マルチエージェント系における競合共進化型学習の性能評価

マルチエージェント系における競合共進化型学習の性能評価

玉城 清政* 玉城 斉** 山田 孝治*** 遠藤 聡志***

The performance evaluation of competitive co-evolution in multi-agent system

Kiyomasa Tamashiro*, Tadashi Tamaki**, Koji Yamada***, Satoshi Endo***

Abstract

One of the important issues in intelligent systems and robotics is to develop an efficient method to control multi-agent system. In order to work multi-agent system well as problem solver, it's so significant to create cooperative behaviors among the agents. In the multi-agent system, the behaviors of cooperation emerged as the results of suitable role learning by each agent. In this paper, we evaluate reinforcement learning, genetic algorithm and competitive co-evolution algorithm from the viewpoint of adaptability to different environment as the learning method of multi-agent system, and discuss the property of each technique.

Keywords: Reinforcement learning, Genetic algorithm, competitive co-evolution

1. はじめに

知的システム工学やロボティックス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．従来の単一エージェントによる問題解決法では，エージェントは，複雑な環境下で現在の状態を認識する高度な状態認識モデルが必要であった．これに対して，マルチエージェントシステムでは，複数のエージェントが個々に局所的な状態を認識し，この情報認識に基づいた協調行動を実現することで問題の持つ複雑さに対応する．

本論文では，マルチエージェントの学習法として，従来手法であるⅠ.強化学習，Ⅱ.強化学習に代わる学習手法として進化的計算の手法である遺伝的アルゴリズム，Ⅲ.競合共進化アルゴリズムのそれぞれを異なる環境への適応性の観点から計算機実験により評価し，各手法の性質を検討す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 マルチエージェントシステムの定義

エージェントはセンサを通して環境を知覚できエフェクタを通して，環境に対して行動をすることができる行動主体と定義できる [1].

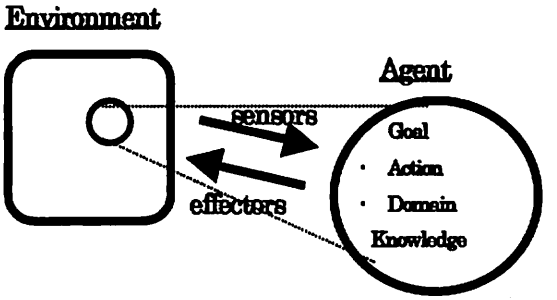

図1：エージェントの定義

マルチエージェントシステムでは，複数のエージェントは同一あるいは異なる目的と環境に対する領域知識が与えられ，センサを通して得られた知覚系列に対して行動を選択する．エージェントとの協調や敵対といった相互作用的振る舞いを学習することでシステムの目的を達成する．

2.2 エージェントの学習システム

エージェントシステムが適応性を実現するためには，エージェントが環境から得た情報に対して合理的行動を学習する必要がある．エージェントのプログラム中の判断処理は知覚系列に基づいて内部知識を参照にしながら取るべき行動を決定し，一般的定式化が可能である．基本的に，この判断プログラムについては，与えられた知覚系列から取るべき行動を求める写像として与えればいいことになる．したがって，エージェントにとって学習とは起こりうる各々の知覚系列について，高い性能指標を実現するために写像を更新することと置き換えられる．

受理：1999年6月7日・

*大学院理工学研究科情報工学専攻
(Dept. Information Engineering, Graduate School of Science and Engineering)
**大学院理工学研究科卒業
(Graduated, Dept. Information Engineering, Graduate School of Science and Engineering)
***工学部情報工学科
(Dept.of Information Engineering,Fac. of Eng)

2.3 強化学習

強化学習は，本来動物心理学や動物行動学の分野で用いられた用語である．典型的には，報酬という特別な入力を手掛かりとして行動パターンを学習する場合に用いられる．強化学習の目的は，報酬を最大化するように知覚された状態から行動への写像を生成することである．

学習者は，他の多くの機械学習のようにどの行動を出力すべきかが与えられるのではなく，どこ行動が最も高い報酬を得るかを試行錯誤に探索する．このような単純な原理によって動作する強化学習では，学習対象に対する明示的なモデルを持たずに学習することが可能である．

強化学習において，学習者はある環境のなかで行動を起こすエージェント，例えば，自律移動ロボットなどが想定される．学習者は各時間ステップにおいて知覚として与えられる状態から行動を決定する．ここで状態とは，学習システムにとっての外部からの入力である．実際にとった行動に対して環境から報酬あるいは罰が与えられるが，報酬の大きさは多くの場合，過去数ステップの行動系列に対して決定される．学習の目的は，ある時間長にわたる報酬の重み和を最大化することである．報酬と罰を合わせて強化信号（reinforcement）と呼ぶ．報酬を正の強化（positive reinforcement），罰を負の強化（negative reinforcement）と呼ぶこともある．

