

協調追跡のための提携における利得構成の進化的学習

佐藤秀樹、渥美雅保

創価大学工学研究科情報システム学専攻

hsatoh@t.soka.ac.jp, matsumi@t.soka.ac.jp

あらかし複数のエージェントが協力してタスクを解決する場合、タスクをより合理的に解決するエージェント集団の形成が必要となる。協力ゲームの理論における提携は、合理的なエージェント集団の形成に対する指針を与えている。我々は、複数のハンターエージェントが複数の獲物を追跡する拡張追跡タスクを例として、どの追跡タスクについて誰といかなる利得配分で提携を形成するかという問題と、提携のもとで追跡行動を選択する問題とを、遺伝的プログラミングに基づき並行的に学習するアーキテクチャを提案した。本研究では、提携に基づく協調追跡行動が進化的に学習可能であることをシミュレーション実験により明らかにする。

キーワード 提携、遺伝的プログラミング、ゲーム理論

Evolutionary Learning of Payoff Configurations in Coalition for Cooperative Pursuit

Hideki SATO, Masayasu ATSUMI

Division of Info. Sys. Sci., Graduate school of Eng., Soka University

hsatoh@t.soka.ac.jp, matsumi@t.soka.ac.jp

Abstract In multi-agent environment, it is important to make a group of agents that solve each task rationally. The concept of coalition, which is proposed in the cooperative game theory, provides the rational criteria for making agents' group. We propose an evolutionary learning architecture based on the genetic programming that learns a rational coalition among agents and pursuit behavior under the coalition concurrently. In this paper, by using the extended pursuit problem as an example, we show that coalitions are formed and the effective pursuit behavior appears under the coalition.

key words Coalition, Genetic Programming, game theory

1 はじめに

複数のエージェントが協力してタスクを解決する場合、タスクをより合理的に解決するエージェント集団を形成することが必要となる。協力ゲームの理論における提携は、合理的なエージェント集団の形成に対する指針を与えている。

我々は、複数のハンターエージェントが複数の獲物を追跡する拡張追跡タスクを例として、どの追跡タスクについて誰といかなる利得配分で提携を形成するかという問題と、提携のもとで追跡行動を遂行する問題とを、遺伝的プログラミング[1]に基づき並行的に学習するアーキテクチャを提案した[2]。

本研究では、提携とそれに基づく協調追跡行動が進化的に学習可能であることをシミュレーション実験により明らかにする。

2 拡張追跡タスク

拡張追跡タスクは、追跡タスク[3]を次の2点で拡張したものである。

- (1) 無限に広がる2次元格子上に、2つの獲物エージェント(獲物)と6つのハンターエージェント(ハンター)がいる。
- (2) あるハンターがある獲物を捕らえたときに得られる利得は、ハンターによって異なり、かつ捕らえる獲物によっても異なる。

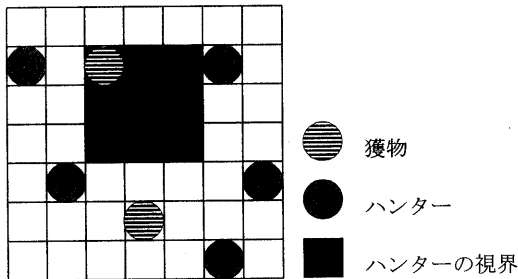


図1 拡張追跡タスク

各ハンターは一度に1グリッドずつ移動し、動かずに留まることも可能である。各グリッドには、同時に複数のエージェントが入ることはできない。また、

各ハンターには視界が等しく与えられており、獲物の4近傍を囲むことで獲物を獲得できる。また、獲物は常にランダムに行動する(図1)。

ハンターの目的は、獲物を捕らえることにより、できるだけ多くの利得を獲得することであり、そのためには、エージェント間提携の形成と追跡行動との学習が各エージェントに要求される。

3 提携公募アーキテクチャ

3.1 提携

提携[4]とは、1つの協力単位として行動することが可能なグループであり、形式的にはプレイヤー全体の集合の部分集合をいう。即ち n 人のプレイヤーの集合 $N = \{1, 2, \dots, n\}$ に対して、その部分集合 $S (\subseteq N)$ を提携と呼ぶ。

