自律ロボットの行動獲得のための制御アーキテクチャ*

(大域結合カオス系を用いた強化学習)

中村 陽一郎^{*1}, 大倉 和博^{*2}, 上田 完次^{*2}

Control Architecture for Behavior Acquisition of Autonomous Robots (Reinforcement Learning Using Globally Coupled Chaotic System)

Yoichiro NAKAMURA, Kazuhiro OHKURA and Kanji UEDA

An autonomous robot that acts in a complex environment needs to learn which action should be performed in each external state. Therefore, we propose a new control architecture for autonomous robots. The control architecture includes a learning system, which consists of globally coupled chaotic elements. A chaotic element has dynamics which are designed so that the elements can collectively execute reinforcement learning. Although each element continuously updates its internal state according to its intrinsic dynamics, this local processing collectively determines an action of the robot and allows it to interact with its environment. The result of this interaction returns to the elements as a payoff, which changes the dynamics. We carried out computational experiments for a navigation task. The results show that the chaotic elements self-organize so that an autonomous mobile robot exhibits such behaviors as goal-reaching, wall-following and collision avoidance.

Key Words: Mechatronics and Robotics, Moving Robot, Learning, Behavior Acquisition, Reactive System, Reinforcement Learning, Chaos, Synchronization

1. はじめに

建設現場や海中, さらにはオフィスや一般家庭など, 複雑かつ動的に変化する環境下で活動する自律ロボットが求められている. こうした複雑な環境に対しては, ロボットが遭遇する状態を設計者がすべて予期し, 各状態に対し動作ルールを準備しておくことは困難である. また, 初期状態から目標状態までの動作列を予め決定しておくことは, 環境変動に対し意味をなさない.

これに対し, サブサンプリング・アーキテクチャ⁽¹⁾に代表される行動ベース型ロボットは, 環境の状態に応じて即応的に動作し, 複雑で変化する環境に対しても頑健かつ柔軟に対処できる. しかし, 即応的に動作するロボットがタスクを達成するためには, 即応性ととともに動作の合目的性が要求される. したがって, タスクの遂行に有効な行動を自律的に獲得していく学習機能が必要となり, 強化学習や進化的手法による行動獲得の試みがなされている⁽²⁾⁽³⁾⁽⁴⁾.

本論文では, 行動ベース型の自律移動ロボットを対象に, 即応性と合目的性という二つの機能の統合的実現に向けて, 新しい制御アーキテクチャを提案する.

この制御アーキテクチャの目的は, 目標状態に至る動作列を強化学習により自律的に獲得することにある. そして, 強化学習の機構に大域結合カオス系⁽⁵⁾を適用する. カオスを応用した情報処理機構としてはニューラルネットが一般的であり, ロボットの学習に適用した例もある⁽⁶⁾. 一方, 大域結合カオス系については工学的応用があまりなされていないものの, クラスタリングやカオス遍歴など, この系が有する様々な特徴は, 知的システム構築への新たな要素技術になりうる. そこで, 本研究では, 大域結合カオス系が系全体として強化学習をするように, 系を構成するカオス要素にダイナミクスを与えることを試みる. すなわち, 環境に関する知識や動作ルールが予めなくてもロボットが環境を探索できるように, 多様な動作を即応的に発生する仕組みと, タスクの達成に有効な動作を導くカオス要素が自己組織するような仕組みを, カオス要素のダイナミクスに組み込む. そして, この制御アーキテクチャの有効性を検証するために, 自律移動ロボットのナビゲーション・タスクについて計算機実験を行い, 結果を考察する.

なお, 本論文では, ある環境状態において自律ロボットが出力するエフェクタ操作の集合を, 動作(action)と呼ぶ. また, 一連の状態一動作列として観察されるロボットの運動パターンを, 行動(behavior)と呼ぶ.

* 原稿受付 1996年9月20日.

*1 正員, 日立造船(株)技術研究所 (〒551 大阪市大正区船町2-2-11) (神戸大学自然科学研究科).

*2 正員, 神戸大学工学部 (〒657 神戸市灘区六甲台町).

Fig.1 Chaotic itinerancy

2. 大域結合カオス系の適用

2・1 適用の動機 生命関係学の概念によれば、生命システムが外部環境の無限定性に対処していくためには、内部状態の無限定性、すなわち内的自由度が不可欠である⁽⁷⁾。つまり、生命システムは、それ自体が複雑で無限定なカオス的な状態を内包していて、その中から多様な秩序状態を導き出すことができるため、未知の環境や状態変化に対してもシステムの状態を状況に応じて決めることが可能となる。

一方、行動ベース型ロボットに関しても、独立に機能するモジュール間の相互作用から行動が創発するという認識がある。このような独立に機能するモジュール間の相互作用の実現形式としては、場を通じた相互作用が適している。実際、ダイナミクスを持つ複数の要素が場を通じて相互作用することにより、歩容のような秩序的な時空パターンを自己組織化する様々な自律分散型システムが提案されている⁽⁸⁾。これに加えて、行動を学習するためには、秩序(本研究ではロボットの行動)を維持しようとする要素や、新たな秩序を生みだそうとする要素が、自己組織化した秩序の拘束を受けながらさらに相互作用を行い、系全体としてよりよい秩序を形成するように発展していくことが求められる。そのためには競合と協調の機構が必要になる。