形式的には，時刻tにおける強化信号の大きさをr_tとすると，学習者の目的は，現在から未来にわたる強化信号の重み和，式（1.1）

$$v_t = \sum_{i=t}^{\infty} \gamma^{i-t} \cdot r_i \qquad (1.1)$$

を最大化することである．ただし，γは$0 \le \gamma \le 1$の定数であり割引率（discount rate）と呼ばれる．$r_t > 0$の場合には報酬，$r_t < 0$の場合には罰とする．$\gamma = 0$の場合は，現在の強化信号のみに着目し未来を無視することになる．逆に$\gamma = 1$では，どんなに遠い未来でもよいから大きな報酬が得られる方がよいことになる．すなわち，γの値の大小によって，どのくらい先の未来までを考慮するかが決まる．しかし，未来の報酬は観測できないので，一般には過去から現在までの強化信号の重み和，式（1.2）

$$\hat{v}_t = \sum_{i=0}^{t} \gamma^{t-i} \cdot r_i \qquad (1.2)$$

をv_tの近似として利用する．

強化学習は問題に付けられた，名前であるため，その実現方法はさまざまなものが提案されている．

2.4 Q-learning

代表的な強化学習アルゴリズムに Q-learning［2］がある．Q-learning ではルールと呼ばれる状態と行動の組に対する重みを見積もる．この重みをQ値と呼び，状態と行動の組から重みを導く関数をQ関数と呼ぶ．ルールとはエージェントの知覚可能な状態$x(\in X)$と選択可能な行動aの組み合わせである．状態xにおいて，行動$a(\in A)$をとるルールに対するQ値は，$Q(x,a)$と記述される．本実験では現在の状態において，Q値を最大にするルールを選択する方法を用いる．エージェントが時刻$t-1$の状態x_{t-1}において行動a_{t-1}を選択した結果，状態がx_tとなり，強化信号r_tが得られたとすると，Q値は以下の式(2)によって更新される．ここでAは現在の状態において可能な行動の集合とする．Xは現在の状態において知覚可能な状態の集合とする．

$$Q_t(x_{t-1}, a_{t-1}) = (1-\alpha) Q_{t-1}(x_{t-1}, a_{t-1}) + \alpha (r_{t-1} + \gamma \max_{k \in A} Q_{t-1}(x_t, a_k)) \qquad (2)$$

ここでαは学習率であり，$0 < \alpha \le 1$なる定数である．γは割引率とする．もし$\alpha = 0$の場合，Q値は更新されずエージェントは学習を行わない．右辺第1項は，以前のQ値がどの程度の割合を占めるのかを表し，αの高いエージェントほど，以前のQ値の占める割合が減少する．右辺第2項はこの時刻$t-1$で得られた報酬と現在の状態から見込まれる最大のQ値であり，αの高いエージェントほどQ値が大きく更新される．

3．進化的計算

進化的計算（ＥＣ）は，生物の進化過程を模倣した計算論的学習方法である．ＥＣのすべての枠組みに共通するオペレータは選択，交叉または突然変異があり，これらによって適応的な解探索が可能である．

3.1 遺伝的アルゴリズム（GA）［3］

ＥＣの代表的な手法に，遺伝的アルゴリズム（ＧＡ）がある．GA は，個体集合がある環境に適応するために，固定の環境からの評価をもとに進化し，適応能力の高い個体を獲得するメカニズムである．

3.2 GA の概要

現実の生物は，遺伝→発生→適応→淘汰のサイクルを繰り返し，ダイナミックに変化する環境に適応し進化して現在の姿になっていると考えられる．生物の進化過程にヒントを得た GA も同様の過程を経ることで，進化を促している．ＧＡのプロセスは以下のようになる．

Step1: ランダムに初期集団を生成．
Step2: 評価関数を用いて，各個体の適応度を計算．
Step3: 集団の評価を行い，終了条件を満たしていれば終了．
Step4: Step2 で計算された適応度に基づいて個体を選択（淘汰）．
Step5: 選択された個体に対し，交叉を実行．
Step6: 突然変異を実行．（再び Step2 へ）