3.2 交渉集合型特性関数

提携形成においては、各ハンターがより多くの利得を得られる合理的な提携を形成することが重要である。ここで、合理的であるとは、その提携のもとでの各ハンターへの利得配分が、各ハンターにとって同意できる個人合理的利得構成であることをいう。このような個人合理的利得構成はM安定であるといわれる。このとき、各提携の提携値は、交渉集合型特性関数により決定される[4]。

拡張追跡タスクにおける交渉集合型特性関数を次のように定式化する。ハンターの集合を N 、許容提携を $B (\subseteq N)$ 、提携 B の持つ共同戦略の集合を $S_B = \{S_{p_1}, S_{p_2}\}$ (ただし、 p_1, p_2 は獲物で、 S_{p_i} は B のメンバーが獲物 p_i を追跡することを表す)、提携 B に属するプレイヤー h_i の利得関数を $f_{h_i}: S_B \rightarrow \mathbb{R}$ (実数)、 $f_{h_i}(S_{p_1}) = v_{i1} \in \mathbb{R}$ 、 $f_{h_i}(S_{p_2}) = v_{i2} \in \mathbb{R}$ とする。このとき、交渉集合型特性関数を次のように定義する(式1)。各ハンターエージェントは、この特性関数のもとで、できるだけ大きな提携値を与えてくれる提携を形成しM安定な利得配分を達成すべく交渉を行う。

$$v(B) = \max_{S_{p_j} \in S_B, h_i \in B} \sum f_{h_i}(S_{p_j})$$

$$= \begin{cases} 0 & (|B| < 4) \\ \max_{h_i \in B} \{ \sum_{h_i \in B} f_{h_i}(S_{p_1}), \sum_{h_i \in B} f_{h_i}(S_{p_2}) \} & (|B| = 4) \\ \max_{B' \subset B, |B'|=4} \max_{h_i \in B'} \{ \sum_{h_i \in B'} f_{h_i}(S_{p_1}), \sum_{h_i \in B'} f_{h_i}(S_{p_2}) \} & (|B| > 4) \end{cases}$$

式1 交渉集合型特性関数

3.3 提携公募アーキテクチャ

エージェントの合理的提携行動は、交渉集合型特性関数が与えられれば求めることも可能である。しかし、他人の利得関数が未知であるような不完全情報下では、エージェントは特性関数をあらかじめ知ることができない。従って、各エージェントには合理的な提携、即ちM安定な利得配分を、提携行動を通じた環境との相互作用の中で試行錯誤的に学習することが要求される。そこで、本研究では、提携形成の学習を行うために、次の提携公募アーキテクチャ（図2）を提案し、導入する。

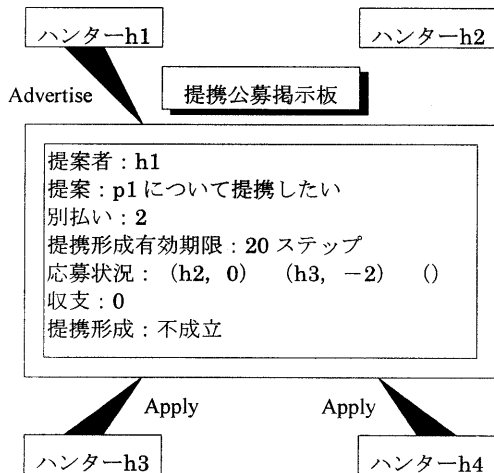


図2 提携公募アーキテクチャ

提携公募アーキテクチャにおいては、提携は、提携公募掲示板を介した提携の公募と応募によりなされる。即ち、ハンターは捕獲したい獲物に対する提携形成を公募し、その公募に対して他のハンター

が応募することにより提携が形成される。公募や応募をするときには、自らが支払ったり、もしくは受け取りたい別払い額を同時に提示する。応募者が4人になり、かつ別払い額の収支がプラスであるとき、それら4人のハンターエージェント間に提携が成立する。

