

マルチエージェントを用いた組織形成に関する基礎的研究

乾 岳史*

櫻井 成一朗**

*東京工業大学大学院社会理工学研究科社会工学専攻

**東京工業大学大学院情報理工学研究科計算工学専攻

自律エージェントを複数協調させることによって問題解決を図るマルチエージェントシステムの研究が行われているが、その多くは問題解決の効率向上を目標としたものであり、組織形成の仕組みについては明らかになっていない。本研究では、マルチエージェントシステムを用いて動的組織形成の仕組みを明らかにすることを目的とする。コンピュータシミュレーションにより、エージェント間の通信路の結合状態と静的組織形成との関係について調べた結果、組織としての秩序の形成には至らないことが明らかになった。そこで、環境の状態に依存して形成される動的組織形成に注目し、通信プロトコル、強化学習のプロセスと組織形成との関連についてシミュレーションすることによって、組織形成に対する基礎的知見を得ることができた。

A basic study on formation of collaborative behavior in a multi-agent pursuit game.

INUI Takeshi*

SAKURAI Seiichiro**

* Department of Social Engineering, Tokyo Institute of Technology

** Department of Computer Science, Tokyo Institute of Technology

This paper focuses on dynamic formation of collaborative behavior in a multi-agent pursuit game. The dynamic formation of collaborative behavior is achieved by the re-inforcement learning through the interaction with the environment. It is shown that the weighted rules function as triggers for the emergence of commanders which direct the other agents.

1 はじめに

既存の組織理論では、環境と組織形態との適応関係について多くの研究がなされているが、その多くは実証研究であり、組織形態を形成するプロセスについては解明が進んでいない。

これに対して、人工知能の分野では、環境との相互作用を考慮したマルチエージェント系の研究成果が近年増加している [4, 5, 6, 7, 8]。個々の適応エージェント (以下エージェント) としては単純な能力しか持たないエージェント群でも、環境中で行動させ、相互に影響を与えながら複雑な組織的行動が創

発されるのである。しかし、目標達成という観点から協調行動の創発を研究したものがほとんどで、組織の形成について研究したものはほとんどない。

エージェントが協調行動をとるためには、何らかの形で情報を共有する必要がある。情報共有の方法の一つとして、エージェント間でのメッセージ交換がある。通常は、設計者が環境に対する知識からメッセージを設計し、メッセージの意味付けも同時に行う。文献 [1] では、強化学習の環境適応性に注目して、強化学習を用いてエージェント間で各メッセージの意味付けを共有し、協調行動をとることを明らかにしている。更に、回線構造とメッセージ種類数

について実験をおこない、部分的に制限のある回線構造が有効であることを述べている。また、協調行動のためにはある程度のメッセージ種類数が必要であることも述べている。

文献 [2] では、環境を動的、静的とに区別し、静的環境下では通信回線によらない学習で十分に問題解決を行うことが可能であることを明らかにしている。しかし、動的環境下では、組織をアプリアリに与えて、協調的行動が起こることがあるとだけ述べており、組織形成には至っていない。文献 [3] でも、組織的行動と呼べるような行動は学習できていない。

本研究では、エージェントがメッセージの意味付けを学習して、協調行動を創発する過程を検証することで、組織形成の基礎的知見を得ることを目的とする。

2 マルチエージェントシステム

2.1 適応エージェント

エージェントとは、自ら得た情報を基に、自ら判断を下す事によって、自律的に行動する個体を表す。豊富な知識を与えることも無く、得られる情報も限られている場合には、各エージェントの達成可能な目標も限られており、多くの場合その単純な目標の達成さえも失敗してしまう。個々のエージェントの能力が限定的であっても、集団として協調することができれば、より高度な目標達成が可能となる場合がある [6,7]。すなわち、個の知ではなく、集団の知を実現する事によって、より高度な目標達成を可能とするのがマルチエージェント系である。

