F-002

# 動的環境に適応する自律学習システムの一考察

## Adaptive Learning System

中村峰幸 †　吉川雅弥 †　寺井秀一 †
Mineyuki Nakamura　Masaya Yoshikawa　Hidekazu Terai

従来からニューラルネットワーク(Neural Network)を用いた学習が広く使われている。しかし、従来手法では学習に必要な教師信号という情報が人間の作業によって集められる必要があるために自律学習には向かない。そこで本論文では事前予想が不可能な環境における自律動作を実現するために、行動制御と評価予測を行う2つのNNと遺伝的アルゴリズム（Genetic Algorithm 以下：GA）を用いた知能処理システムを提案する。その評価のために、植物と昆虫で構成される絶滅しやすい不安定な擬似生態系を対象とし本提案手法の、有効性を確認した。

## 1．　はじめに

本論文では事前予想が不可能な環境における自律動作を実現するために、行動決定と評価予測を行うための2つのNNを用いた知能処理システムを提案する。2つのNNの共進化により自律した学習を可能とする。

その評価のために、植物と昆虫から構成される絶滅しやすい不安定な擬似生態系を対象とし、昆虫を駆除するロボットを導入することにより、植物・昆虫のより安定した共生を目標とする。

この自律学習システムをロボットに実装する場合、消費電力の高い汎用的なCPUよりも、専用ハードウェア（以下：HW）を用いる事で消費電力を抑えることができ、高速処理が実現できる。以上より本研究では駆除ロボットの自律学習システムHW化を目指し、まずソフトウェア（以下：SW）により有効性の確認し、学習方式、進化方式をそれぞれ比較し実装すべき方式の検討を行った。HW化においては評価回路部分を設計し人間の反応速度を超える結果を得た。

## 2．　提案モデル

多くの場合、専門の知識をNNに取り込むことで特性が得られている。しかし、専門の知識を得ることが困難な場合や、専門ごとに知識が異なる場合には、専門の知識を前提とせずに自動的にルールを構成・学習する必要がある。

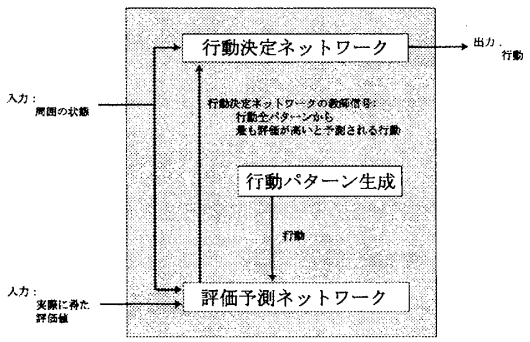



図1　提案モデル

バックプロパゲーション（Back-Propagation 以下：BP）[1]アルゴリズムを使用したNNの学習は入力ベクトルとそれに対応する出力ベクトル（教師信号）を使って学習をする。対象とする問題により、教師信号は変化するために、その取り扱いが複雑化している。そこで今回、教師信号の簡略化、及び自律した学習システムを構築するため、従来の学習に加え、予測という概念を付加する。

従来の1つのNNでは行動決定・学習しか出来なかったが提案する自律学習システムは図1の様に、2つのNNを並列動作させることにより、予測機能を備える。

## 3．　対象モデル

対象モデルはセル空間が20×20のセルオートマトン(Cellular Automata 以下：CA)を用いた植物、昆虫、そして提案する自律学習システムをそなえた昆虫駆除ロボットを20台導入した擬似生態系モデル[2]とした。植物、昆虫、駆除ロボットの関係は図2の様になっている。

このモデルでは、駆除ロボットは昆虫を駆除することによってエネルギーを得ることができ、昆虫は植物を食べて増殖する。植物は周囲8近傍の植物の密度によって増殖、死滅するが、周囲9×9の範囲以内に昆虫がいないと受粉が行えない為に死滅する。(受粉要素)

そのため駆除ロボットを用いて植物と昆虫が共存する環境を構築する。

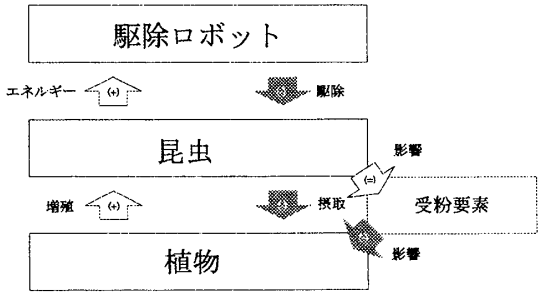



図2　対象モデル

† 立命館大学

## ４．　自律学習システム

駆除ロボットに適用した自律学習システムは学習に NN、進化に GA を採用した。NN は３層型ネットワークで構成し、非線形関数にシグモイド関数を、学習アルゴリズムに BP をそれぞれ用いた。行動決定ネットワークは入力層 11、中間層 10、出力層６で構成し、評価予測ネットワークでは入力層 17、中間層 13、出力層７で構成する。中間層はタンブ法[3]により決定した。また、GA は個体数 64 とした。