以下では，これらの設定方法や種類について述べる．

コーディング

ＧＡで解を探索するには，始めに対象問題の解候補を数字またはアルファベット等の文字列で表現する必要がある．問題に応じて，何をどのようにコーディングするかは異なるが，遺伝子の表現に探索は大きく作用されるため，慎重にコーディングを設計する必要がある．

適応度関数

ＧＡにおいて，個体を進化させるためにも各個体を評価できる適応度を求める必要がある．適応度を計算するための指標が，適応度関数である．集団の振る舞いは適応度関数によって左右されるため，適応度関数を適切に設定する必要がある．

選択 (select)

選択は集団における適応度の分布に従って，次世代に生存する個体群を確率的に決定する．実装の方法は様々であるが，本研究で用いた選択法のみについて説明する．

1. ランク方式
   ランク方式は，適応度によって各個体をランク付けし，あらかじめ各ランクに対して決められた確率で子孫を残せるようにする．各個体は，その適応度ごとにランキングされており，選択確率は適応度にはよらず，ランクに依存する．
2. エリート保存法式
   エリート保存方式は，集団中で最も適応度の高い個体をそのまま次世代に残す方法である．これにより，現世代で高い適応度を示した個体は，遺伝子操作の影響を受けずに次世代への存続が可能となる．

本研究では，エリート保存方式によって異なる遺伝子を保存後，ランク方式による選択を行うという方法をとった．エリート保存方式では，優れた個体から進化した個体は次世代においても高い適応度を得ると考えられる．また，ランク方式により極端に高い評価値をもつ個体の再生産が抑制されるため，個体の多様性が維持できる．この二点から適応的に解の探索が可能になると考えられる．

交叉 (crossover)

交叉は，二つの親の遺伝子を組み替えることによって新しい個体を生成する操作である．両親の優れた部分形質をうまく組み合わせて，子に継承させることに成功すると，探索における飛躍をもたらす．

突然変異 (mutation)

突然変異は，遺伝子を一定の確率で変化させる操作である．あまり大きな変異確率に設定するとランダム探索と化してしまうが，ある程度の変異は必要である．突然変異がない場合は，初期の遺伝子の組み合わせ以外の空間を探索する事はできず，従って求められる解の質にも限界が出てくる．一般的に，突然変異は固定された確率で各遺伝子が変化するように設定する．

GA はこれまで述べたような遺伝子操作を行うことで，適応的に解の獲得が可能であると考えられ，進化モデルとして最適化問題に適応され，有効性が示されてきた．

4. 比較実験 I

適応性の観点からマルチエージェントの学習法について強化学習とＧＡを比較考察する．

4.1 追跡問題の設定

追跡問題はマルチエージェント標準問題の一つとして Benda らによって提案された問題であり，複数のハンターエージェントが獲物エージェントを捕獲することを目的としたモデルである [4]．

追跡問題を以下のように設定した (図2)．空間を 5×5 の大きさの 2 次元トーラス空間とし，エージェントの数をハンター 2，獲物１とする．獲物捕獲条件は，２つのハンターと獲物との距離が１となった状態とする．

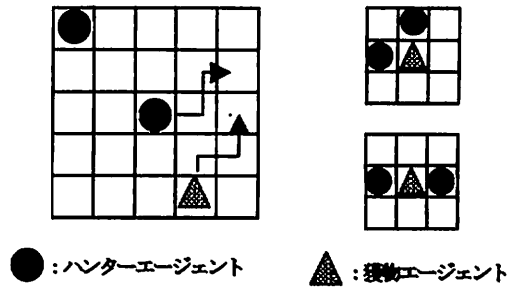

図2：追跡問題

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う．また，エージェントは同一マス上には存在できない．ハンターには，知覚として他のエージェントとの相対位置が与えられる．自分を中心として環境全体を知覚でき，ハンターと獲物の区別が可能である．獲物は時間遷移のみを知覚でき，予め与えられた長さ８タイムステップで記述された行動系列に従って行動することを繰り返す単純なエージェントとして実装する．獲物の行動パターンは約 40 万通り存在し，ハンターには事前に獲物の行動系列の情報は与えられないため，全ての事例を事前に学習することで獲物の捕獲行動を学習することは困難である．

100 ステップを上限として配置し，試行を繰り返しエージェントの学習を行う．また，各エージェントの初期配置は，獲物を固定としてハンターが獲物からの距離を 4 でハンター同士の距離が２となる位置に配置される．５×５マスの場合，配置のパターンは４通りとなる，４つのケース全てに対して学習，評価を行う．

4.2 各エージェントの設定

図3に学習事例として与える獲物エージェントを示す．Type-A，Type-B，Type-Cの3種類の獲物エージェントを用意し，4ケースの初期配置に対して対戦させ，獲物を交互に切り替えて学習及び評価を行う．

