

模倣学習と共に進化に関する研究

村上 幸一[†] 佐藤 浩[†] 生天目 章[†]

† 防衛大学校 情報数理学科
〒239-8686 神奈川県横須賀市走水 1-10-20

E-mail: †{g40046, hsato, nama}@nda.ac.jp

あらまし 本論文では、個々の相互作用がタカ・ハトゲームとして表されるエージェント社会での、共進化の問題を取り扱う。個々のエージェントがハト派戦略及びタカ派戦略をもち、そしてランダムな方法で選択された相手と、相互作用をする場合には、ハト派戦略とタカ派戦略を一定の割合でとる状態、すなわち進化的に安定な戦略(ESS)の割合の分布状態で、均衡することが知られている。しかし、この安定した状態は最適な均衡解ではない。本論文では、各エージェントが相互に学習を行うことにより、全員がハト派戦略をとるようになり、最も望ましい状態になることを示す。また、エージェントが、どのようなメタルールを獲得することで、そのような最適解を定理するように共進化をしたのか、そのプロセスを明らかにする。

キーワード タカ・ハトゲーム、模倣学習、共進化、人工社会

Co-evolutions through Imitation Learning

Yukikazu MURAKAMI[†], Hiroshi SATO[†], and Akira NAMATAME[†]

† Dept. of Computer Science, National Defense Academy
Hashirimizu 1-10-20, Yokosuka-shi 239-8686, JAPAN

E-mail: †{g40046, hsato, nama}@nda.ac.jp

Abstract It is an interesting question to answer how the society gropes its way towards efficient equilibrium in an imperfect world when self-interested agents learn each other in order to improve the rules of interaction. In this paper, we focus on the co-evolution of meta-rules of strategic interactions in the negotiation situations. We formulate negotiation situations as hawk-dove games. It is known the mixed strategy will result in an efficient equilibrium in hawk-dove games. In this paper, we introduce co-evolutionary dynamics with local matching and investigate the role of collaboration learning. We show that all agents gradually learn to behave as doves, which result in social efficiency.

Key words hawk-dove games, mutual learning, mimicry, local matching, co-evolution

1. はじめに

生物の進化は、その集合体である社会の進化にどのような影響を及ぼしているのであろうか。その疑問を解き明かす研究として、近年、人工社会や、マルチエージェント・システムなど、さまざまな研究が行われている[1][2]。これらの研究の多くは、生物（以後、計算機上で自律的に振舞う主体を指すものとし、エージェントと呼ぶ。）がもつ、生き残るために戦略（以後、単に戦略と呼ぶ。）を、いくつかのパターン（例えば、“必ず協調する”“必ず裏切る”“しつべがえし”など）に既定し、それらを対戦させることによって、社会の進化の傾向をつかむ研究が多い。しかし、すべてのエージェントが、それぞれ固有の戦略をもつ場合に、エージェント同士を対戦させると、どのような社会へと、進化していくのか？そのような社会の進化に関する研究は、まだ少ないようだ。

そこで、本論文ではそのような戦略をもったエージェントが、どのように共進化していくかに、焦点をあてて研究を行う。また、エージェントの学習方法として、模倣学習を用いることで、社会的に効率的な状態が、形成することができることを明らかにする。特に、短期的な視点と、長期的な視点からの模倣学習について検討し、それら2つの学習の差異についても明らかにしていく。

2. タカ・ハトゲーム

2.1 タカ・ハトゲームの定義

エージェントが多数存在する社会において、各エージェントは1対1で相互作用する。（これを「対戦」と呼ぶ。）本論文では、これらの対戦をタカ・ハトゲームとして、定式化する。

タカ・ハトゲームでは、各エージェントが対戦時に取りうる戦略として、次の2種類を設定する。

タカ派戦略 S_1 ：傷つくか相手が逃げ出すまで、戦いを挑む。

ハト派戦略 S_2 ：争う気がないことを主張し、相手に戦いを挑まれるなら逃げる。

タカ・ハトゲームの利得行列を、表1に示す。各エージェントは、利益Vを目の前にし対戦を行う。また、対戦に負ることによる損失をCとする。ここで、損失が利益を上回る（ $C > V$ ）ものとする。も