具体的な処理フローを図３に示す。

Step1：初期個体を生成する。
Step2：評価する１個体を入力し、個体の遺伝子（結合荷重）を元にネットワークを構成する。
Step3：行動決定ネットワークにロボットの環境（周囲の昆虫の有無、エネルギー、遭遇履歴）を入力し、最も評価値の高い行動を求める。
Step4：最も評価値の高い行動を教師信号として行動決定ネットワークの学習を行う。また一方で、実際の行動から得られた評価値を元に評価予測ネットワークの学習を行う。
Step5：Step3~4 を学習回数分繰り返す。
Step6：全個体について Step2~5 を行う。
Step7：選択・交叉・突然変異（GA 処理）を行う。
Step8：Step2~7 を世代数分繰り返す。

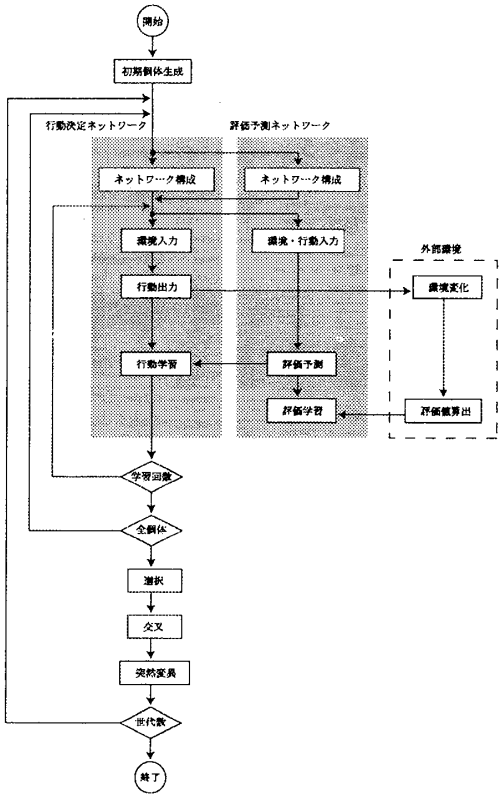



図３　処理フロー

## ５．　学習方式と進化方式の比較・検討

ここでは、ハードウェア実装の為に学習方式と進化方式の比較・検討について述べる。

まず、学習方式の比較・検討を行った。ＮＮによる学習は学習、睡眠、覚醒を繰り返すことにより行う。今回、睡眠に当たる部分を「10 回に１回ランダムに行動させる」事にした。睡眠無し、睡眠有りのエリート個体の行動内容を表１，表２に示す。

表１から睡眠無しでは世代を経て行動決定ネットワークが局所解に陥っていると考えられる。表２の睡眠有りの方では行動の多様性を保ち、局所解を回避しているのが読み取れる。

次に進化方法による評価値の変化では表3, 表4のようになる。ここでは獲得形質（結合荷重）が遺伝するラマルク型＋学習、遺伝しないダーウィン型＋学習、学習のみ、GA を用いた進化のみにより比較をした。評価は「昆虫の探索と直面する環境条件に適した昆虫の駆除」を元に行われ、評価値算出式（評価関数）を式（1）とした。

$$s(t+1) = s(t) + A \cdot Ins(t) + B \cdot (Ins(t+1) - Ins(t)) + C \cdot (Eng(t+1) - Eng(t)) \ldots (1)$$

ここでA，B，Cは定数、Ins(t)は時刻 t における周囲の昆虫の有無、Eng(t)は時刻 t におけるエネルギー残量、s(t)は時刻 t における評価値である。

表１　エリート個体の行動内容(睡眠無し)

| 世代数 | 2 | 160 |
|---|---|---|
| 駆除回数 | 729 | 1461 |
| 左移動 | 1450 | 3202 |
| 上移動 | 3094 | 0 |
| 下移動 | 230 | 0 |
| 右移動 | 3791 | 5337 |
| 無行動 | 706 | 0 |