大域結合カオス系では、カオスと引き込みという非線形振動子特有の二つの性質によって、上記のような内部状態の無限定化と秩序化が生じる。そして、この働きによって、場を通じた競合と協調が実現される。

2・2 大域結合カオス系 多数のカオス要素が場を介して大域的に結合したこの系に関しては、クラスタリング、カオス遍歴、階層構造の創発などの興味深い性質が金子によって見出されている⁽⁹⁾。クラスタリングは、複数のカオス要素が相互作用によって引き込

み同調する現象で、同時に複数のクラスタ(同調関係)が存在し得るが、不安定性が強ければクラスタは安定せず、時間的に同調関係が変化する(カオス遍歴)。

図1に、3個のカオス要素が大域的に結合した系における遍歴的挙動を示す。3個の要素が引き込んで一時的に一つのクラスタを形成するが、その後二つのクラスタに分離している様子がわかる。要素のダイナミクスは次式の離散時間方程式で表される。

$$x_i[t+1] = F(x_i[t]) + \varepsilon \left\{ \frac{1}{N} \sum_j x_j[t] - x_i[t] \right\} \cdots \cdots (1)$$

カオス性を有した非線形写像 $F(x)$ の具体的な形は、第5章で示す。カオス要素は、全要素の振動の大域平均場を通じて全要素と間接的に結合している(右辺第2項)。この結合は、要素の振動を平均場に近づけるが、要素のカオス性は両者の差を拡大する。この二つのはたらきが競合することによって、カオス遍歴が生じる。

このように、大域結合カオス系では、カオス要素群が振動場を通じて相互作用する結果、カオスによる内部状態の無限定化と、引き込みによる秩序化が起こる。また、大域場とのみ結合し、要素間を直接結ぶ結合がないため、要素の追加や削除が容易である。一方、第5章で示すように、特定の要素間の同調をロックすることもできる。以上の特徴から、特に複数の要素から構成されるシステムにおいて、大域結合カオス系は次のような機構になり得る：(1)要素間での多様な関係の形成、(2)要素どうしの協調と競合、(3)要素間の関係の記憶。本研究では次章で示す分散型制御アーキテクチャによる強化学習に、上記三つの機構を応用する。

3. 自律移動ロボットの制御アーキテクチャ

計算機実験に適用する自律移動ロボット(図2)を具体例として扱い、制御アーキテクチャを説明する。

3・1 状態の知覚 図2に示すように、ロボットは2個の光センサと7個の近接センサ、および5個の衝突センサを有している。後述する学習システムには、光センサと近接センサの計9個の出力を2値化した9ビットのストリングが、状態入力として与えられる。

3・2 動作の規定 独立に作動する要素的なエフェクタ駆動操作を設け、操作どうしの同期関係のパターンによりロボットの動作を規定する。図2では、左右の車輪を回転させる3種類の操作 $O = \{o_1, o_2, o_3\}$ を設けている。具体的には、 o_1 ：左車輪の前回転(回転量 w_p)、 o_2 ：右車輪の前回転(回転量 w_p)、 o_3 ：左右両輪の後回転(回転量 w_b)である。3個のエフェクタ操作間の関係に応じて、左右2輪の差動機構によって前進、微速前進、後退、左・右回転、左・右旋回の $2^3-1=7$ 種類の動作が得られる。すなわち、各エフェクタ操作の出力を

Fig.2 Control architecture of the mobile robot

$o'_i (= \{0, 1\})$, 動作パターンを $A = \{o'_1, o'_2, o'_3\}$ と表すと、動作パターンに対する右車輪と左車輪の回転量 ϕ_R, ϕ_L は、それぞれ次式で決定される。

$$\left. \begin{aligned} \phi_R &= w_f o'_3 - w_b o'_2 \\ \phi_L &= w_f o'_1 - w_b o'_2 \end{aligned} \right\} \dots\dots\dots (2)$$

回転操作の回転量は $w_f > w_b$ と設定している。したがって、たとえば $A = \{1, 0, 1\}$ は前進動作、 $\{0, 0, 1\}$ は左旋回、 $\{0, 1, 1\}$ は左回転、 $\{1, 1, 1\}$ は、微速前進となる。

このように、個々の操作は単純でも、その組み合わせにより得られる動作は多様となる。学習システムは、状態入力に対してエフェクタ駆動操作間に関係付けを行い、動作を決定する。この関係付けの機構として大域結合カオス系を適用する。