し、タカ派戦略をとる者がタカ派戦略をとる者と出会うとき、お互いに目の前の利益を獲得するために争いを始める。その結果、一方は勝利し利益Vを獲得し、また一方は敗北し損失Cを被ることになるので、この対戦での期待利得は、 $(V - C)/2$ となる（図1.(c)）。この場合、期待利得は負になる。

ハト派戦略をとる者がハト派戦略をとる者と出会うとき、お互いに誇示牽制し合い、そして相手が強く出たときには逃げる。したがって、このケースでの期待利得は $V/2$ となる（図1.(d)）。

一方で、タカ派戦略をとる者とハト派戦略をとる者が出会うとき、ハト派戦略をとる者は逃げるので、タカ派戦略をとる者はVの利益を獲得し、ハト派戦略をとる者の利得はゼロになる（図1.(a), (b)）。

表1の利得行列の、混合戦略の下での均衡解は（図1.(e)）の点で求まる。

進化プロセスにより、どのような均衡解が選択されるかは重要なことであるが、以下に代表的な事実を述べる。

（1）進化的に安定な戦略（ESS）

ESSとは、集団を構成しているすべての個体が、その戦略を採用しているときに、突然変異によって

表1. タカ・ハトゲームの利得行列 ($0 < V < C$)

The other's Own's strategy strategy	S_1 (Hawk)	S_2 (Dove)
S_1 (Hawk)	$(V-C)/2$	0
S_2 (Dove)	V	$V/2$

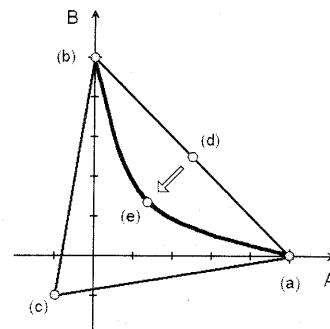


図1. タカ・ハトゲームの利得関係

生じる他のどのような戦略も、その集団の中に進化的に侵入できないとする戦略のことである。[4]そして、ここでは混合戦略によるナッシュ均衡解と同じであり、(図1 (e))の点にあたる。しかしながら、ESSは最適解ではない。

(2) リプリケータ・ダイナミクス (RD)

RDは、個々のエージェントの相互作用が、進化ゲームと呼ばれるゲーム論的な枠組みで与えられているときに、そのマクロ特性を時間発展作用素として導くことのできる手法である。

エージェントは、二つの代替案、例えばタカ・ハトゲームの場合には $\{H, D\}$ という代替案の中から、意思決定を行い、その選択比率に応じた形で適応度が決まり、増殖する。[5]

RDも、ESS(そして、混合戦略によるナッシュ戦略)して収束することが知られている。

3. エージェントの社会モデル

本論文もふくめた、エージェントベースのコンピュータモデリングは、その中に登場するエージェントが、“変化のない外生的な効用関数”をもつ合理的な行為者であるため、現実の生物、特に人間とは隔たりがあることは確かである。しかし実社会での実験が困難である以上、非常に有用な実験方法の一つであることは、間違いない。

3.1 エージェント社会の構成要素

本論文で行うコンピュータシミュレーションには、エージェント、環境、ルールという3つの基本的な構成要素が絡んでいる。まずこれらについて説明する。

(1) エージェント

エージェントとは、計算機上で自律的に振舞う主体を指すが、エージェントは、内部状態と行動ルール(後述)を持っている。エージェントの内部状態には、そのエージェントの一生を通じて一定な、先天的といえるものもあれば、外部環境や他のエージェントとの相互作用を通じて変化する、後天的といえるものもある。

このシミュレーションで先天的な状態としては、エージェントの名前(かつ、人工社会での存在位置)などがそれにあたるだろう。

また、後天的な状態として、エージェントは、これまでの対戦履歴、獲得した資産、戦略を決定する

知識(戦略決定表)などが挙げられる。

エージェントは、進化の手法として模倣学習を用いる。模倣学習とは、自分の周囲において最も成功した者の戦略をうまく真似るというやりかたである。本シミュレーションでも同様に、エージェントは、周囲でもっとも成功しているエージェントの戦略決定表を、まるごとコピーする(図2)。

このように、相手との関わり合いの中から、何かを学びとり、自らの行動様式を逐次改良していくことは、広義の意味での進化といえよう。こうした進化の過程において、模倣学習は、生物が進化していくための極めて有効な手段であると言える。