表２　エリート個体の行動内容(睡眠有り)

| 世代数 | 2 | 160 |
|---|---|---|
| 駆除回数 | 1504 | 1483 |
| 左移動 | 1507 | 1442 |
| 上移動 | 2450 | 2529 |
| 下移動 | 1523 | 1529 |
| 右移動 | 1523 | 1544 |
| 無行動 | 1493 | 1473 |

表３　進化方法毎の評価値変化（睡眠無し）

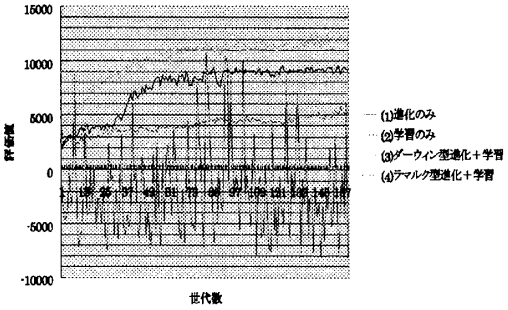



表４　進化方法毎の評価値変化（睡眠有り）

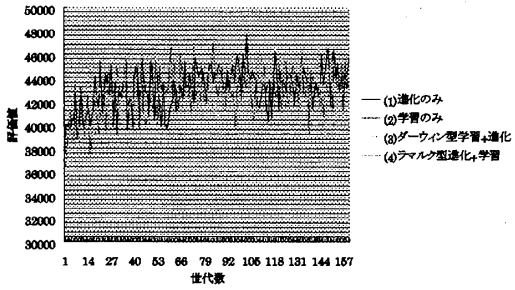



表3, 表4の評価値をみると睡眠有りの方が睡眠無しよりも評価値は倍以上高く、振幅が大きく、さらに収束が早い。また、進化毎の評価値は学習のみの評価値が低く、ダーウィン型進化＋学習、進化のみ、ラマルク型＋学習の順で評価値に大きく差があったが、順位は変わらないが殆ど差がなくなった。これは、解を導きだしたとも考えられることもできるし、評価予測ネットワークが局所解に陥ったともと考えられる。しかし、不安定な擬似生態系を対象にしているので早期世代で良い評価値を得られたことは評価できる。今回はここで解を導きだしたとする。

ここで500ステップ後の植物・昆虫の生存率を表５に示す。これにより学習・進化が進んでいることを確認する。ロボットと導入することにより、擬似生態系は安定していっているのが分かる。また、睡眠を行うことにより擬似生態系の安定は増している。

次に進化方式について、まず、選択方式について述べる。ここでは、エリート保存、トーナメント、ルーレットについて比較を行い、結果を表６示す。これからエリート保存方式が高い評価値を得ていることが分かる。

交叉方式は、一点交叉、二点交叉、一様交叉について比較を行い、結果を表７に示す。交叉方式では一様交叉が振幅も少なく、評価値が高い。

表５ 植物･昆虫の生存率

| | ロボ無 | 睡眠無 | | 睡眠有 | |
|---|---|---|---|---|---|
| 世代数 | × | 2 | 160 | 2 | 160 |
| 生存率 | 10% | 31.30% | 46.90% | 43.80% | 65.60% |

表６　択方式毎による評価値の変化

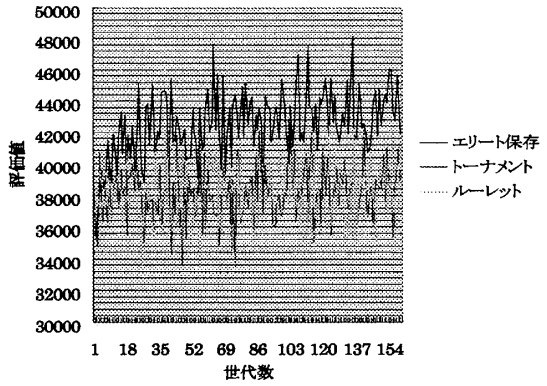
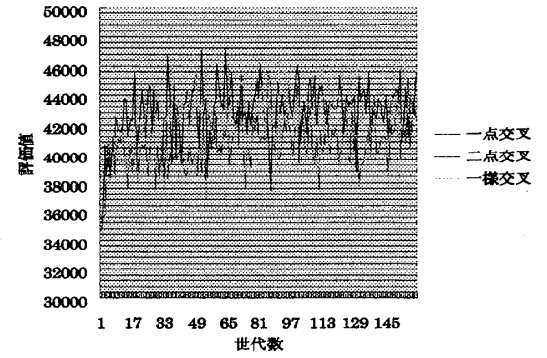


