

## 分類子システムを用いたエージェントの適応戦略獲得

江谷典子、武田英明、西田豊明

奈良先端科学技術大学院大学

情報科学研究科 知能情報処理学講座

e-mail:noriko-e,takeda,nishida@is.aist-nara.ac.jp

設計者が計算機内部には最小限の知識記述だけを行い、環境に応じてその知識を計算機が書き換えて行動を修正していくような適応行動のとれる情報処理機構を構築することを目指している。ここでは、創発の発現機構としての可能性がある進化論的計算手法の一つである分類子システムを衝突回避問題に適用し、設計者の意図的な知識の埋め込みをしないで、適応度の高い行動戦略ルールを獲得する方法を考察するために、3つのタイプの環境モデルおよび分類子システムのルールリストの数についてシミュレーションし、勝ち回数を適応度の一評価とした結果を示す。

## Using Classifier System to Learn Adaptive Strategies for Collision Avoidance

Noriko Etani, Hideaki Takeda, Toyooki Nishida

Graduate School of Information Science

Nara Institute of Science and Technology

This paper presents an evolutionary algorithm evaluation for the development of a strategic behavior of a cell agent in 4x4 grid. It is discussed that the key for behavior emergence exists in a environmental model which a system receives from 4x4 grid environment. A model which is introduced an evolutionary computing method is applied to organize a cell agent's behavioral condition-action rules without a designer's knowledge. It is stressed on the importance of finding a new rule during the evolution rather than fitness.

## 1 はじめに

計算機の処理速度や記憶容量の増加と小型軽量化が進むにつれ、多様な環境での利用や多くの知識処理を行わせたいという要求が増大している。その中で、特定分野の専門知識による問題解決器としてエキスパートシステムが利用されてきた。しかし、知識獲得の困難さや状況依存性により、状況変化によるルールの書き換えや異なる分野への応用には自動化は困難であり、多くの人手を要して書き換えられている [12]。また、ロボットが生産工場から実世界エージェントとして日常生活の場やオフィスや学校という人間や自然環境という動的な環境へ進出していくためには環境への適応機能を備えた計算機が求められるであろう。そこで、設計者が計算機内部には最小限の知識記述だけを行い、環境に応じてその知識を計算機が書き換えて行動を修正していくような適応行動のとれる情報処理機構を構築することが不可欠である。しかし、複雑な振舞いを完全に予測し、制御するための適応的プログラムそのものを予め計算機に格納しておくことは不可能であり、プログラムあるいは処理機構が創発的に適応的なアルゴリズムを獲得していくアプローチが期待される。

ここでは、創発の発現機構としての可能性がある進化論的計算手法を衝突回避問題に適用し、設計者の意図的な知識の埋め込みをしないで、適応度の高い行動戦略を獲得する方法について考察を行い、その実験結果を示す。

## 2 衝突回避問題への適用

ここでは、視覚を持つ自律移動エージェントを想定して、衝突回避問題という  $4 \times 4$  の柵目上に配置されたセルが衝突をさせて目標達成するという動的な環境をシミュレートするプログラムの開発を行い、評価を行った。

### 2.1 問題設定

$4 \times 4$  の柵目 (図 1) 上における衝突回避問題の環境条件を次のように定義する。

	13	14	15	
first player (Random)	8	9	10	11
	4	5	6	7
0	1	2	3	
	second player (Classifier System)			

図 1: セルの初期配置

- ・環境は、 $4 \times 4$  の柵目から構成されている。ただし、その端は連結していない。
- ・柵目上を移動するセルは6つあり、2つのグループに分かれている。各グループの初期位置は、実験対象の分類子システムである赤グループが柵目番号 {1,2,3}、対戦相手で「ランダム動作」をする青グループが柵目番号 {4,8,12} である。
- ・1 試行中1ステップ毎に、一つのグループの一つのセルだけが移動できる。「前」「左」「右」「留まる」の移動方向から一つだけを選択し、ひと柵のみ移動ができるとする。同じグループあるいは相手グループのセルがいる位置には移動はできない。
- ・この試行の目的は、各グループの全セルが初期位置の端から反対側の端にたどりつくことである。先に全セルがたどりついたグループが勝ちとし、いち試行は終了する。

### 2.2 モデル化

Holland の提唱した分類子システムを設定問題に適用する計算機上の処理モデルを示す (図 2)。6つのセルを含む  $4 \times 4$  柵目を環境と設定し、その環境情報を解釈する機構を通して、環境メッセージへ情報がエンコーディングされる。これを環境モデルとする。分類子システムを用いるグループはこのモデルをシステム内部処理へ取り込む。一方、ランダム動作のグルー

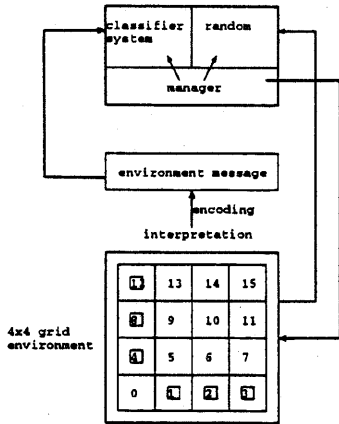


図 2: モデル化

ブは直接環境情報を参照し、3つのセルのうちランダムに「前進」できるセルを選び行動させる。。分類子システムを用いるグループとランダム動作を用いるグループの順番制御はマネージャが行う。本マネージャプログラムの役割は、

