

# 強化学習を用いたマルチエージェントシステムにおける協調行動獲得\*

玉城 斉 遠藤 聡志 山田 孝治†

琉球大学工学部‡

## 1 まえがき

知的システム工学などの分野において、マルチエージェントのコンセプトが効果的な問題解決の枠組みとして注目されている。マルチエージェントシステムが効果的に問題を解決するためには、エージェント間での協調行動の獲得が重要となる。マルチエージェントシステムにおける協調行動は、個々のエージェントが解決すべき問題に対する各々の役割を適切に学習することで実現される。本研究では、知的かつ効果的なマルチエージェントシステムを設計するために基礎的実験を設計し、その結果について考察する。

## 2 エージェントの定義

マルチエージェントシステムでは、大規模・複雑な問題に対して処理の要素を各エージェントに分担させることで問題解決を行なう。本論文では以下のようにエージェントを定義する。

$$\text{Agent} = \{A, \mathcal{L}\}$$

ここで、 $A, \mathcal{L}$  は、

$A$ : 現在の状態において可能な行動の集合。

$\mathcal{L}$ : 状態に対応した最適行動を学習する学習機構。

である。エージェントは環境情報を自己の状態として認知する。そして、その状態に対して可能な行動集合から次の最適行動を過去の学習に基づいて選択する。

本稿では、エージェントの学習に代表的な強化学習法である Q-learning [1] を採用する。

## 3 Q-learning

Q-learning では、ルールと呼ばれる状態  $x$  と行動  $a(a \in A)$  の組に対する重み  $Q$  値を見積もることで学習を行なう。Q 値は、状態に対して与えられる報酬によって更新される。本実験では現在の状態において、Q 値が最大であるルールを選択する方法を用いる。

\*Acquisition of cooperative behaviors using Q-learning in multi-agent system

†T.Tamashiro S.Endo K.Yamada

‡Faculty of Engineering, University of the Ryukyus

## 4 追跡問題

追跡問題は、 $n \times n$  の大きさの 2 次元格子空間上で、複数のハンターエージェントが、ランダムに行動する獲物エージェントを捕獲することを目的としたマルチエージェント系のモデルである。

本研究では、獲物捕獲条件を、2つのハンターエージェントが獲物エージェントを両側から挟んだ状態とし、エージェント間の協調が問題解決に必須なモデルとして追跡問題を採用する。

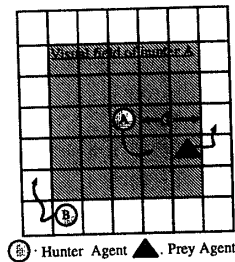


図 1: 追跡問題

エージェントは各タイムステップ毎に 1 マス移動、停止するといった行動を行なう。また、エージェントは同じマス上に同時に存在できる。

ハンターエージェントには、視覚として他のエージェントとの相対位置が与えられる。視野の大きさは、自分を中心とした  $(2d+1) \times (2d+1)$  マスである。また、ハンターと獲物の区別が可能である。

獲物捕獲までを一試行とし、捕獲後に再配置し試行を繰り返しエージェントの学習を行なう。

## 5 計算機実験

### 実験 1: 役割学習の確認

実験 1 では、追跡問題におけるハンターエージェント間での役割分化の有無の確認を目的とする。

ここで、環境の大きさは  $11 \times 11$  とする。ハンターエージェントおよび獲物エージェントの数は、それぞれ 2, 1 とする。ハンターエージェントの能力パラメータ

として、学習率  $\alpha$ 、視野の深さ  $d$  を用いる。この実験では、 $\alpha = 0.4, d = 5$  とする。

結果を図2に示す。図より捕獲行動を学習したエージェントの間でエージェント A は右から、B は左から主に獲物を捕獲している。このことから、役割学習を行ったことが分かる。

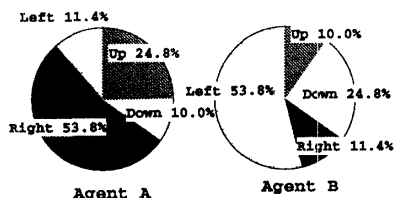


図2: エージェントの獲物捕獲位置の割合

実験2-1: 機能差の実験 (学習率)