2.2 協調行動

エージェントは個別に環境からの情報を得て、それをもとに行動を決定する。マルチエージェントとして集団の知を実現するためには、エージェント間で情報をやりとりし、エージェント同士の競合を解消し、全体として有効に機能するようにしなければならない。マルチエージェント系の協調を実現するための方法として、強化学習 [4,5,8,9] がある。強化学習では、各エージェントに対してあらかじめ協調機能を備えることなく、環境から得た報酬によって協調行動が創発される。

本研究では、メッセージに対して何ら意味を持た

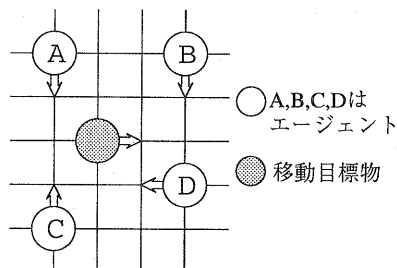


図 1: 2次元追跡問題

せず、エージェントは学習によってメッセージの意味共有を進める。メッセージの意味共有とは、エージェントの発したメッセージが特定の状況での特定の行動に対応しているということが当事者間で理解され意思伝達が可能になることである。協調行動とは、意味共有がなされたメッセージを活用して、個々のエージェントが適切に行動し、問題を解決していく行動である。意味共有も学習するエージェントでは、意味共有とそれに基づく行動の学習は同時に進んでいくことになる。エージェントの協調行動が形成されることを、マルチエージェント系で組織が形成されたものとする。

2.3 2次元追跡問題

2次元追跡問題とは、次のようなゲームである。無限に広がる平面上に、4つのエージェントと移動目標物が存在する。エージェント群が移動目標物を4方から取り囲み捕捉することがゲームの目的である。各エージェントは2次元の格子状の空間を移動し、エージェントと移動目標物は交互に移動を繰り返す。このエージェントと目標物の動きを1サイクルとする。エージェントが目標物の上下左右に位置し、目標物が移動できないときエージェントは目標物を捕捉したとする。エージェントは規定のサイクルのあいだに目標物を捕捉しなければならず、捕捉できなければそのゲームは失敗となる。

本研究では、論文 [1] と同様、エージェント D は他のエージェントにない特殊な能力を持ち、D が目標物に隣接した場合には、目標物は移動できなくなる。

2.4 意思決定

各エージェントは、環境の情報を知るためのセンサーとメッセージ通信のための通信回線を持つ。また、センサーから得られた情報と通信回線上のメッセージに対応して意思決定するためのルールを持つ。このルールは、(s,b,i,o)の組で表し、重みを強化学習によって割り当てる。sはセンサー値、bは行動目標、i/oは入/出力メッセージを表し、各々整数値によって与える。

このルールは、図2に示されるような形で表すことができる。表の各要素は、行動-送信メッセージの組とそのルールに対する重みの値である。エージェントは、センサー情報と受信メッセージによって決まるルールのリストの中から、一番重みの重いルールを自分の意思決定として選択する。

初期状態のルールの重みは均等に与えられるので、ルールリストからどのルールに基づいて行動するかは無作為に決定される。後述する強化学習を用いた各ルールに対する重み付けがなされると、各エージェントの行動が変化し、ある状況では特定の行動をとるようになる。したがって、各エージェントが協調して目標を達成するための知識を陽に与えずとも、環境との相互作用によって目標達成が可能になる。

エージェントのセンサーの状態の数S、メッセージの種類数M、行動の数Bとした時、エージェントの持つルールの数は $S \times (M+1) \times B \times M$ である。強化学習の結果、特定の状況で特定の意思決定がなされるようになった時点では、実際に意思決定によって選択されるルールの数は、およそ $S \times (M+1)$ となる。なお、本論文では学習後実際に意思決定によって選択されるルールのことを有効ルールと呼ぶ。

2.5 メッセージ通信

文献[1]をもとに、メッセージ通信は次のように行われるものとした。

1. 各エージェントは、各サイクル毎にセンサーの情報を更新する。
2. この時点では、参照すべきメッセージはまだ発せられていないので、センサー情報のみから、意思決定する。次に、その意思決定に用いられたルールの重みをメッセージの重みとして、回線に送信する。

		センサー信号			
		0	1	2	4
メ ッ セ ー ジ	なし	左 8:-0.10	右16:-0.05	下11:-0.15	
		左 4:-0.10	左16:-0.05	左11:-0.15	...
		