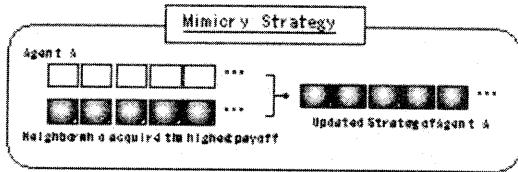


図2. 模倣学習

本研究では、短期的な模倣学習及び長期的な模倣学習の2つの視点からシミュレーションを行う。短期的な模倣学習では、周囲8近傍の中で、最も利得の高いエージェントの戦略(タカ派戦略もしくは、ハト派戦略)を次回の戦略とする。これに対して、長期的な模倣学習では、エージェントは以下のようないくつかの内部構造を持つ。(図3)

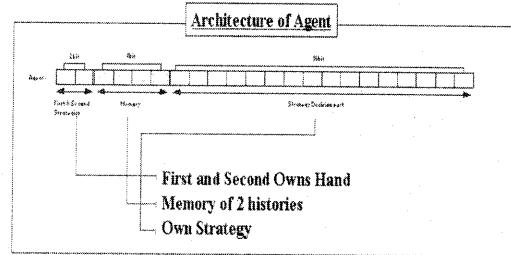


図3. 相互作用のためのメタ・ルール

各エージェントの内部構造を、22ビットからなる配列($P_i, i \in [1, \dots, 22]$)で表現し、各配列の構成は次のように定める。

- ・第1ビットおよび第2ビット(P_1, P_2)は、各世代における1回目および2回目に出す戦略
- ・第3ビットから第6ビット($P_3, P_4 \in [3, \dots, 6]$)は、過去2回分の、自分と相手の対戦履歴

・第7ビットから第22ビット($P_i, i \in [7, \dots, 22]$)は、戦略決定表をあらわし、3回目以降の対戦に参照される。この部分の初期値に乱数を入れることによって、各エージェントに、ランダムな戦略決定表を与えている。

また図4に、戦略決定表と対戦履歴との関係についての一例を示す。

No	Strategy of the past				strategy at t	
	t-2		t-1			
	Own	Opp	Own	Opp		
0	H	H	H	H	H	
1	H	H	H	D	D	
2	H	H	D	H	H	
3	H	H	D	D	D	
4	H	D	H	H	H	
5	H	D	H	D	D	
6	H	D	D	H	H	
7	H	D	D	D	D	
8	D	H	H	H	H	
9	D	H	H	D	D	
10	D	H	D	H	H	
11	D	H	D	D	D	
12	D	D	H	H	H	
13	D	D	H	D	D	
14	D	D	D	H	H	
15	D	D	D	D	D	

図4 戰略決定表（メタ・ルール）

(2) 環境

ここで環境とは、エージェントが相互作用する空間を示す。それゆえ、エージェントは環境の中に存在しているが、環境はエージェントとは別個の媒体を指す。本モデルでは、場を格子状に構成してモデル化する。

具体的には 20×20 の二次元平面に、最大400人のエージェントを配した社会を考える。さらに、その左右の端と上下の端をつなげ、どのエージェントも等しく相互作用できるようにする。

(3) 行動ルール

エージェントにはいくつかの行動ルール（以下、ルール）がある。

ここでは、自分の周りにいるエージェントと相互作用をするための“対戦ルール”および、周辺のエージェントから学び、より高い利得を得るために“学習ルール”をもつものとし、対戦と学習のサイクルをもって1世代とした。以下に、各ルールの詳細を示す。

対戦ルール：

- ・自分の一つ前に他のエージェントがいるなら、1対1で対戦を行なう。
- ・対戦でとる戦略は、自分の過去2回の対戦履歴を参照して戦略決定表より、これを選ぶ。
- （短期的模倣学習では、前回、周囲で最も利得の高かったエージェントの戦略を選択）
- ・これを繰り返し、1世代に一人と20回の対戦を行なう。（短期的模倣学習では、1回）
- ・同様に、この1対1の対戦を周囲の8近傍のエージェントと行なう。

学習ルール：

- ・その世代に對戦を行なった相手エージェントの中で、最も多い資産をもつエージェントを最も良い決定をしているエージェントとして認識する。
- ・そのエージェントの戦略決定表を参考に学習を行なう。
- ・学習法は完全模倣を用いる。