- ・ 2グループの実行権を制御する。
- ・ 4x4の柵目を環境と定義し、柵目上の情報を環境メッセージへエンコードする。
- ・ action メッセージの指示に従い柵目上のセルの動きを制御する。
- ・ 試行中の各種データの表示とファイルセーブを行う。

### 3 適応戦略獲得の評価実験

「勝ち回数」を一つの適応度を評価する尺度とし、ルールリストには初期設定ルールの記述を行わず、試行回数6000回を行い、次の点について評価実験を行う。

#### 3.1 環境モデル

柵目およびセルの配置を環境として獲得する時の環境モデルについて次の3種類の設計を考えた。

RED			BLUE		
1	2	3	1	2	3
***					

4bits:a cell position number of a grid

図 3: TYPE 1

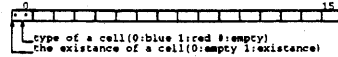


図 4: TYPE 2

#### ・ TYPE1(図 3)

各セルの属性として柵目番号を扱う環境モデルである。赤グループのセル位置番号と青グループのセル位置番号から構成されている。各位置番号は、4ビット表現にて行う。

#### ・ TYPE2(図 4)

視覚領域を16に分割し、各分割領域の属性としてセルの存在を考える環境モデルである。最初のビット群は15の柵目についてセルの有無を示す。次のビット群は15の柵目についてセルの種別が赤・青・空状態かを示す。2ビット表現にて行う。

#### ・ TYPE3(図 5)

16の柵目に着目したセルの有無とその種別、およびセル間の関係属性。人間が相手の手を読み行動選択を行う場合を模写した環境モデルである。a. 動作させるセルの直前にセルがある、b. 動作させるセルの前方にセルがある、c. 動作させるセルは移動先がない、d. 動作させるセルは相手セルの直前にいる、e. 動作させるセルは相手セルの前方にいる、f. 動作させるセルは相手セルが全く動けない位置にいるという関係属性である。一属性は、6ビット表現にて行う。

#### 3.2 ルールリストサイズ

分類子システムのルールリストの数について、環境モデル TYPE2を用いた場合、2

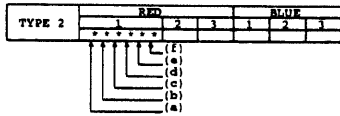


図 5: TYPE 3

000、4000、8000の場合の勝ち回数を検討する。これは、強いルールを生成する上で記憶数が影響を与えているのかどうかを考察するためである。

## 4 実験結果の考察

### 4.1 環境モデル

図 6、図 7、図 8 は、X 軸に試行回数、Y 軸に勝ち回数を表した TYPE1,TYPE2,TYPE3 の結果である。勝ち回数を比較した時、TYPE2 が一番適応度の高いルール生成に貢献している。TYPE1 は単に樹目番号を属性として扱っているために遺伝的アルゴリズムによりルールを生成しても、セルの経路を短縮する上では意味のない設計となるからと考えられる。また、Goldberg [2] は、定義長が短く低い次数で、しかも fitness が平均以上のスキーマは他のスキーマと組合わさって、さらによいスキーマが誕生するという仮説 (building block 仮説) を提案している。ここで、スキーマ定理より、スキーマ H の定義長 ( $\delta$ ) および次数 ( $O$ ) を次のように定義する。

$$O(H) = (\text{全長 } L) - (*\text{の数})$$

$$\delta(H) = (H\text{の最左と最右の非*の記号間の距離})$$

crossoverにより生き残る確率は、染色体全長(長さ  $k$ )の可能な crossoverpoint の数 ( $k-1$ ) と定義長より

$$P_s = 1 - \frac{\delta(H)}{k-1}$$

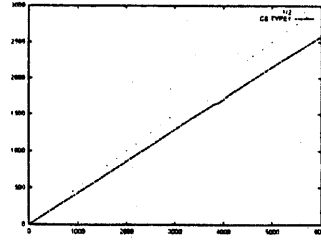


図 6: TYPE1

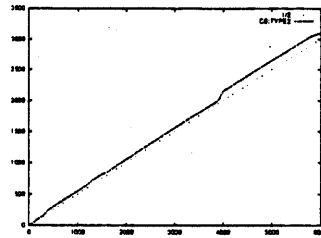


図 7: TYPE2

*mutation* 確率を  $P_m$  とし、ある位置の遺伝子に確率  $P_m$  で一様に作用するので、スキーマの具体値 (0 または 1) の部分のみに *mutation* が起こらない確率すなわちスキーマの生存確率は  $(1 - P_m)$  である。具体値は  $O(H)$  であり、生存確率は

$$(1 - P_m)^{O(H)}$$

従って、*buildingblock* 仮説より、環境モデルの表現ビット数の少ない TYPE2 が適応度の高いルールを生成する可能性を示している。

### 4.2 ルールリストサイズ

図 9、図 10、図 11 は、X 軸に試行回数、Y 軸に勝ち回数を表したルールリストサイズ 2000、4000、8000 の結果である。勝ち回数を比較したところ、ルールリストのサイズに依存する関係はないと考えられる。