	1	下 7:-1.10	右14:-4.95	下 1:-1.19	
		左16: 842	右 1: 0.00	右 2: 0.00	...
		下 2: 0.00	右 2: 0.00	右 3: 0.00	
	2	
		左 5:-1.05	左12:-2.45	右 1:-0.99	
		上15:-0.10	上15:-1.09	左14:850.0	...
	3	右 4:-0.10	左16:-1.15	右 9:-0.20	
		
		右 6:-1.09	左 9:-2.45	右 7:-1.19	

図 2: ルールベース

3. 次に、エージェントは参照可能な通信回線があれば、メッセージを参照する。参照可能な回線を複数持つ場合には、メッセージの重みの重い方を参照する。そして、自分が前の意思決定の重みと比較し、自分の発したメッセージの方が重ければ、過去の決定を維持し、新しい意思決定は行わない。そうでなければ、受け取ったメッセージとセンサーから意思決定を再度行い、4に移る。
4. 新しい意思決定をし、メッセージを送信する。送信メッセージの重みは新しい意思決定のルールの重みと参照したメッセージの重みの和である。3に戻る。

全てのエージェントが新しい意思決定をしなくなった時点で、そのサイクルの最終的な決定とする。

2.6 強化学習

強化学習とは、特定の行動に対してだけ報酬を与えるという試行を繰り返すことによって、特定の条件下で特定の行動をとるという行動パターンが強化される、条件付けと呼ばれる適応現象の模倣である。

本研究では、文献[1]と同様に、環境から局所的報酬と大域的報酬の2種類の報酬を与え、各エージェントが環境から得た報酬に従ってルールの強化を行う。

報酬の配分は、エージェントの行動系列が目標達成に寄与したと考え、行動系列で用いられたルール

それぞれに対して公比 $1/2$ の強化関数による Profit-sharing 法 [8] を用いて報酬を分配する。通信でメッセージを参照したエージェントから発信したエージェントへも、公比 $1/2$ で報酬を伝播させる。その結果、目標達成に寄与したと考えられるルールは、メッセージのやり取りの方向にも強化される。メッセージの伝達に対して報酬を伝播させることで、メッセージを用いた行動の強化が促される。

2.7 意味共有・競合解消・協調行動

エージェントの間で意味共有が行なわれるとは、エージェント間でやりとりされるメッセージがそれぞれのエージェントの意思に基づいて発信され、受信者がその意思を認識して行動することである。ただし、エージェントの意思、認識は観察者が解釈した結果であり、エージェントが環境に対して具体的に認識しているわけではない。

目標達成によって報酬が分配され、行動の強化が進むと、特定の状況の特定のメッセージに対して特定の行動があらわれる。ルールからみるとセンサーとメッセージの組に対して行動目標と送信メッセージの組が 1 対 1 に対応するようになった状況である。そのような状態をエージェント間での意味共有がなされた状態と考え、エージェントがある意思の元でメッセージを送信し、その意思を反映して受信者が行動決定していると解釈できる時、エージェント間で意味共有に基づく組織が形成されたとする。

競合解消とは、メッセージ通信によってエージェントが同一の行動目標に進むのを回避することである。競合解消の学習はエージェントの意味共有と同時に進むが、意味共有を通して組織形成が行なわれたことが競合解消に結び付くという保証は必ずしもない。

協調行動とは、意味共有によりメッセージを用いた競合解消が可能になった結果創発されるエージェント系全体としての行動のことである。競合解消と同様に、目標達成に寄与するとは必ずしも限らない。

本論文ではエージェントの行動は、各サイクル毎に決定されるので、エージェント系に創発される協調行動は各サイクルのメッセージ通信の過程と結果に表れる。エージェントは、メッセージが送信されていない状態では、センサー情報のみから意志決定するがその行動はエージェント間での調整がなされていない。エージェントはその調整をメッセージ交