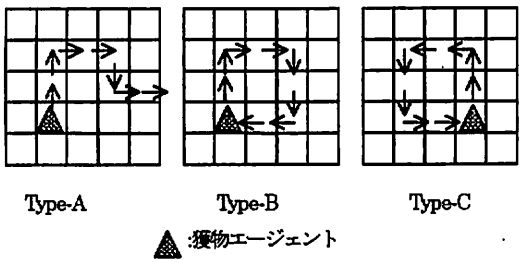

図3：学習対象の獲物の行動アルゴリズム

4.2.1 強化学習エージェントの設定

強化学習エージェントの設定において，Q-learningのパラメータは以下のように設定し式(2)に対応しており，10万試行の学習を行うものとする．

学習率 α	0.04
割引率 γ	0.9
Qの初期値	0.1
獲物捕獲時の報酬	+1.0
獲物未捕獲時の報酬	-0.1

表1：Q-learningのパラメータ

4.2.2 GAによる学習エージェントの設定

遺伝子を図4のように全ての知覚状態に対する行動を停止，上，下，右，左に対応させ0～4の整数値のテーブルとしてコーディングする．GAの進化的オペレータについて交叉オペレータは一点交叉，個体の選択にはランク方式とエリート保存方式を用いる．エリート保存方式により最大適応度を持つ遺伝子の保存，ランク方式により適応度が極端に異なる遺伝子の淘汰率を下げることによる遺伝子の多様性の維持が期待できる．ランク方式による各個体の適応度は，最大の評価を得た個体から順に一番目の個体の適応度を1.0とした公比0.9の等比減少列として与えられる．GAの各パラメータは表2のように設定した．

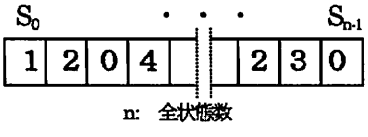

図4：ハンターエージェントの遺伝子コーディング

集団数	100
突然変異率	0.005
交叉率	0.2

表2：GAのパラメータ

ハンターエージェントの評価値は平均捕獲ステップ数が短いほど高くなるように式(3)のように決定した．

ステップ数の2乗の平均を評価値としたことで，平均のステップ数が同じ個体においては捕獲ステップ数が極端に低いケースをもつ個体ほど評価値が高くなるように設定した．これにより，短いステップ数で獲物を捕獲する構造をもつ個体ほど次の世代に残る確率が高くなり，個体に対する最適性が向上すると考えられる．

$$Evolution\ (\text{評価値}) = \left\{ \sum_{n=0}^{N} \frac{(\max step - step(n))^2}{\max step^2} \right\} / N \qquad (3)$$

但し，　$\max step$ ：最大ステップ
$steps(n)$ ： n試行目のステップ数
N ：1個体毎の試行回数

4.3 エージェントの学習実験

強化学習エージェントに対して，一回の試行最大ステップを100ステップとして，10万試行の学習を行った．学習の過程において1,000試行毎の平均ステップ数の計算を行い，結果を図5に示す．

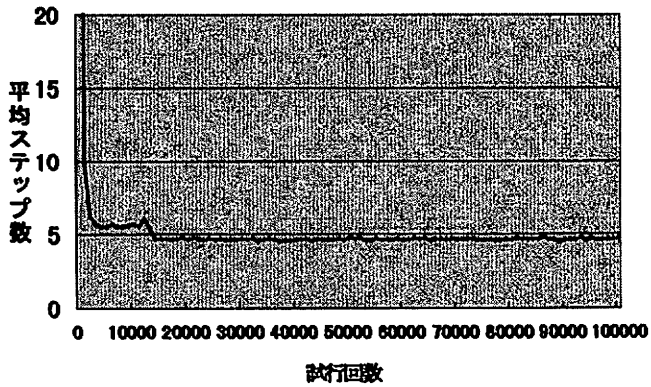

図5：強化学習エージェントの学習の様子

図5では，10万試行後の平均ステップ数は4.7前後まで減少し獲物を効率よく捕獲していることが分かる．すなわち，強化学習によりハンターが学習事例に対して学習を行ったことが確認された．

