

# フェロモン情報を用いたマルチエージェント システムの協調学習\*

1 X-3

加藤 新吾      松尾 啓志      岩田 彰†

名古屋工業大学電気情報工学科‡

## 1 はじめに

複雑かつ動的な問題を多くの自律的なエージェントが何らかの協調動作をすることで解決しようと、マルチエージェント研究が進められている。本研究ではエージェント間の情報伝達的手段としてフェロモンを用いた協調学習の一方法を提案し、その効果を考察する。なお、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 Profit Sharing を採用し、行動選択としてルール選択を採用する [1]。

## 2 追跡問題

$n \times n$  格子状トラスの環境を設定し、複数のハンターエージェントと獲物エージェントをランダムに配置し初期状態とする。各エージェントは、予め決められた順番で行動し、上下左右の方向に一マス進むかまたは停止の行動をする。同一のマスに複数のエージェントが存在することは許されない。

ハンターの視界は  $m \times m (n \geq m)$  で、自分の周囲  $m^2 - 1$  マスに含まれる獲物・自分以外のハンターの情報と、現在自分が位置するマスに置かれているフェロモン情報の2種類の情報を獲得可能である。また、自分以外の個々のハンターの区別はしない。

獲物の視界は  $1 \times 1$  で、上下左右の直線方向のみを観測でき、視界内にあるハンターから遠ざかる方向に逃げる行動を選択する。視界内にハンターが存在しない場合にはランダムに行動する。

全てのハンターが獲物に隣接した状態を目標状態とし、全てのハンターに報酬が与えられる。

## 3 フェロモン情報の利用

### 3.1 エージェント間の情報伝達

エージェント間の情報伝達的手法としてリアルタイム通信を行う方式がある。しかし、この手法は扱う問題が

大きくなるにつれて通信コストが激増するという欠点がある。そこで、本研究ではエージェント間の情報伝達的手段としてフェロモン情報を用いることを提案する。

### 3.2 フェロモン情報

フェロモンは環境を観察したエージェントによって反射的に放出され、いったん放出されたフェロモンは生成したエージェントから独立した物質となり、環境中から徐々に蒸発していく。本稿においてフェロモンが持つ情報は、フェロモンが生成されたとき獲物がどこにいたかという情報とフェロモン自体の量（フェロモンが生成されてからどれだけの時間が経ったかという情報）である（図1）。

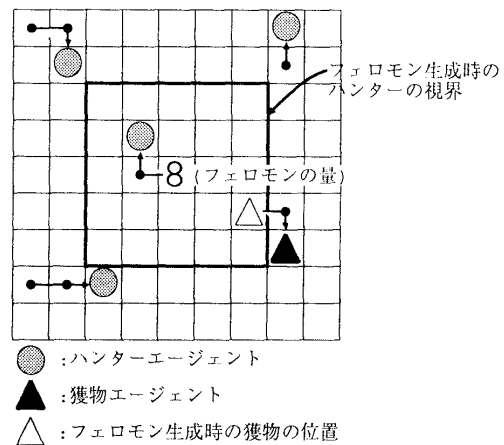


図1: フェロモン情報 (2ステップ前に置かれたフェロモンの量とその情報について)

### 3.3 フェロモン情報の利用

獲物の位置が分からない状態ではハンターエージェントが有効な戦略を取ることは困難である。つまり、追跡問題においてハンターエージェントが何よりも早急にとるべき行動が自らの視界内に獲物を納めることである。

本稿では、ハンターエージェントは視界内に獲物がいるときは現在の視界を感覚入力として利用し、視界内に

\*Cooperative Learning for Multiagents System using Pheromone

†Shingo Kato, Hiroshi Matsuo, Akira Iwata

‡Department of Electrical and Computer Engineering, Nagoya Institute of Technology

獲物がないときは自分がいるマスのフェロモン情報を感覚入力として利用するハイブリッドな状態識別手法を提案する(図2)。本手法により、視界内に獲物がない時には獲物に近づく行動を、視界内に獲物がいるときには他のハンターエージェントと協調的に獲物エージェントを追い込む行動を選択することが期待できる。