換によって行い、この調整過程全体が、学習によって強化されている。

3 メッセージ通信に基づく協調行動

本章では、メッセージ通信によって創発された協調行動を観察し、エージェント間でやり取りされるメッセージに対する意味付けを行う。その上でルールに対する考察により、協調行動がどのように創発されるかについて述べる。

3.1 実験条件

行動目標 エージェントの環境に対する行動は、文献 [1] を参照し、エージェントが目標物の上下左右のどこを目標点とするかを決定することとする。

実際の移動方向の決定は次のように行なった。

- 行動目標とセンサーの値から、プログラムされた方向に移動する。
- センサー情報だけで、目標点にいることは分からない場合もあるが、ここではエージェントが目標点にいる場合にはそこにとどまるものとした。
- 目標物に隣接しているが、目標点とは反対の地点にいる場合など、移動方向の可能性が複数ある場合には、それらの中から無作為に選ぶ。

したがって、基本的にエージェントは目標に近づいていくことになる。終盤においては目標点によってはエージェントから遠ざかる移動もありえる。

報酬 報酬は文献 [1] を参照し、以下のような大域的報酬と局所的報酬を与えるものとした。

大域的報酬は、ゲームの最後に与えられる。ゲームの目標を達成した場合には、正の報酬 y を与える。ゲームの目標を達成できなかった場合には、まず、エージェントが目標物に隣接しているかどうか判断する。隣接していない場合には、そのエージェントと目標物との距離 d をもとに、 $z(0 < z < -y)$ の値を与える。隣接している場合には、目標位置の競合があるかどうか判定し、目標の競合がある場合には、負の報酬 w を与える。局所的報酬は、各サイ

表 1: エージェントに与える報酬の設定

$$\begin{aligned}
 x &= 0.05 \\
 y &= 20.0 \\
 z &= y \times \left(\frac{2}{1 + \exp((d-1)/y)} - 1 \right) \\
 w &= 0.05
 \end{aligned}$$

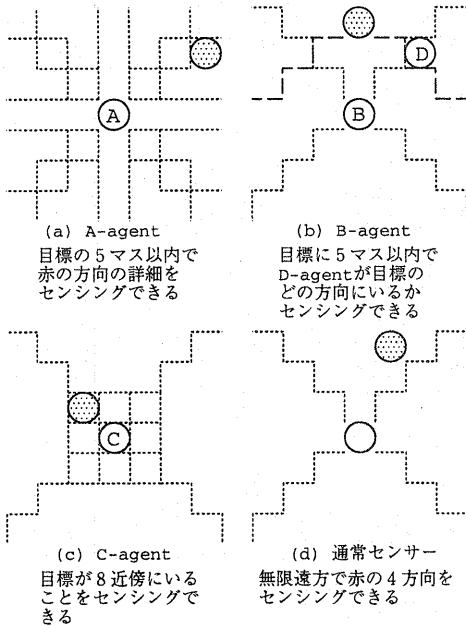


図 3: エージェントのセンサーの設定

クルごと、エージェントの目標物との位置が縮まっていなければ、負の報酬 $-x$ を与える。具体的な値は、表 1 に示した。

センサー エージェント A, B, C は表 2 に示すような異なる能力を持ったセンサーを持つ。エージェント D は特別なセンシングは行わず、通常センサーの情報のみ受け取る。

エージェントの持つセンサーは、それぞれセンサーの能力、センサーのとり値が異なるので、エージェントの行動もその影響を受けるものと考えられる。

回線構造 エージェント系の回線構造は図 4 のように全てのエージェントが自分以外の全てのエージェントのメッセージを参照可能な設定とした。

エージェント間でやり取りされるメッセージの種

表 2: エージェントのセンサーの設定

通常センサー	目標物の方角を上下左右で 4 方向を知覚。値は 0 から 3 の値をとる。
エージェント A	通常センサーの他、目標物に 5 マスまで近付いた時点で、目標物の位置について図 3(a) に示すような 16 方向を知覚。値は 0 から 19 の値をとる。
エージェント B	通常センサーの他、目標物に 5 マスまで近付いた時点で、自分から見た目標物の方角とあわせて、エージェント D が目標物からみてどちらの方向にあるかを知覚。値は 0 から 19 の値をとる。
エージェント C	通常センサーの他、目標物の図 3(c) に示す 8 近傍にいることを知覚。値は 0 から 11 の値をとる。