GAにより1,000世代まで学習を行った．図6にGAの最大評価値の推移を示す．

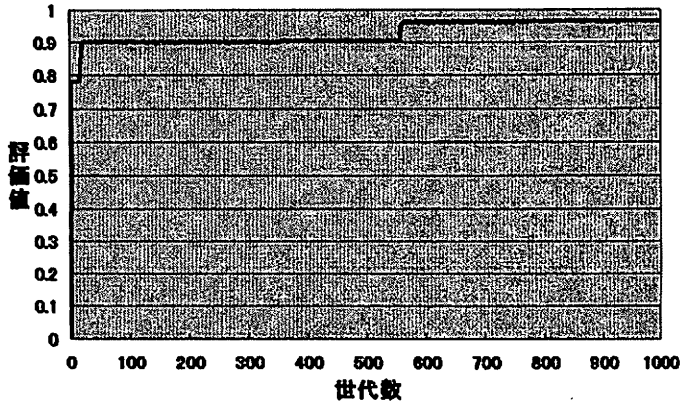

図6：GAの評価値の推移

図6より，20世代以降で評価値0.9という高い評価値に到達し，560世代以降で評価値0.95を越える個体が確認され，行動ルールの収束が見られた．すなわち，GAによってハンターの学習が確認された．

2つの結果から強化学習およびGAによってエージェントの学習が行われたことを確認した．次に，学習の結果に対する性能評価を行い，2つの学習手法を比較する．

4.4 評価実験の設定

獲物エージェントは，学習事例として3種類の獲物を設定した．また，評価事例として学習事例に類似又は正反対の行動系列を持つ3種類の獲物を図7に設定した．

Type-DはType-Bの半分の行動を停止に，Type-Eも同様に，Type-Cの行動の半分を停止にしたものである．最後にType-FはType-A全行動の上下左右を反転させたものである．

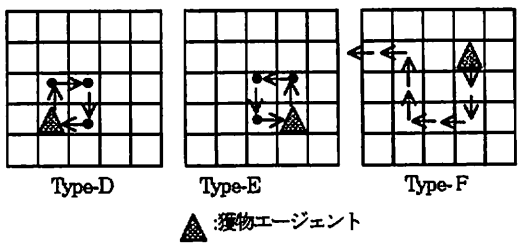

図7：学習評価用の獲物の行動アルゴリズム

4.5 実験結果

図3の学習に用いた3種類の獲物エージェントと図7の新たに設定した3種類の獲物エージェント，計6種類の獲物エージェントに対してそれぞれ100試行の毎のシミュレーションを行った．6種類の獲物エージェントに対する平均捕獲ステップ数を図8に示す．

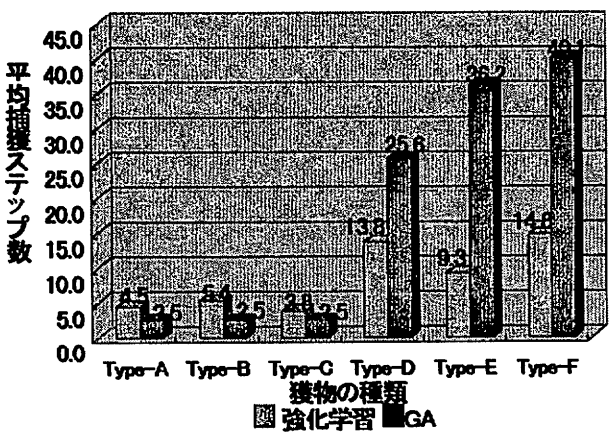

図8：学習及び未学習事例に対する平均捕獲ステップ数

図8より，強化学習，GAともに学習事例の方が平均ステップが低く，学習事例に対する効果的学習が行われた．一方，未学習の獲物にたいしては捕獲ステップ数が増加し，性能が悪化する傾向が見られた．GAによって学習したエージェントは，学習に使用した3種類に対してはいずれも2.5ステップと強化学習よりも短いステップ数を示している．しかし，未学習の獲物に対しては強化学習よりも低い性能を示した．これは学習パターンがどちらも短いステップの行動で捕獲できるため，GAにとって最適化すべき行動ルールの数が少なく，異なった獲物の行動パターンへのルールが獲得できなかったためと考えられる．一方，強化学習において異なる獲物エージェントの行動は試行錯誤的でランダムに近いことから，学習の初期において様々な状況についての学習の結果を得ていたことが考えれれる．

5 比較実験II

比較実験Iの結果より，GAによる学習において，固定の評価関数では，他のエージェントの行動の影響を含む評価が出来ず学習の汎化能力に乏しいことが分かった．そこで，進化的計算手法の一つであり生物間の相互作用によるアルゴリズムを適用し強化学習と比較する．

5.1 共進化型学習 [5]

競合共進化アルゴリズムを図9に示す．

Step1：遺伝的に区別された個体をもつ集団 Population1 (P1)，Population2 (P2) を作る．

Step2：P1 の全ての個体は，P2 からサンプリングされた全ての個体に対し，優劣比較を行い，その結果を P1 の個体の評価値とする (1st turn).