4 提携における利得構成の進化的学習

4.1 利得構成

ハンターエージェントが、獲物を捕らえたときに獲得する利得は、提携を構成するメンバー、及び捕獲した獲物により異なる。

ハンター $h_i (i=1, \dots, 6)$ がある4人提携 B を形成して、ある獲物 $p_j (j=1, 2)$ を捕らえる場合の提携値 $v_{p_j}(B)$ 、及び各ハンターが獲得する利得 $v_{ij} = f_{h_i}(S_{p_j})$ は、ハンター h_i の獲物 p_j に対する価値 w_{ij} と、別払い額 $s_{B,i}$ により、次のように定められるとする。

$$v_{p_j}(B) = \sum_{h_i \in B} w_{ij}$$

$$v_{ij} = f_{h_i}(S_{p_j}) = f_{h_i}(w_{ij}, s_{B,i}) = w_{ij} - s_{B,i}$$

ここで、 w_{ij} はあらかじめ定められているものとし、 $s_{B,i}$ は別払いを支払うとき正、受け取る時負の値をとる。

これより、エージェントが、より多くの利得を獲得するためには、次の2つの要件が必要となる。

- (1) 提携値が大きい提携を形成する。即ち、 w_{ij} により決まる提携値がより大きくなる獲物とメンバーの組み合わせを求める。
- (2) 別払い額の収支が均衡するようにする。即ち、 $s_{B,i}$ により決まる収支が0となるような別払い額の組み合わせを求める。別払い額の収支が0となる時、 $v_{p_j}(B) = \sum_{h_i \in B} v_{ij}$ が成り立ち、 v_{ij} は提携 B のもとで利得構成を与えることになる。

即ち、2つの獲物 p_1, p_2 に対して、大きな提携値を与えるハンターの組合せが異なる場合には、それぞれの獲物に対して、適切な提携を行うことが大きな利得を得るための必要条件となるし、また、別払

い額の収支を0とすることにより別払い額を無駄にしないことが大きな利得を得るための必要条件となる。

4.2 利得構成の進化的学習

ハンターエージェントには、4.1より、より大きい提携値が獲得できる獲物とメンバーの組合せを求めること、及び提携のための別払い額の収支を0とする別払い額の組合せを求めることが要求される。そのため、利得構成の遺伝的プログラミングに基づく進化的学習においては、ハンター $h_i (i=1, \dots, 6)$ は、提携公募掲示板の公募・応募状況、及びグリッドワールドにおける獲物や他のハンターの位置に応じて、

- (1) どの獲物に対して公募するか、もしくは応募するか。
- (2) 別払い額をいくりにするか。

を決定する木(提携形成木)を保有し、その木の適用において形成された h_i を含む獲物 p_j に対する提携 B に対して、次の適合度関数

$$F = \frac{\sum_{h_i \in B} (w_{ij} - s_{B,i})}{4} = \frac{\sum_{h_i \in B} w_{ij} - \sum_{h_i \in B} s_{B,i}}{4}$$

の値をその木の適合度として評価する。ここで、 w_{ij} 、 $s_{B,i}$ は4.1で述べたとおりとする。第1項は、より大きい提携値を獲得できる獲物とメンバーの組合せを探索することにも貢献し、第2項は提携のための別払い額の収支を0とする別払い額の組合せを探索することにも貢献する。そして、この適合度関数の値を最大化するような提携を形成することを可能とする提携形成木を進化的に求める。

5 エージェントアーキテクチャ

追跡行動と提携形成を並行的に学習するエージェントのアーキテクチャを図3に示す。各エージェントは、提携形成のためのGP木(提携形成木)と追跡行動のためのGP木(追跡行動木)を持つ。

提携形成木は、環境状態(グリッドワールドにおける獲物や他のハンターの位置)と提携公募掲示板

の公募・応募状況を観測し、ある獲物の捕獲に関する提携の公募または応募を提携公募掲示板に対して行う木である。一方、追跡行動木は、環境状態の観測と提携公募掲示板に記されている提携が成立しているかどうかの情報に基づき、東・西・南・北のいずれかに移動するか、その場にとどまるかの選択を行う木である。