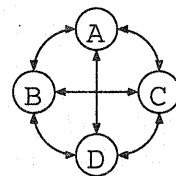


図 4: エージェント間の回線構造

類数は 64 に設定した。これは各エージェントが自らの移動目標をバケツリレー方式に伝送する際には、最大 2^6 ビットの情報が必要となるので、その伝送を可能とするために 64 種類とした。

3.2 動的組織形成

予備実験においては、静的な組織の形成を確認することができなかったため、本研究ではメッセージ交換を通して行動が規定される動的組織の創発について焦点を絞り考察する。

シミュレーション結果として実際に観測された行動を図 5 に示す。左右がそのサイクルのそれぞれのエージェントと目標物の位置である。右はエージェントがメッセージ交換を行う前の目標点を示し、左はメッセージ交換後の目標点を示す。間のグラフは、

表 3:

サイクル	動的に形成された組織
サイクル 12	D(左 43) → B(右 43)
	B(右 43) → C(上 34)
	C(上 34) → A(上 19)
サイクル 13	D(左 43) → B(上 11)
	B(上 11) → C(下 38)
	C(下 38) → A(左 51)
サイクル 14	D(左 43) → A(上 10)
	A(上 10) → C(右 1)
	C(右 1) → B(左 60)
サイクル 15	D(左 43) → B(上 11)
	B(上 11) → C(右 64)
	C(右 64) → A(下 63)
サイクル 16	D(左 43) → A(上 43)
	A(上 43) → B(右 43)
	B(右 43) → C(右 20)
サイクル 17	D(左 43) → A(上 10)
	A(上 10) → C(右 1)
	C(右 1) → B(左 60)

一つの要素が移動目標と送信したメッセージの組で示されるそれぞれのエージェントの意思決定を表す。グラフの矢印は矢印の元のメッセージが参照されたことを示す。図 5 に示した行動では、表 3 に示すように、各サイクル毎にまたメッセージ交換の 1 ステップ毎に動的に組織が形成されていることを観察できる。

表 3 に示したように、各々のサイクルにおいて三つの動的組織が形成されていることがわかる。

サイクル 17 でのエージェント D から A へのメッセージ送信では、D の発したメッセージを受けて A が適当な方向へ移動するという行動を見ることができ、このような形で D の発したメッセージと現在の自分の位置から考えると現在位置にとどまるべきだ、というメッセージとして意味共有ができていてと解釈することができる。A からのメッセージに応じて C が競合解消した行動目標を決定し、C のメッセージによって B が決定するという、意味共有の結果みられる動的組織の行動によって、エージェント全体としての目標達成が可能になっている。