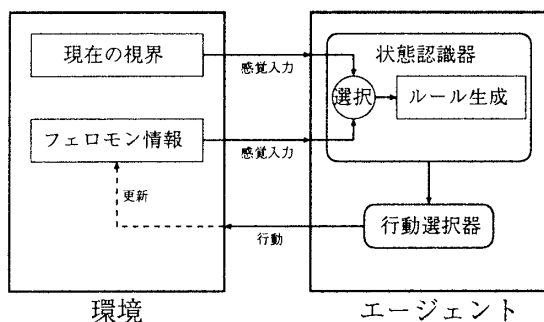


図2: 感覚入力の選択

## 4 実験

### 4.1 実験方法

9×9格子状トラスの環境にハンター4体と獲物1体をランダムに配置する。エピソードごとに、ハンターと獲物はランダムに配置し、100万エピソードの繰り返しを1試行とする。

ハンターのパラメータとして視界を $m \times m$ 、各ルールの初期重みを0.1、報酬を1に設定し、公比0.2の等比減少関数を強化関数に採用した。また、フェロモンの量は生成されたとき10で、1ステップ毎に1ずつ減少していく。獲物の視界は5×5とする。

### 4.2 実験結果および考察

次の3つの条件下でそれぞれ10試行実験を行なった。それぞれの学習曲線を図3に示す。

- 視界5×5、フェロモン使用せず
- 視界7×7、フェロモン使用せず
- 視界5×5、フェロモン使用(提案手法)

図3から、フェロモンを使用した場合は200エピソード位で使用しない場合よりも早く目的状態に達することが確認できる。視界が5×5の場合に比べて約2倍広い7×7の場合と比較した場合でも、ほぼ同程度の政策に収束可能と同時に収束速度が早いことが分かる。

また感覚入力の状態数を比較した場合、視界を5×5から7×7に広げた場合に、状態数が10倍以上も増加しているのに比べて、フェロモンを使用した場合はわずか1%の増加にとどまっている。この点から、フェロモ

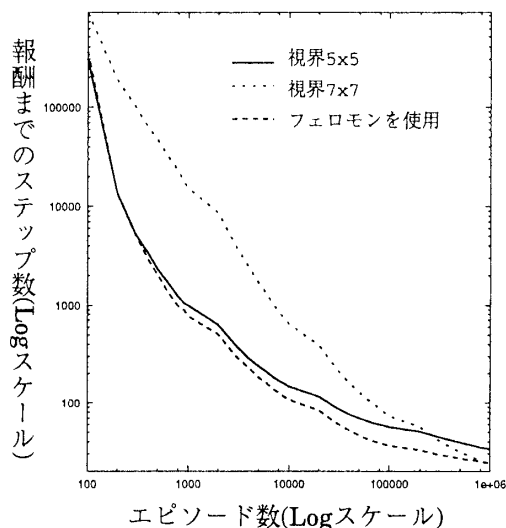


図3: 学習曲線(報酬に至るまでのステップ数とエピソード数の関係)

	感覚入力の状態数	状態数の比 (aを1とした場合)
a	390625	1
b	394625	1.01
c	5764801	14.76

表1: 感覚入力の状態数の比較

ンを使用することにより状態空間を効率よく利用できていることが分かる。

## 5 まとめ

本稿では、エージェント間の情報伝達にフェロモン情報を用いることを提案した。また、実験によりフェロモン情報が有効に利用されていることを示した。

## 参考文献

- [1] 荒井 幸代, 宮崎 和光, 小林 重信, マルチエージェント強化学習の方法論-Q-Learning と Profit Sharing による接近-, 人工知能学会誌, Vol.13, No.4, pp.609-617, July.1998.
- [2] 川村 秀憲, 山本 雅人, 鈴木 恵二, 大内 東, エージェント間の協調動作に基づく分散的探索手法, 第6回マルチ・エージェントと協調計算ワークショップ (MACC'97)