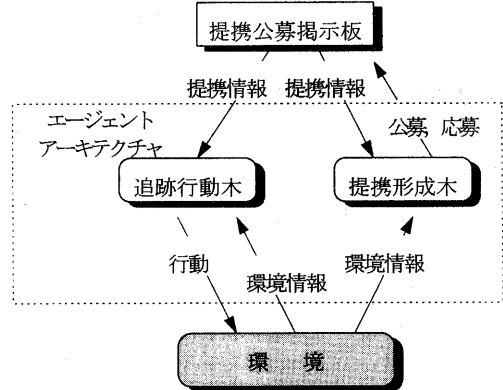


図3 エージェントアーキテクチャ

6 実験

6.1 提携の学習

まず、合理的な提携の学習が可能であることをシミュレーション実験により明らかにする。そのため、ハンターは、獲物の追跡は行わず、出現する2つの獲物のいずれに対して、いかなる別払いのもとで提携すべきかについてのみの学習を行う。

6.1.1 実験の枠組み

本実験では、6つのハンターが2つの獲物のいずれかに対して形成する提携の評価を行う。表1に、2つの獲物に対する各ハンターの価値を示す。別払い額として支払ったり、受け取ることができる額は、各価値の10, 20, 30, 40%のいずれかとする。表2に、獲物 p_0 に対する、各ハンターの価値、および各ハンターが提携を公募したときに獲得可能な最大提携値、及び、そのときの収支を示す。例えば、表2のハンター0の欄は、次のことを示している。獲物 p_0 に対して、ハンター0が公募し、ハンター1, 2, 3が応募することにより、ハンター0, 1, 2, 3の

提携が形成される場合、提携値は、各ハンターの獲物 p_0 に対する価値の合計、即ち、180となる。この提携において、ハンター0の公募における別払い額が24（獲物に対する価値の40%）、ハンター1,2,3の応募における別払い要求額がそれぞれ、10,8,6（獲物に対する価値の20%）の場合、別払い額の収支が0となる。関数と終端記号および実験パラメータは表3、表4のように設定した。なお、本実験では、lil-gp[5]を拡張して用いている。

	hunter0	hunter1	hunter2	hunter3	hunter4	hunter5
獲物 p_0	60	50	40	30	20	10
獲物 p_1	10	20	30	40	50	60

表1 価値

	hunter0	hunter1	hunter2	hunter3	hunter4	hunter5
価値	60	50	40	30	20	10
最大提携値	180	180	178	150	100	-
収支	0	0	2	0	0	-

表2 価値と提携値

Function	
who_ad_p0	公募ハンターにより分岐
who_ad_p1	公募ハンターにより分岐
ap_num_p0	応募ハンターの数により分岐
ap_num_p1	応募ハンターの数により分岐
ap_pay_p0	公募別払い額と応募別払い額の差により分岐
ap_pay_p1	公募別払い額と応募別払い額の差により分岐
bal_p0	収支と応募別払い額の差により分岐
bal_p1	収支と応募別払い額の差により分岐

Terminal	
ad_10_p0	別払い額10%で公募する
ad_10_p1	別払い額10%で公募する
ad_20_p0	別払い額20%で公募する
ad_20_p1	別払い額20%で公募する
ad_30_p0	別払い額30%で公募する
ad_30_p1	別払い額30%で公募する
ad_40_p0	別払い額40%で公募する
ad_40_p1	別払い額40%で公募する
ap_10_p0	別払い額10%で応募する
ap_10_p1	別払い額10%で応募する
ap_20_p0	別払い額20%で応募する
ap_20_p1	別払い額20%で応募する
ap_30_p0	別払い額30%で応募する
ap_30_p1	別払い額30%で応募する
ap_40_p0	別払い額40%で応募する
ap_40_p1	別払い額40%で応募する
ignore	何も行わない

表3 関数と終端記号

ステップ数	50	木の最大深さ	7
提携形成有効期限	15	交叉率	0.8
人口	100	再生率	0.19
世代	1000	突然変異率	0.01

表4 実験パラメータ

6.1.2 実験結果

図4に、50ステップ（回）の提携公募もしくは応募行動のうち、獲物 p_0 の提携のために各ハンターが公募した回数の変化、図5に獲物 p_1 の提携のために各ハンターが公募した回数の変化を、10世代ごとに平均して示す。図6に、図4の公募に対する各ハンターの応募回数の変化、図7に、図5の公募に対する各ハンターの応募回数の変化を、10世代ごとに平均して示す。公募に対しては、必ずいずれかのハンターにより応募がなされ、提携が成立するため、図4と図5の公募回数は、提携成立回数にもなっている。