各サイクルの 3 つの組織においては、18 組織中 2 個の組織を除き、指令を発したエージェントの目

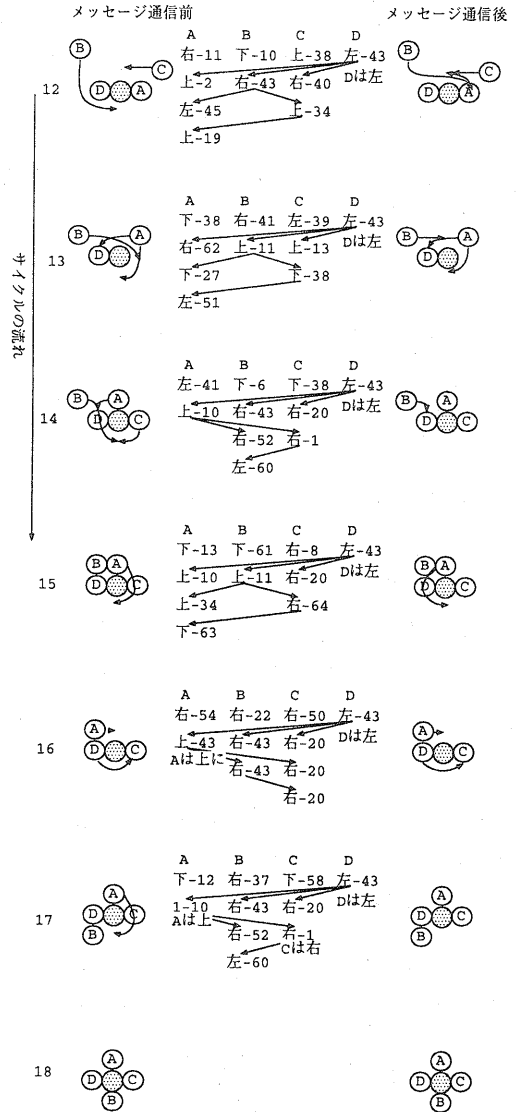


図 5: 学習されたエージェントの行動

標とは異なる目標に設定されていることがわかる。この事実は各小組織の内部では、競合解消を行うことで目標が達成しやすくなるであろうという期待の下で、協調行動を行っていると思えることができる。これはまた各小組織内で情報が閉じており、局所的な判断しか下せていないということの証でもある。この実験における組織形成においては、同一環境下では重み最大のルールを持つエージェントが指令の発信者となり、更にその受信者の中で重み最大のルールを持つエージェントが次の小組織における指令の発信者となっていることがわかる。逆に指令の発信者たる能力を持つエージェントは重みの大きなルールを持つエージェントに限られており、センサー能力だけから指令の発信者として機能するエージェントは実験した範囲では確認できなかった。以上のことから、強化学習によって動的組織を形成するための機能がルールに対して重みを与えることによって実現されていると考えることができる。

次にこの動的組織形成をルールの具体的重みから検証する。表4に、この図5の局面で使われる有効ルールを示す。表の項目は、参照メッセージ*i*、行動*b*、送信メッセージ*s*、ルールの重みで、エージェント名の横にセンサー値を示した。表に示さなかったルールは、重みが0か負のもので、有効な学習がなされた後であればメッセージ通信の過程では他のエージェントには参照されないエージェントがこのルールに基づいて行動した結果が、図5である。

エージェントが最初に目標を達成するのは、確率的結果である。その時に使用された発信元のルール、それを参照したルールが強化され、次に同じ状況になった時には、そのルールが選択されるようになる。エージェントの強化はこのように進む。そこで、あるルールに基づく目標達成が繰り返されれば、ルールが強化されて大きな重みを持つようになり、その状況でのエージェントの発言力が強まる。各ルールは環境依存性であるので、それ以外の状況での発言力とは無関係である。

学習後のエージェントの行動は確定的で、特定の状況では必ずそれに応じた行動が行われる。これらの行動はエージェント群全体の行動に対する報酬に基づく強化の結果であるが、その行動が環境に適應していることは保証されない。目標を捕捉するという目的の達成率の違いは、確定的に捕捉できるような状況がどの程度起こるか、ということによって決定される。

表 4: 有効ルール

A センサー値:7			
i	b	o	重み
2	1	20	7.50
6	1	54	310.06
19	1	13	43.05
20	3	54	16.75
21	1	1	19.35
23	1	30	17.50
25	1	20	3223.7
29	1	20	1712.9
32	1	8	17.30
35	1	24	6021.26
38	1	36	19.10
41	1	5	35.90
42	1	54	976.65
43	1	10	1336.5
48	1	37	1369.4
60	1	52	111.34
62	1	7	5079.01
64	3	63	14.90

B センサー値:14			
i	b	o	重み
1	2	60	893.40
9	1	53	563.28
10	0	52	468.60
13	0	15	135.95
17	0	62	79.80
26	1	64	134.35
31	2	23	759.51
43	0	43	145.16

C センサー値:8			
i	b	o	重み
6	0	6	8292.82
10	0	1	905.40
11	0	64	25.50
13	0	19	17.80
20	0	22	476.35
43	0	20	140.10
59	0	49	114.35
64	0	28	1951.9