3.2 利得パラメータの設定

シミュレーションには、ハト派戦略とタカ派戦略どちらも等しくとった時に均衡する場合を境に、以下の3つのパラメータを用いる。

ケース(1) $(V, C) = (2, 10)$

(1)は、対戦による利得が少なく、そして戦いのコストが高い場合でハト派戦略が有利な場合。

ケース(2) $(V, C) = (2, 4)$

(2)は、どちらの戦略も等しくとった時に均衡する場合で、ハト派戦略とタカ派戦略に有利不利がない場合。

ケース(3) $(V, C) = (10, 12)$

(3)のケースは、対戦で獲得できる利得が多く、損失コストは少ないので、タカ派戦略が有利な場合である。

4. 模倣学習のシミュレーション結果

4.1 短期的な視点からの模倣学習

以下に、短期的な模倣学習におけるシミュレーション結果を示す。ケース(1)及びケース(2)では、初期配置を、すべてのエージェントがタカ派戦略をとり、中心の一人だけがハト派戦略をとるように、また、ケース(3)では、エージェントの3/4がハト派戦略を

とり、残り $1/4$ がタカ派戦略をとるように配置した。

ケース(1)

対戦による利得が少なく、戦いのコストが高いケース(1)では、エージェントが学習を行わずに、自己の利益を追求すると、ハト派戦略をとるエージェントが全体に占める割合 80% (以降、ハト派 $\ast\%$ と表す)、タカ派 20% でナッシュ均衡となり、このとき、対戦において各エージェントの獲得できる利得の和は 0.8 になる。

これに対し、シミュレーション結果では、30世代までにすべてのエージェントがハト派戦略をとるよう進化した(図5)。この場合、双方のエージェントは 1 の利得を獲得する。これは、各エージェントが次第にお互いに争うことを嫌うようになることを示し、社会的に効率的な状態である。

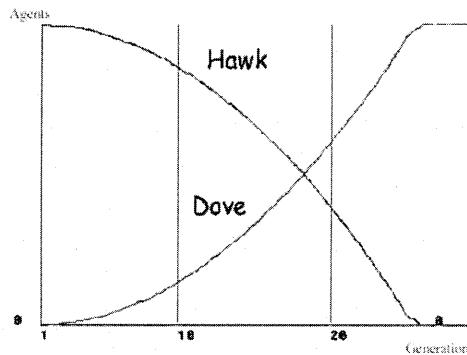


図5. $(V, C)=(2, 10)$ における戦略比

ケース(2)

また、ハト派戦略とタカ派戦略に有利不利がない場合ケース(2)では、学習を行なわない場合の均衡は、タカ派戦略とハト派戦略、どちらの戦略も等しくとったときで、獲得利得は 0.5 となる。

一方、シミュレーション結果では、30世代までにほとんどのエージェントが、ハト派戦略をとるように進化した(図6)。しかし、ケース(1)よりも目の前の利益が大きいため、完全に収束するには至っていない。そしてそれぞれ、概ね 1 の利得を得る結果となった。これも同じく、各エージェントが次第にお互いに争うことを嫌うようになることを示し、社会的に効率的な状態であると言える。

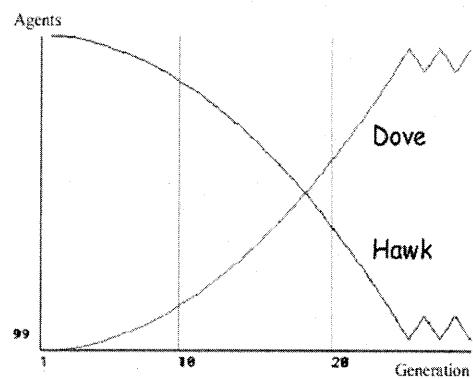


図6. $(V, C)=(2, 4)$ における戦略比

ケース(3)

タカ派戦略が有利なケース(3)では、学習を行なわない場合の均衡は、タカ派 83% 、ハト派 17% のときで、獲得利得は 0.83 となる。

しかしシミュレーションの結果では、5世代までに、すべてのエージェントがタカ派戦略をとるようになった(図7)。これにより、すべてのエージェントは -1 の損失を得るだけになり、争いを避ける事で得られた 5 の利得を逃す結果となつた。