また、表5は各獲物に対して形成された提携において、最大提携値を獲得した提携での各ハンターの利得構成を示している。

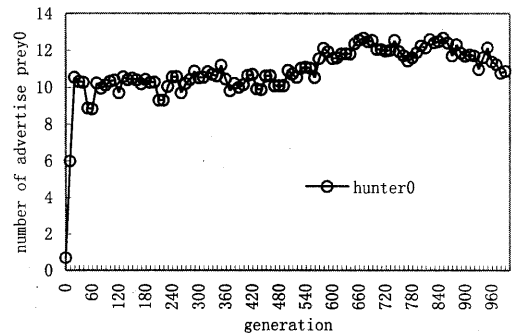


図4 獲物 p_0 に対する公募回数

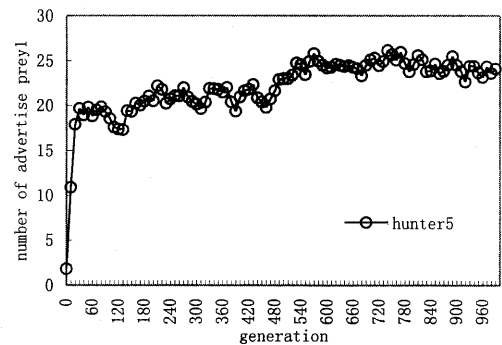


図5 獲物 p_1 に対する公募回数

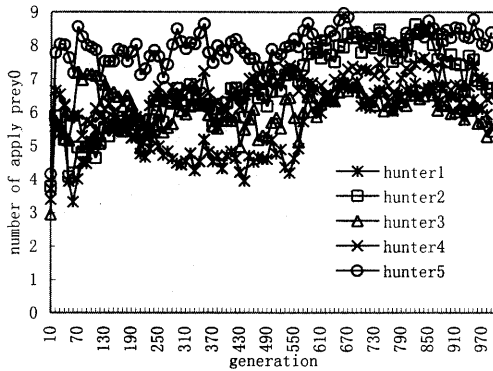


図6 獲物 p_0 に対する応募回数

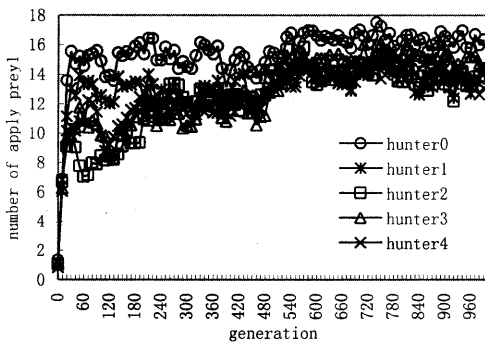


図7 獲物 p_1 に対する応募回数

獲物0						
	hunter0	hunter1	hunter2	hunter3	hunter4	hunter5
価値	60	50	40	30	20	10
公募別払い額	24	-	-	-	-	-
応募別払い額	-	10	8	6	-	-
利益	36	60	48	36	-	-

獲物1						
	hunter0	hunter1	hunter2	hunter3	hunter4	hunter5
価値	10	20	30	40	50	60
公募別払い額	-	-	-	-	-	24
応募別払い額	-	-	6	8	10	-
利益	-	-	36	48	60	36

表5 利得構成

6.1.3 考察

実験結果より、世代が進むにつれて、獲物 p_0, p_1 に対する公募・応募回数、即ち、提携回数が増えていることがわかる。50ステップの提携行動において、約40回の提携が行われるようになった。また、表5に示すように、獲物 p_0, p_1 に対する提携成立の大多数において、最大提携値を与え、かつ公募別払い額と応募別払い額の和の差で表される収支が0

となる利得構成が達成された。

これより、大きな提携値を獲得し、かつ、収支が0である提携の学習が可能であることが明らかになった。

6.2 提携と追跡行動の学習

次に、合理的な提携形成と追跡行動が並行的に学習可能であることをシミュレーション実験により明らかにする。

6.2.1 実験の枠組み

本実験では、6つのハンターが2つの獲物のいずれかに対して形成する提携と追跡行動の評価を行う。なお、関数と終端記号および実験パラメータは、表6、表7のように設定した。