D センサー値:2			
i	b	o	重み
0	2	43	38136.1
7	1	31	0.05
38	2	45	19.80
46	0	19	0.05

エージェントがメッセージ通信によって行っているのは、自分のセンサー情報を送り、また、相手のセンサー情報を受け取ることで、過去の捕捉できた時の行動パターンを再現することである。それぞれのエージェントの行動目標には競合があり、メッセージ通信によっても、競合解消が必ず行なわれるという結果は得られなかった。これはエージェントの移動目標位置が競合していても、移動の過程で捕捉した状況になれば良い、という問題設定であるという理由も考えられる。

4 結論

強化によって、環境の状態と行動が対応するようになった時点では、エージェント間のメッセージ通信も特定の対応がつけられている。このような状態がエージェントで意味共有が可能になった状態と考えられる。

組織はエージェントが意味共有によって結び付いたものとした。エージェントの組織形成は、偶然成功した試行で使われたメッセージが役に立つもの

とされ、それが繰り返されるうちに、状況に応じたメッセージのやりとりが生まれる。その状況に応じたメッセージのやりとりに基づく行動がエージェントの組織行動である。ここで形成される組織は偶然有効とされたメッセージによって調整されるものであり、形成された組織によって、有効な問題解決が行なわれることは保証されない。

メッセージ通信の起点となるエージェントは、センサー情報のみから意思決定する。このエージェントの決定は、エージェント全体の行動の行動方針と見ることもできる。起点となるエージェントの決定はセンサーのみから決定されるので、センサー情報を伝えているとみることもできる。エージェントは、メッセージに対する学習によって、環境の状態をより多く知ることができるようになる。

メッセージの有効性の一つは、メッセージ通信を行うことで、エージェントが得られる情報が増えるからと考えられる。メッセージはエージェント同士のセンサーの拡張という役目を持ち、また有効ルールの増加をもたらし、系全体として安定して動くことが可能になる。エージェントは学習によって、他のエージェントの状況を間接的に知ることができ、特に終盤においてはその情報を競合解消に役立てることができる。

本研究の課題としては、形成された組織の協調行動の環境への適応の度合いが低かった。エージェントの通信方法、学習方法には限界があり、まだ改良の余地がある。さらに、本研究で扱ったゲームでは、目標達成に協調行動が必ずしも必要ではなかったため、より複雑なゲームに基づく研究が必要とされることも指摘しておく。

参考文献

- [1] 松浦・嘉数 非均質 agent 系における組織的行動の形成: 2 次元追跡問題における考察 情報処理学会論文誌 Vol.38 No.6 pp.1083-1093,1997
- [2] 佐伯 素彦 組織的行動の自律的形成に関する基礎的研究-multi-agent に拠るシミュレーション実験- 東京工業大学修士論文,1998
- [3] 乾 岳史 Multi-Agent による協調行動の自律的形成に関する基礎的研究- 2 次元追跡問題によるシミュレーション- 東京工業大学卒業論文,1999
- [4] 畝見 達夫 強化学習「最近の機械学習」人工知能学会誌 Vol.9 No.6 pp.830-835,1994
- [5] 畝見 達夫 強化学習エージェントの集団行動 MACC'93 pp.137-150,1993
- [6] 大沢 英一 協調プランニングの動的組織再編とメタレベル整合戦略-追跡ゲームにおける考察- コンピュータソフトウェア Vol.12 No.1 pp.511-519,1995
- [7] 大沢 英一 問題空間の変化に適応する協調的組織スキーマ-追跡ゲームによる考察- MACC'92 pp.105-120,1992
- [8] 荒井・宮崎・小林 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 pp.609-617,1998
- [9] 山村・宮崎・小林 エージェントの学習 人工知能学会誌 Vol.10 No.5 pp.683-689,1995
- [10] 山田 誠二 適応エージェント 共立出版株式会社,1997
- [11] 石田 亨,片桐 恭弘,桑原 和宏 分散人工知能 コロナ社,1996