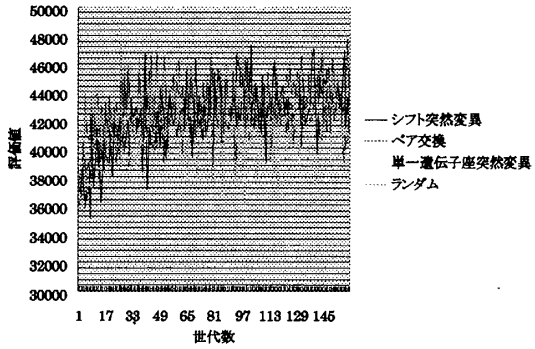

表７　交叉方式毎による評価値の変化




突然変異方式では、シフト突然変異、ペア交換、単一遺伝子突然変異、上記３つをランダムで行ったものの比較を行い、結果を表８に示す。単一遺伝子突然変異以外はどれも差がないのが読み取れる。

以上から、自律学習システムのハードウェア実装方式を考察すると、突然変異方式についてはFPGA実装段階で回路規模・速度を比較しシフト突然変異かペア交換のどちらかを選ぶ。

表 8　突然変異方式毎による評価値の変化




## 6．　回路設計

この自律学習システムをハードウェア(HW)化することにより実装面積が小さくなり、また高速な CPU などの消費電力が高い部品を使わなくて済む為に実装面で利便性に富む。そのような理由から今対象モデルにおける駆除ロボットの自律学習システム HW 化を目指し、今回はその進化処理部分を除く評価回路部分の設計を行った。

HW では NN の各層間の入出力を符号なし 8 ビットで表現、結合荷重を符号付き 11 ビットで表現した。

使用 FPGA ボードは Xilinx Vertex2　XC2VC6000 をターゲットとし。Synplicity 社 Synplify Pro で論理合成を行った。その結果、動作周波数 60.5MHz で動作することが確認できた。この論理合成から得た動作周波数から行動を出力するのに掛かる時間の見積もりを表 9 に示す。

一般に人間の応答速度が 0.2～0.3 秒と言われているので環境入力の 0.96μs、行動と行動の間隔の 74.26μs は十分と考えられ。進化処理部分を拡張しても時間的には余裕があると考えられる。

## 7．　まとめ

本論文では事前予想が不可能な環境における自律動作を実現するために、行動制御と評価予測の 2 つの NN を用いた知能処理システムを提案した。擬似生態系において学習方式と進化方式の比較・検討を行い、有効性と HW 化への見積もりを得た。

回路設計では動作速度を十分確保できそうな見積もりを得た。

現状では評価関数の決め方によりシステムが左右される状態にあるので評価値、評価関数の取り扱い方を考えてより汎用的なものに変えていきたいと考えている。

表 9　プロセスごとの必要なクロック数と応答時間

| プロセス | 必要クロック数 | 時間(μs) |
|---|---|---|
| 個体入力 | 479 | 7.91 |
| 環境入力→行動出力 | 58 | 0.96 |
| 行動出力→行動学習 | 342 | 56.5 |
| 行動学習→予測学習 | 102 | 16.8 |

## 参考文献

(1) Rumelhart, D.E., and Hinton, G.E., and Williams, R.J.: "Learning representations by back-propagating errors", Nature, 323, pp. 535-536 (1986)

(2) 久保田直行, 三原正雅, 小島史男, 福田敏男: 擬似生態系のための共進化ロボットの行動獲得, 知能と情報(日本知能情報ファジィ学会誌), Vol.15, No.1, pp.88-97(2003)

(3) Bahman Kermanshahi: ニューラルネットワークの設計と応用, 株式会社昭晃堂, pp.42-45 (1999)

(4) 杉本徳和, 鮫島和行, 銅谷賢治, 川人光男: 複数の状態予測と報酬予測モデルによる強化学習と行動目標の推定, 電子情報通信学会論文誌, Vol.J87-D-II, No.2, pp.683-694 (2004)

(5) 電気学会 GA・ニューロを用いた学習法とその応用調査専門委員会: 学習とそのアルゴリズム ニューラルネットワーク 遺伝的アルゴリズム 強化学習, 森北出版株式会社, pp. (2002)

(6) Bahman Kermanshahi: ニューラルネットワークの設計と応用, 株式会社昭晃堂, pp.42-45 (1999)