Step3：P2 の全ての個体は，P1 からサンプリングされた全て (2 nd turn)の個体に対し，優劣比較を行い，その結果をP2の個体の評価値とする．

Step4： Step2，3 の評価値を元に，P1，P2 に対して進化的オペレータを適用する (3 rd turn)．

Step5：終了条件が満たされないならばStep2に戻る．

本アルゴリズムにおいて，step2～step5 までを競合共進化1世代とする．

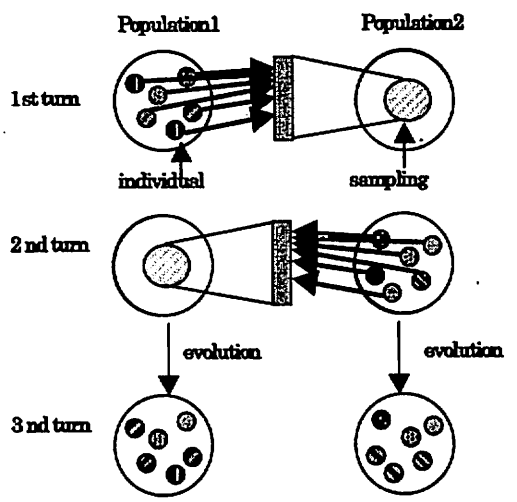

図9：競合共進化モデル

本研究では個体のサンプリングの方法を前の世代で得られた個体の評価値から上位 10 個体の選択とする．このように，互いに対して，評価の高い個体を見つけ合うことを続けた結果，多様な解空間の探索が可能となりさまざまな

問題に対応できる集団の形成が行われた.

競合共進化では，進化過程に基づいて変化する評価値を元に，両集団が常に相手に対して優位を目指すため，相互作用的に系全体が進化すると考えられる．従って，競合相手により評価が異なる問題に対して適応的に解集団の獲得が可能であると考えられる.

マルチエージェントシステムの相互作用には，協調と敵対があり，エージェントの関係が異なれば相手の行動の違いによる環境の変化は大きくなると考えれる．敵対する目的を持つエージェントが存在し，対象となるエージェントの行動が予想できない場合においても，学習の対象を固定せずに学習を行える競合共進化アルゴリズムはマルチエージェントシステムの敵対の関係にあるエージェントの振る舞いをそのまま表現できる．また，相互作用的にエージェントの行動が変化するため異なる環境に対しても適応性をもつ集団の獲得ができる.

5.2 共進化の設定

ハンターエージェントの設計に関しては GA と同様であるが，獲物エージェントは競合する集団となることから自らも学習を行う.

獲物エージェントは行動系列を遺伝子にコーディングする．獲物の突然変異率は 0.2 とした．これは獲物の 1 世代での突然変異による行動の変化がほぼ確実に起こすための処理であり，これによりエリート保存以外のプロセスで同じ個体が複製されないようになる．結果として，獲物側の集団の多様性がより大きくなる.

エージェントのサンプリング数はエリート保存方式によって得られた遺伝的に異なる上位 10 個体とする．互いに上位 10 個体と対戦後,進化的操作を行い次の世代に進む.集団数を 100，共進化世代数を 1,000 とした.

5.3 学習実験

図 10，図 11 はそれぞれ共進化ハンター側集団の最大評価値の推移，獲物側集団の最大評価値の推移を示す.