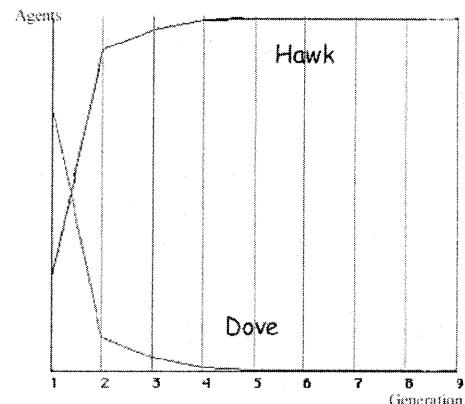


図7. $(V, C)=(10, 12)$ における戦略比

4.2 長期的な視点からの模倣学習

以下に、長期的な模倣学習のシミュレーション結果を示す。初期状態では、すべてのエージェントはランダムな戦略決定表を持つ。また、1世代に行なう20回の対戦のうち、ハト派戦略とタカ派戦略をとつた割合で、エージェントを5等分にグループ分け

した。

ケース(1)

対戦による利得が少なく、戦いのコストが高いケース(1)の場合について、長期的な模倣学習でのシミュレーション結果を図8に示す。

このケースでは、ほぼ20世代を超えたあたりで、すべてのエージェントがハト派戦略をとるようになり短期的な模倣戦略と同様に、社会的に効率的な状態になった。また、400種類あった、エージェントの戦略は長期的な学習によって、表2に示す6種類に集約された。また、各エージェントは、15番目の戦略において必ずハト派戦略をとるようになった。

ケース(2)

ハト派戦略とタカ派戦略に有利不利がない場合ケース(2)の場合について、長期的な模倣学習でのシミュレーション結果を図9に示す。

このケースでは、短期学習の場合、すべてのエージェントがハト派戦略にはならず、完全に1には収束

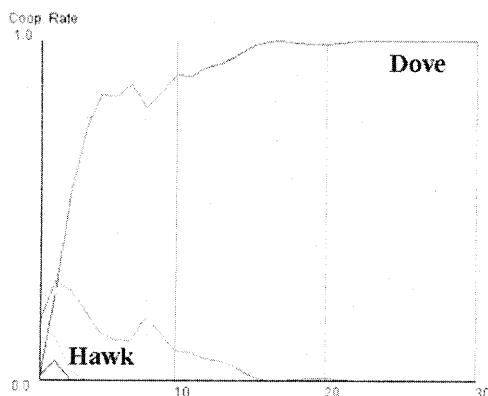


図8. 長期的な模倣学習における人口比 : (V, C)=(2, 10)

表2. 共有された6種類のメタ・ルール

	TYPE 1	TYPE 2	TYPE 3	TYPE 4	TYPE 5	TYPE 6
0	H	H	H	H	H	H
1	H	H	H	H	H	H
2	H	H	H	D	H	D
3	H	H	D	H	H	H
4	D	H	H	D	D	D
5	D	H	D	D	H	D
6	H	H	D	H	H	D
7	H	D	D	H	H	H
8	D	H	H	D	D	D
9	D	H	D	H	H	H
10	D	H	H	H	D	D
11	D	H	H	H	H	D
12	D	D	H	D	D	D
13	D	H	H	H	D	D
14	D	H	H	D	D	H
15	D	D	D	D	D	D
agent number	113	139	74	70	2	2

しなかつたが、長期的な学習を行うことにより、約10世代すべてのエージェントがハト派戦略をとるようになり、社会的に効率的な状態になるに至った。また、400種類あった、エージェントの戦略は長期的な学習によって、表3に示す10種類に集約された。また、各エージェントは、15番目の戦略において必ずハト派戦略をとるようになった。

ケース(3)

短期的な模倣学習では、タカ派戦略が社会全体を占めることになったケース(3)の場合について、長期的な模倣学習によって、シミュレーションした結果を図10に示す。

短期的な模倣学習と同様、ここでも、当初タカ派が有利な状況であるため、タカ派を選択するエージェントが多いが、第5世代の付近で逆転し、30世代付近では、すべてのエージェントがハト派を選択するようになった。これは、短期模倣学習の際の利得-1はもちろん、ナッシュ均衡における0.83も大きく