実験2-1では、エージェントに機能差を持たせた場合の役割学習へ影響を検討する。機能差に Q-learning の主要なパラメータである学習率を用いる。学習率を  $\alpha_A = 0.2, \alpha_B = 0.4$  とする。

図3に獲物捕獲位置の推移を示す。4,000 試行目において、異種の系は役割分化しているが、同種の系では、まだ役割分化は見られず異種の系ほど役割学習の速度が速いことが分かる。

また、異種の系と同種の系のステップ数での比較を行なうと、常に同種の系より短いステップで獲物を捕獲し、役割学習の効率化による性能向上が見られた。

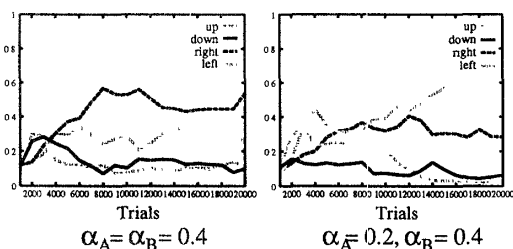


図3: 学習率異種及び同種の系におけるの捕獲位置の推移

実験2-2: 機能差の実験 (視野)

実験2-2では、エージェントの機能差として、視野を用いる。環境の大きさを  $20 \times 20$  とする。視野の深さは、 $d_A = 2, d_B = 4$  とし、同一視野のエージェント系 ( $d_A = d_B = 3$ ) と比較実験を行なう。

図4に獲物捕獲位置の推移の比較を示す。異種の系では役割が明確に現れ安定しているが、同種の系では、役割獲得後に変化が見られ役割が不安定である。

実験3: 余剰エージェントの実験

実験3では、余剰のエージェントを加えた場合のシステムの振る舞いを検討する。ここでは、ハンターエー

ジェントの数を3とし、実験1の結果と比較する。

図5より、エージェント C には明確な役割学習が見られたが、エージェント A, B は行動に差異が無く役割学習が見られなかったことが分かる。

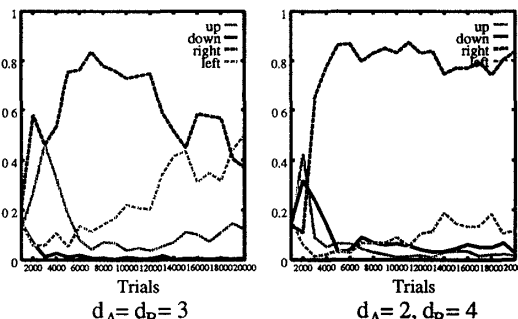


図4: 視野を制限した場合における獲物捕獲位置の推移

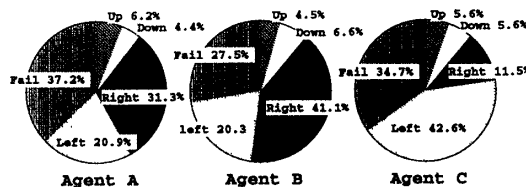


図5: 3つのハンターエージェントの役割

6 おわりに

本報告では、マルチエージェントによる協調行動獲得のため、役割学習に関する基礎計算機実験を行なった。その結果、以下の結論を得た。

- 獲物捕獲行動を学習したエージェント系では、役割学習が見られた。
- 機能差の異なる系を構成することによって、各エージェントの行動の違いが明確に現れ、役割学習が容易になる。
- 余剰エージェントを加えた場合、役割学習に悪影響を及ぼす。

参考文献

[1] Watkins.C.J.C.H, and Dayan.P, Technical Note:Q-Learning, *Machine Learning*, Vol.8, No.3, pp.279-292, 1992.  
 [2] Tan.M, Multi-agent Reinforcement Learning: Independent vs. Cooperative Agents, *Proceedings of the Tenth International Conference on Machine Learning*, pp330-337,Morgan Kaufmann, 1993.  
 [3] 河石 勇, 山田 誠二, 豊田 順一, 異種学習エージェント系における経験の共有と学習効率, *MACC'95 Online Proceedings* 研究会資料シリーズ No.1, 1995.