6.2.2 実験結果

図8に、追跡行動のみの学習による獲物の捕獲数と、追跡行動と提携形成の並行学習における獲物の捕獲数の比較を示す。図9に、獲物 p_0 の提携のために各ハンターが公募した回数の変化、図10に、獲物 p_1 の提携のために各ハンターが公募した回数の変化を示す。図11に、図9の公募に対する各ハンターの応募回数の変化、図12に、図10の公募に対する各ハンターの応募回数の変化を示す。なお、各図とも10世代ごとに平均して示している。

6.2.3 考察

実験結果より、追跡行動のみよりも、提携に基づく追跡行動がより多くの獲物を捕獲できることが確かめられた。提携については、6.1.3と同様に、世代が進むにつれて、獲物 p_0, p_1 に対する公募・応募回数、即ち、提携成立回数が増えていることがわかる。1500ステップの提携・追跡行動において、捕獲した獲物の約半数について提携が形成されていることがわかった。

また、提携のみの学習では、1つの獲物に対して、1つのハンターのみが公募を行うのに対し、提携と追跡行動の学習では、1つの獲物に対して、複数のハンターが公募を行うようになることがわかった。

これは、ハンターが利得を獲得するためには、獲物の価値が大きくなるとも、その獲物の近くに位置するハンターのグループで提携を形成し、獲物を捕獲することが必要となるためである。その場合、そのグループにおいて最大の提携値を獲得できる提携を形成することが、合理的な提携形成である。従って、常に最大提携値を獲得する提携を形成するのではなく、環境状況や公募・応募状況に適していて、その提携のもとで、収支が0になるような提携が形成されている。

追跡行動木			
Function			
if_p_north if_p_south if_p_east if_p_west	視界内の獲物の位置により分岐		
if_h_north if_h_south if_h_east if_h_west	視界内のハンターの位置により分岐		
if_coalition_p0 if_coalition_p1	提携形成の有無により分岐		
Terminal			
north south east west stay	動く		
提携形成木			
Function			
if_view_h0 if_view_h1 if_view_h2 if_view_h3 if_view_h4 if_view_h5	視界内のハンターにより分岐		
if_view_p0 if_view_p1	視界内の獲物により分岐		
who_ad_p0 who_ad_p1	ap_num_p0 ap_num_p1	ap_pay_p0 ap_pay_p1	bal_p0 bal_p1
Terminal			
ad_10_p0 ad_10_p1	ad_20_p0 ad_20_p1	ad_30_p0 ad_30_p1	ad_40_p0 ad_40_p1
ap_10_p0 ap_10_p1	ap_20_p0 ap_20_p1	ap_30_p0 ap_30_p1	ap_40_p0 ap_40_p1
ignore			

表6 関数と終端記号

グリッドワールド	30*30	人口	100
ステップ数	1500	世代	1000
ハンター	6	木の最大深さ	7
獲物	2	交叉率	0.8
ハンターの視界	5	再生率	0.19
提携形成有効期限	15	突然変異率	0.01