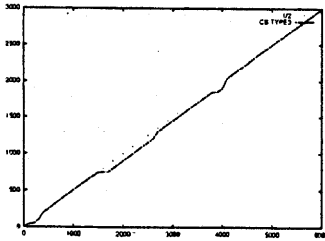


図 8: TYPE3

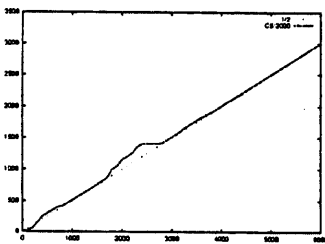


図 9: 2000

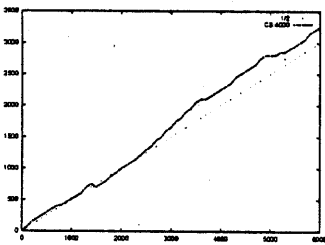


図 10: 4000

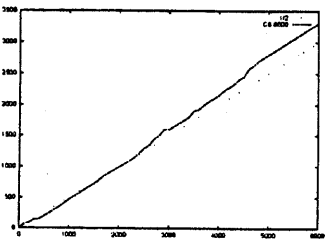


図 11: 8000

### 4.3 まとめ

基本的な分類子システムを用いたシミュレーション結果から、何らルール記述をしていない状態から6000回数の試行を通して、視覚領域を分割し、各領域の属性としてセルの存在を捉えた場合、 $1/2$ 以上の勝ち回数を獲得するところまで適応的ルールの生成が可能であることが示唆された。一方、実環境で動作するエージェントを想定した場合、内部におけるmatch/postの繰り返し処理の効率を上げ、即時処理が行える機構を取り組むことや、ルールリストには予め設計者の知識を記述するなどして処理機構そのものの適応および学習効率を上げる試みが必要である。

## 5 課題

本研究では視覚を有する自律移動エージェントを想定し、環境への適応的ルールの生成を可能とする進化論的計算手法の一つである分類子システムを衝突回避問題に適用し、視覚の捉える環境モデルから獲得する行動戦略により強いルール生成が行われたことを「勝ち回数」という一評価により考察を行った。設計者が初期設定として何らルールを直接記述することなく、環境との相互作用から方向付けられた結果であり、いち「創発」のプロセスと解釈することができる。ただし、創発のプロセスは単に環境への適応度の高さだけを評価するのではなく、何らかのルールの発見が行われていることも解析する必要がある。今後は、この解析とともに、設計者が初期設定として知識記述を行い試行した結果、設計者が考慮していなかったルールを発見することができるのかを試しつつ、リアルタイム性があり適応戦略を獲得できるような情報処理機構を構築していきたい。

## 参考文献

- [1] JOHN H. Holland. ADAPTION IN NATURAL AND ARTIFICIAL SYSTEMS. The MIT Press(1992).

- [2] David E. Goldberg. GENETIC ALGORITHMS in Search, Optimization Machine Learning. Addison-Wesley Publishing Co.(1989).
- [3] George Kampis. SELF-MODIFYING SYSTEMS IN BIOLOGY AND COGNITIVE SCIENCE. A New Framework for Dynamics, Information and Complexity. IFSR International Series on Systems Science and Engineering. PERGAMON PRESS(1991).
- [4] Alan C. Schultz. Using Genetic Algorithm to Learn Strategies for Collision Avoidance and Local Navigation. Navy Center for Applied Research in Artificial Intelligence(Code 5514). Naval Research Laboratory. Washinton, DC20375-5000, USA(1991).
- [5] Mark Pendrith. On Reinforcement Learning of Control Actions in Noisy and Non-Markovian Domains. School of Computer Science and Engineering. The University of New South Wales. UNSW-CSE-TR-9410(1994).
- [6] 北野宏明 編. 遺伝的アルゴリズム. 産業図書 (1993).
- [7] 伊庭斉志 著. 遺伝的アルゴリズムの基礎. オーム社 (1994).
- [8] Lawrence Davis 編. 遺伝的アルゴリズムハンドブック. 森北出版株式会社 (1990).
- [9] Rolf Pfeifer. 招待論文 Symbols, patterns, and behavior: toward a new understanding of intelligence. 1996年人工知能学会全国大会(第10回). pp.1 - pp.15(1996).
- [10] Yukinori Kakazu. ロボティクスにおける創発と進化. 日本ロボット学会誌. Vol.15 No.5. July. pp.2 - pp.6(1997).
- [11] 伊庭斉志. 遺伝的プログラミングとロボット. 日本ロボット学会誌. Vol.15 No.5. July. pp.7 - pp.10(1997).
- [12] 西田豊明. 情報指向の人工知能と知識コミュニティ. 情報処理学会. 情報メディア研究会資料. pp.41 - pp.48(1994).