標を与えて報酬関数を設定することは易しいが、エージェント毎に対して設定することは難しいと考えられる。よってシステム全体に対して定義された報酬をエージェント毎に分配する必要がある。どのエージェントにどれだけの量の報酬を与えるべきか、という問題を**報酬分配問題**という。本研究では、マルチエージェント強化学習の枠組と報酬分配問題を解決する手法の提案を行なう。

2　マルチエージェント強化学習の枠組の提案

本研究で提案するマルチエージェント強化学習の枠組を図1 に示す。ag_i はエージェントを表し、それぞれ自律的に現在の世界の状態 $s(\in S)$ の観測、行動 $a(\in A)$ の実行を行ない、強化学習により最適政策の学習をする。これに対し世界の状態は状態遷移確率関数 P に従い次状態へ遷移する。この状態遷移を $Eval$ が評価する。また実装者は目標 $Goal$ を論理式で表現し与える。報酬関数 r はこの論理式を用いて次のように表現できる。

$$reward(rw) \leftarrow B_1 \wedge B_2 \wedge \cdots$$

述語 $reward(rw)$ は報酬 $rw(\in \Re)$ が発生することを表し、B_i は原子式を表す。報酬分配問題とは、このとき発生した報酬をどのように分配するか、という問題である。

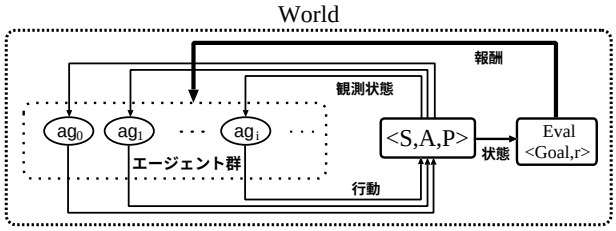

図 1: マルチエージェント強化学習の世界

3　報酬発生条件に基づく報酬分配手法

本論文では、報酬関数内の Horn 節の body 部を**報酬発生条件**とよび、報酬 R 発生時に報酬発生条件内に示されたエージェント全てに直接報酬 R、その他のエージェント全てに間接報酬 μR $(0 \leq \mu \leq 1)$ を分配する方法を提案する。図2は追跡問題のある終端状態を表している。この時報酬関数 r 内の Horn 節が発火し、報酬 100 が発生する。ここでエージェント変数 X, Y に対してエージェント定数 ag_0, ag_1 がそれぞれ bind される。報酬発生条件内に現れたこれらのエージェントが目標達成に必要だったことを示している。

各エージェントが強化学習アルゴリズムとして Profit Sharing を用いた場合の μ の範囲に関する合理性定理を以下のように示した。また μ が次の式を満たすとき報酬が発生しないという結果へは収束しないことを証明した。

$$\mu < \frac{M-1}{M^W\left(1-\left(\frac{1}{M}\right)^W\right)} \cdot \frac{{}_{n-1}C_{x-1}}{{}_{n-1}C_x L}$$

ここで n は全エージェント数、x は直接報酬を得るエージェント数、M は行動のバリエーションの数、W は直接貢献エージェントの最大エピソード長、L は同一感覚入力下に存在する有効ルールの最大競合数である。定理は非合

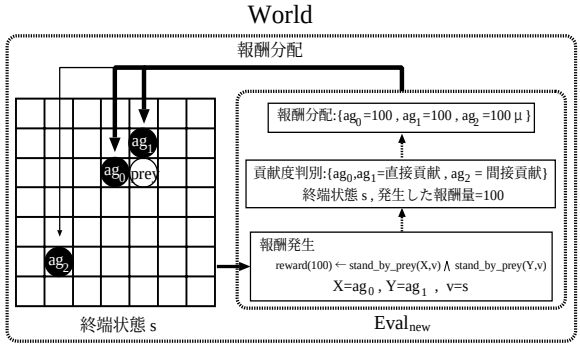

図 2: 報酬発生条件に基づく報酬分配

理的政策の抑制が最も困難な状況を考慮して導いた。全エージェントが等確率で直接報酬を得ると仮定し、各エージェントがある 1 つのルールのみを選択し続けて ${}_{n-1}C_x L$ エピソードで間接報酬を得た時の強化値の合計よりも、${}_{n-1}C_{x-1}$ エピソードで直接報酬を得た時の最初に選択されたルールの強化値が上回るように μ を制限している。

4　実験結果とまとめ

図2にあるような追跡問題を用い、提案手法と従来手法とを比較した。図3にその結果を示す。提案手法は従来手法よりも早く学習した。この理由は、提案手法では複数エージェントに直接報酬を分配していることと、μ の条件が従来手法よりも軽減されたことである。

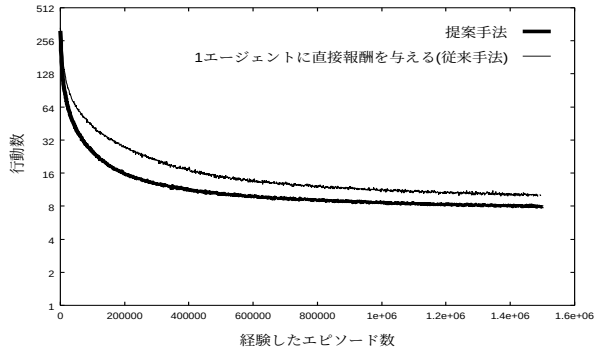

図 3: 1 エピソードあたりの獲物捕捉までの行動数の推移

本研究ではマルチエージェントシステムと強化学習の枠組を提案し、マルチエージェント強化学習における貢献度判別と報酬分配の手法を提案した。また Profit Sharing を用いた時の合理性定理を示した。しかし、従来手法や提案手法では誤った学習へ収束しないことは保証されるが、目標に対する最適解が得られるとは限らない。今後の課題としては、Profit Sharing 以外の学習アルゴリズムを用いた時の μ の考察と、目標に対する最適解が学習可能な手法の提案などが挙げられる。

参考文献

[1] 保知, 松井, 犬塚, 世木: Profit Sharing を用いたマルチエージェント強化学習における直接貢献エージェントに対する報酬分配の提案, 平成 13 年度電気関係学会東海支部連合大会講演論文集, p. 290 (2001).

[2] 保知, 松井, 犬塚, 世木: マルチエージェント強化学習における報酬発生条件に基づく貢献度判別と報酬分配, 2002 年度人工知能学会全国大会 (発表予定).