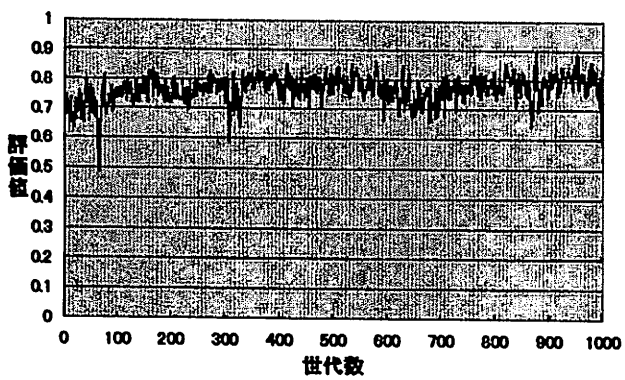

図 10：共進化ハンター集団の推移

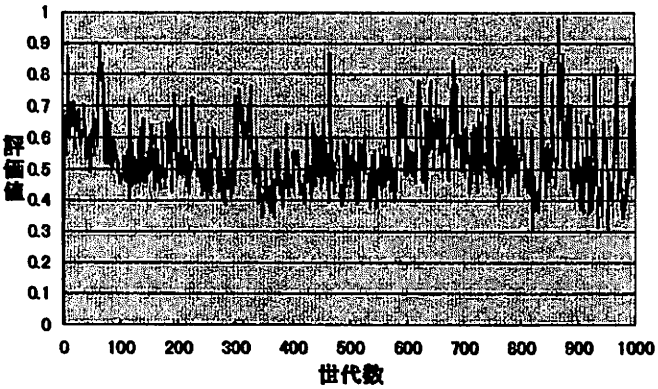

図 11： 共進化獲物集団の評価値の推移

図 10 では，初期のハンターエージェントの評価値は 0.7 前後で振動しているが，徐々に高い値へと推移し常に評価値が 0.8 程度を得られるような集団を形成していることが確認された．図 11 において，評価値の振動が激しいのは，獲物の集団の変動が頻繁に起こっているためと考えれられる.

次に競合共進化の進化過程において得られた獲物の集団に対してハンターの集団が捕獲に行動を獲得しているかを調査した．進化過程において得られた獲物の集団の上位 10 個体を 100 世代毎に選択し，各々 100 試行毎の捕獲シミュレーションを行った．その結果を図 12 に示す.

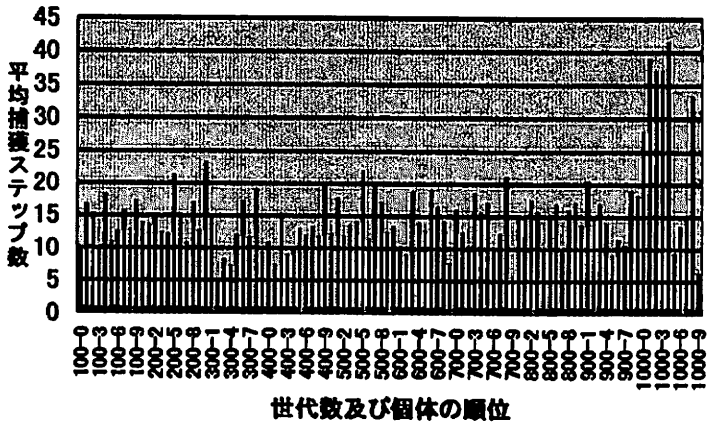

図 12 ：進化過程に得られた獲物に対する性能

図 12 において学習の過程に得られた獲物に対して 25 ステップ以下で捕獲していることが確認できる．また，次の世代に提示された獲物に対しては 40 ステップを超える個体も存在し，学習が必要な獲物が次の世代の評価相手としてあわられることが確認された.

以上の結果より共進化によって学習を行ったエージェントが，進化過程において以前に得られた獲物に対し有効であること，共進化アルゴリズムにより新たな学習事例としての獲物の獲得が行われていることが確認された．したがって，共進化型学習による結果はさまざまな獲物に対応出来ると考えられる.

5.4 評価実験の設定

学習を行った強化学習及び共進化のエージェントに対してランダムに生成した未知の獲物エージェントに対して捕獲実験を行ない性能を比較する．

環境及びエージェントの行動がどのように変化するかについて検討する．評価に用いる獲物行動系列はランダムに生成する．獲物の種類は 1,000 種類とする．それぞれの獲物に対して 200 試行のシミュレーションを行い平均のステップ数をその評価とする．また，共進化エージェントの評価では最終的に得られた集団内で各獲物に対して試験を行いそれぞれに高い性能を示した個体の平均ステップ数を評価する．従って，集団内の１個体ではなく集団全体としての最良の評価が得られる．

5.5 実験結果

図 13 に 1000 種類中の 100 種類の獲物に対する強化学習と共進化により獲得された集団の平均捕獲ステップ数を示す．

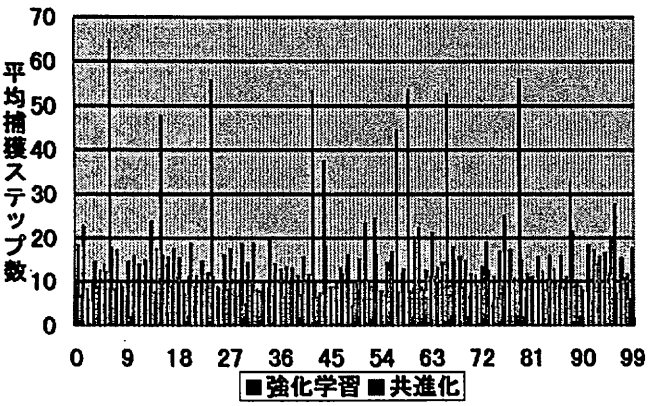

図 13：ランダムに生成した獲物エージェントに対する性能評価

図 13 の結果より，強化学習のエージェントの方が総じて共進化学習エージェントよりもステップ数が低いが，個々の獲物に対して捕獲ステップ数が高く捕獲困難な獲物の存在が確認された．