表7 実験パラメータ

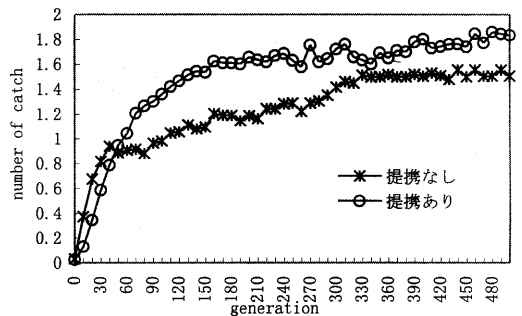


図8 獲物の捕獲数の比較

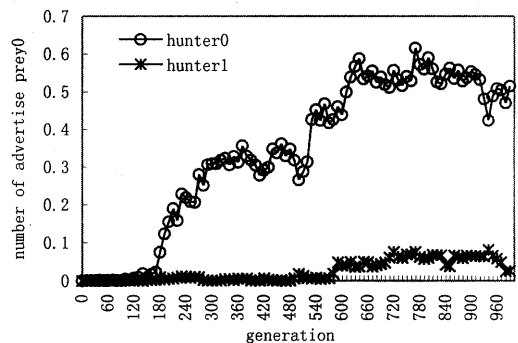


図9 獲物 p_0 に対する公募回数

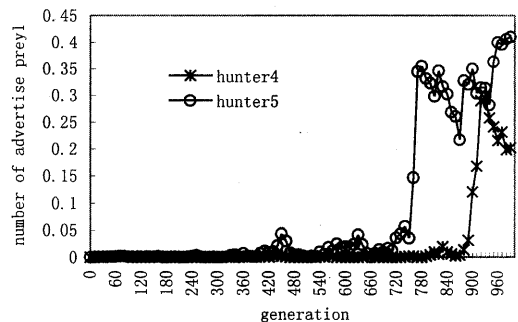


図10 獲物 p_1 に対する公募回数

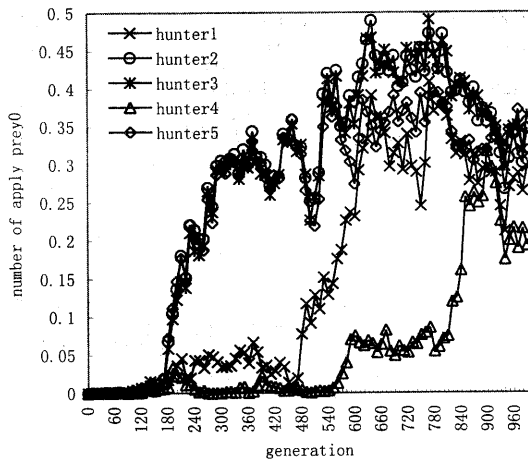


図 1.1 獲物 p_0 に対する応募回数

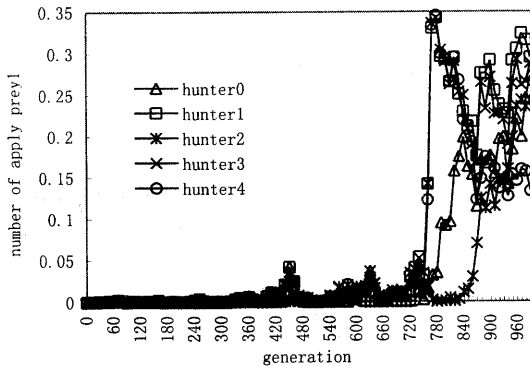


図 1.2 獲物 p_1 に対する応募回数

7 おわりに

本研究では、エージェント間提携形成のためのアーキテクチャとして、提携公募アーキテクチャを提案した。そして、拡張追跡タスクを用いて、提携とそれに基づく協調追跡行動が遺伝的プログラミングに基づき、進化的に学習可能であることをシミュレーション実験により明らかにした。

即ち、第1に、ハンターエージェント間に、最大の提携を与え、かつ別払いの収支を0とする提携行動が、進化的に学習されうることを示した。

第2に、提携に基づく追跡が、提携しない追跡よりも、良い性能を示すことを示した。また、追跡の中の提携においては、必ずしも自分にとって最大の

価値を与える獲物でなくとも、利得を獲得するためには、次善の提携行動を行うようになることを示した。

今後、形成された提携の性質をゲーム論的に、より詳細に分析すること、また、追跡中の提携行動の分析から、動的環境下での提携形成と追跡行動、及びそれらの学習のダイナミクスを明らかにすることが課題である。

参考文献

- [1] Koza, J.: Genetic Programming on the Programming of Computers by Natural Selection, MIT press (1994).
- [2] 佐藤、渥美: エージェント間提携形成に基づく協調追跡行動の進化的学習、情報処理学会第54回全国大会講演論文集(2)、pp.131-132 (1997).
- [3] Benda, M., et al.: On Optimal Cooperation of Knowledge Source, technical Report BCS-G2010-28, Boeing AI Center (1985).
- [4] 鈴木光男: "新ゲーム理論", 勁草書房 (1995).
- [5] Zongker, D., and Punch, B.: lil-gp1.0 User's Manual, Michigan State University (1995)