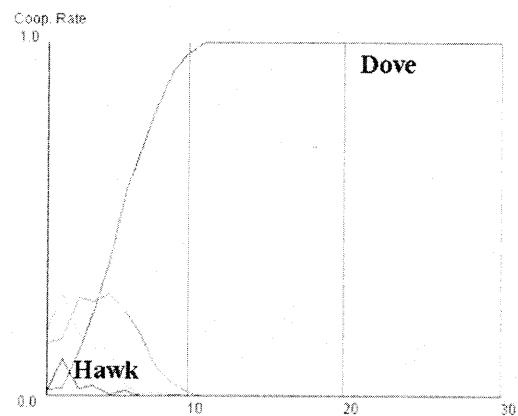


図9. 長期的な模倣学習における人口比 : (V, C)=(2, 4)

表3. 共有された10種類のメタ・ルール

	TYPE 1	TYPE 2	TYPE 3	TYPE 4	TYPE 5	TYPE 6	TYPE 7	TYPE 8	TYPE 9	TYPE 10
0	H	H	H	H	H	D	D	D	H	H
1	H	H	H	D	H	H	H	D	D	D
2	H	H	D	H	H	H	D	H	H	H
3	H	D	D	H	H	H	D	D	H	H
4	D	H	H	H	D	D	H	H	D	D
5	D	H	H	H	D	H	H	H	H	H
6	H	D	D	D	H	H	D	D	D	D
7	D	H	D	H	H	H	H	H	D	H
8	D	H	H	D	D	H	H	H	H	D
9	D	H	D	H	H	H	H	H	H	H
10	D	H	H	H	H	H	H	H	H	H
11	D	H	H	H	H	H	H	H	H	H
12	D	H	H	H	H	H	H	H	H	D
13	D	H	H	H	H	H	D	H	H	H
14	D	H	H	H	H	H	D	D	H	H
15	D	H	H	H	H	H	D	D	D	D
agent number	84	56	33	17	11	95	15	27	53	9

凌ぎ、利得 5 を獲得する。よって、ハト派戦略が有利である場合と同じように、エージェントが次第に争いを避けるように共進化し、また、社会的に効率的な状態を構成することとなった。また、400 種類あったエージェントの戦略は長期的な模倣学習によって表 4 に示す 3 種類に集約された。そして、各エージェントは、15 番目の戦略において必ずハト派戦略をとるようになる。

次に、ケース(3)において獲得された 3 つのメタ・ルールを、再度均等な数だけランダムに配置し、対戦を行ったところ、表 5 のような結果が得られた。これから、TYPE 2 と、TYPE 3 のメタ・ルールでは、本来 TYPE 3 の方が有利であるにもかかわらず、TYPE 1 が存在することによって、TYPE 2 の戦略がもっとも優勢になっていることがわかる。

このように、共同学習の後に獲得された戦略は、ただメタ・ルールを獲得しただけでなく、相補関係が見られる場合もある。

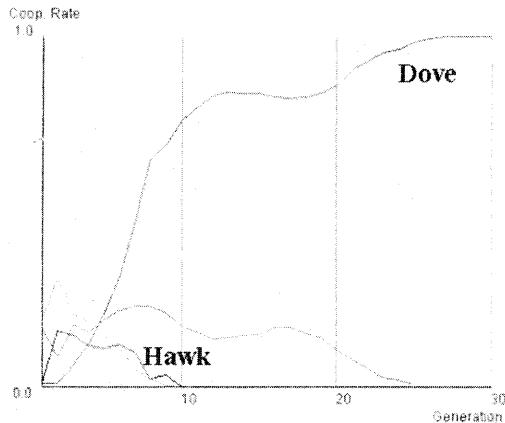


図 10. 長期的な模倣学習における人口比 : (V, C)=(10, 12)

表 4. 共有された 3 種類のメタ・ルール

	TYPE 1	TYPE 2	TYPE 3
0	HHHH	H	D
1	HHHD	H	H
2	HHDH	D	H
3	HHDD	D	D
4	HDHH	H	H
5	HDHD	D	H
6	HDDH	H	D
7	HDDD	H	H
8	DHHH	H	H
9	DHHD	D	H
10	DHDH	H	D
11	DHDD	D	D
12	DDHH	H	D
13	DDHD	H	D
14	DDDH	H	H
15	DDDD	D	D
agent number	93	285	22

表 5. メタ・ルールの相補関係

Strategy Rate	Strategy Type	TYPE1	TYPE2	TYPE3
TYPE1 vs TYPE 2 vs TYPE 3 (33%)	0	292	108	
TYPE1 vs TYPE 2 (50%)	47	353	—	
TYPE 2 vs TYPE 3 (50%)	—	133	267	
TYPE 3 vs TYPE 1 (60%)	38	—	362	