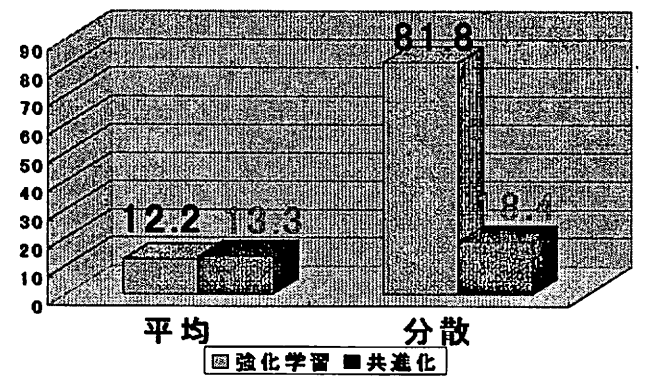

図 14：ランダムに生成した獲物に対する性能評価

図 14 に全ての獲物に対しての捕獲ステップ数の平均および分散を示す．

図 14 より平均においては強化学習エージェントのほうが若干よい性能を示しているが，ステップ数の分散についてみると共進化エージェントの方が圧倒的に小さく，さまざまな捕獲に対して良い性能を示していることが分かる．

以上の結果から，共進化型学習により獲得された集団がさまざまな獲物に対し有効であり，異なる環境への適応性が示された．

6 おわりに

本論文における実験の結果をまとめる．異なる環境に対するエージェントの適応性を評価するため，評価に用いる獲物の行動アルゴリズムが不明な環境として追跡問題を設計した．学習手法として強化学習,遺伝的アルゴリズム及び競合共進化アルゴリズムを用いてハンターエージェントに獲物捕獲行動を学習させた．次の二つの評価実験により，獲物平均捕獲ステップ数について性能比較を行った．

- 比較実験Ⅰ（強化学習—GA）学習事例として用いた３種類の獲物と学習事例と類似，正対する獲物として設定した３種類の獲物を評価事例に対し学習結果について性能比較をおこなった．
- 比較実験Ⅱ（共進化—強化学習）ランダムに設定した 1,000 種類の獲物を評価事例として学習結果について性能比較実験を行った．

比較実験Ⅰの結果より，GA により学習を行ったエージェントは学習事例として与えられた獲物に対して，強化学習で学習を行ったエージェントより低いエージェント，低いステップ数で獲物を捕獲しており，強化学習よりも固定の環境に対して有効に獲物捕獲が可能であることかが示された．一方，強化学習エージェントは学習事例に類似の獲物や異なる獲物に対しても性能が GA のように悪化せず，学習事例とは異なる獲物に対する，適応性が見られた．

比較実験Ⅱの結果より，共進化学習により獲得された集団はランダムに設定した獲物に対して平均ステップ数について性能評価を行った結果，ステップ数の分散を調べると強化学習に比べて小さくさまざまな獲物に対して適応していることが確認された．

このことから，実際の環境が予測できない場合のエージェントの学習に共進化学習が有効であることが示された．また，固定環境における獲物捕獲の性能は GA が高い最適性を示したこと，強化学習では,学習事例を任意に与えた場合でも，異なる問題に対しての適応性が示された．

参考文献

[1] Stuart, Russell.and Peter, Norvig.：エージェントアプローチ人工知能，共立出版株式会社（1997）.

[2] Watkins. C.J.C.H, and Dayan.P :Technical Note :Q-Learning, Machine Learning,, Vol.8,No.3, pp.279-292, (1992).

[3] 北野 宏明：遺伝的アルゴリズム，産業出版（1993）.

[4]Benda, M.Jagannnathan, V.Dodhiawalla,R : On ptimal cooperation of knowledge sources, Technical Report Technical Report BCS-G 2010-28,Boeing AI Center, (1985).

[5]W.D.Hillis : Co-evolution parasites improve simulated wvolution as an optimization procedure. Artificial Life II. Addison Wesley(1991).