5 考 察

エージェントは、模倣学習によって、社会的に効率的な状況をつくり出した。しかし、短期的な学習では、争って奪い取るだけの高い価値のものがあるときには、自らの利益に走るものが出てしまい、それに引きずられる形で、全員が利己に走る社会が形成されてしまった。

しかし、学習の方法を変え、長期的な模倣学習を行なったところ、同じく目前に争うだけの価値のあるものが存在する場合でも、社会的に効率的な状態を形成できた。すなわち模倣学習という限定された方法で、社会全体を最良の状態にするまで共進化させることができた。

また、長期的な模倣学習においては、初期状態において 400 種類あった各エージェントの戦略決定表は、相互作用の結果、数種類に集約された。

ここで、獲得された戦略決定表は、1 種類に集約されるのではなく、数種類の戦略決定表をもつエージェントに集約され、各タイプのエージェント間で均衡状態を保つのは興味深い。これは、世代数を増やしてみても同じような結果が得られ、戦略決定表は数種類に収束したが、それ以上集約されることはない。また、集約されたメタ・ルールには、表 5 のような相補関係が見られる場合もあり、このことから、シミュレーションによって得られたメタ・ルール間では、相関関係が存在するとも考えられる。

また、すべてのケースで、獲得された戦略決定表は、15 番目の戦略において、必ずハト派戦略を選択するという共通のルール（メタ・ルール）を獲得している。

この 15 番目の戦略は”再帰点”でになっており、また、すべてのエージェントがこのルールを学習することにより、確率 1 で必ず再帰する”吸収点”となる。この吸収点の存在により、各エージェントは、ひとたび互いにハト派戦略を選択すると、じ後、ハト派戦略取り続けることとなり、長期的に安定した高い利得を得られることができる。このようなこと

から、社会は共進化へと向かうと考えられる。

このように、長期的な模倣学習では、メタ・ルールを獲得することによって、短期的な模倣学習では社会的に効率的な状況を作り出せないような場合でも、効率的な社会状態を形成することができることがわかった。

6 おわりに

模倣学習によって、エージェントは社会的に効率的な状況をつくり出すことができた。また、長期的な模倣学習は、メタ・ルールを獲得することによって、短期学習では社会的に効率的な状態とならない場合においても、効率的な社会へと共進化させることができた。そして、このメタルールの学習プロセスにおいて、吸収点の存在が非常に大きな意味を持つこともわかった。このような模倣学習による共進化において、獲得したメタルールによっては、他の手法よりも高い協調状態を築くことができるることは興味深い。

一方で、本研究でのエージェントは、移動することなく、初期に配置された位置にとどまる固定的な社会に存在した。これを、移動型の社会にかえることによって、エージェントの共進化にどのように影響を及ぼすかということは、今後の課題である。

文 献

- [1] Esptein J.M. and Axtell R, “人工社会”, 服部正太, 木村香代子, 構造計画研究所, 1999
- [2] Richard J.Gaylord and Louis J.D'Andria: *Simulating Society*, Springer-Verlag New York, Inc. 1998.
- [3] ドーキンズ R, “利己的な遺伝子”, 日高敏隆, 岸油二, 羽田節子, 垂水雄二訳, 紀伊国屋書店, 1991
- [4] Smith, J. M. and Price G.R.: The logic of animal conflict, Nature, 246, pp.15-18, 1973
- [5] 出口 弘, “複雑系としての経済学”, pp.87-102, 日科技連出版社, 2000.
- [6] Murakami Y., Sato H., and Namatame A., “Co-evolution in Negotiation Games”, “Fourth International Conference on Computational Intelligence and Multimedia Applications no24, pp241-245, Yokosuka, Japan, Oct.2001.
- [7] Murakami Y., Sato H., and Namatame A., “Co-evolution in Hawk-Dove Games in a Mobile Environment” The 5th Australia-Japan Joint Workshop on Intelligent & Evolutionary Systems, Otago, New Zealand, pp 19-25 2001
- [8] Uno K., and Namatame A., "Evolutionary Behaviors Emerged through Strategic Interactions in the Large", *GECCO'99 Workshop on Artificial life*,