3・3 学習システムの構成 学習システムは、カオス要素の集合 E から構成される。 E は、カオス遍歴により互いに無限定な同調関係を形成できる要素（無限定要素）の集合 E_{IND} と、同調関係を限定された要素（限定要素）の集合 E_{DEF} からなる。

$$E = \{E_{IND}, E_{DEF}\} \dots\dots\dots (3)$$

カオス要素 $e_i \in E$ は、 D_i をダイナミクス、 X_i を内部状態変数、 mem_i を記憶値として次のように記述される。

$$e_i = \langle D_i, X_i, mem_i \rangle \dots\dots\dots (4)$$

一方、各エフェクタ駆動操作 o_i には、カオス要素の集合であるコラム C_i を対応させる。

$$C_i = \left\{ e_{i1}, e_{i2}, \dots, e_{iN_i(s_n)} \mid e_{ij} \in E_{DEF}, e_{iN_i(s_n)} \in E_{IND}; \right. \\ \left. i = 1, \dots, Nc, j = 1, \dots, N_i(s_n) - 1 \right\} \dots\dots\dots (5)$$

式(5)のように、各コラムには無限定要素が常時1個必ず存在する。特に、初期状態においては、各コラムに

は無限定要素1個だけしか存在しない。 $N_i(s_n)$ は、状態 s_n におけるコラム C_i の要素数である。カオス要素が持つ再生と死滅機能によって、各コラム内の要素数は、状態遷移ごとに変化する。

さて、 R をカオス要素間に同調が生じている関係とすると、異なるコラム C_i, C_j に属する2個の要素間に同調が生じている場合、次のように表される。

$$e_{i1} R e_{jm} \quad (e_{i1} \in C_i, e_{jm} \in C_j) \dots\dots\dots (6)$$

この状態をコラム C_i と C_j の間に関係が形成された状態と考えて同様に R と表すと、 R の関係行列 r_{ij} が各コラム間に定義できる。

$$r_{ij} = \begin{cases} 1 : (C_i, C_j) \in R \\ 0 : (C_i, C_j) \notin R \end{cases} \dots\dots\dots (7)$$

センサ入力を受けると、無限定要素と、以前その状態で形成された同調関係を記憶している限定要素群が、各コラム内で活性化する。そして、ダイナミクスにもとづき相互作用を行い、互いに自己がつくる同調関係を持続させようと競合する。環境状態 s_n において、あるコラム C_i の、他のコラム C_j との最終的な関係は、カオス要素 $e_{il} \in C_i$ のダイナミクス D_{il} によって定まる。

$$D_{il} : s_n \rightarrow r_{ij} \quad (l = 1, \dots, N_i(s_n), j = 1, \dots, Nc) \dots\dots (8)$$

$r_{ij} = 1$ をとった場合、コラム C_i は操作 o_i を発火させ、同時に C_j も操作 o_j を発火させる。ロボットの動作 A は、コラムどうしの関係に対応して、エフェクタ操作の集合 $O = \{o_1, o_2, \dots, o_{Nc}\}$ のべき集合で規定される。

$$A \subset \rho(O) \dots\dots\dots (9)$$

したがって、動作数は関係が生じていない場合 ($R = \phi$) を除く $2^{Nc} - 1$ 個となる。このように、ダイナミクスは状態と動作との間の写像の役割を果たし、同調関係 r_{ij} は状態 s_n での動作ルールに相当する。動作の発火により状態は遷移するが、タスクを完遂するためには、目標達成に有効な動作を導く関係系列を、状態遷移に応じて形成する必要がある。そこで、関係系列を強化学習によって自己組織する仕組みをカオス要素に与える。

4. 大域結合カオス系による強化学習

図3に、強化学習の流れを示す。本文中の番号は、それぞれ図中の番号と対応している。

(1) 学習開始時、各コラム内には無限定要素しか存在しない。センサ入力に対して、無限定要素は同調関係を変化させながら互いに競合する。(2) 競合を経て同調関係が一つだけになった場合、その関係に対応した動作が発火する。(3) 動作を発火させた無限定要素は、その状態でのセンサ入力ストリングを記憶するとともに新たな状態-動作のルールとして同調関係をロックし、限定要素となる。(4) 同時に、動作ルールの生成能力を維持するため、無限定要素をそれぞれのコ

Fig.3 Flow of reinforcement learning

ラムに再生する。(5)すでに経験した状態では、その状態をストリングに記憶していたすべての同調関係が活性化し、無限定要素が一時的に形成する同調関係とともに競合する。(6)競合に勝った同調関係が動作を発火する。(7)ロボットが目標状態に到達した際にのみ報酬が与えられる。(8)同調関係とそれに対応した動作の有効度として、報酬が要素群に分配されていく。(9)一定値以下に有効度を減らした要素は死滅する。

学習システムは、観測した状態に対して無限定要素が同調関係をつくることにより、環境内で即応的に動作ルールを生成していく。動作ルールを記憶した限定要素群と、さらに動作を発生させようとする無限定要素群は、各状態で互いに競合する。競合に勝ち、報酬獲得に寄与する動作を発火して報酬を分配された限定要素は死滅を免れ、系内に存続していく。

5. カオス要素のダイナミクス

以下では、式(10)のガウス型関数、式(11)のシグモイド関数、および式(12)のしきい値処理関数を用いる。

$$f(x, c, \theta) = \exp[-c(x - \theta)^2] \quad (10)$$

$$g(x, c, \theta) = \frac{1}{1 + \exp[(\theta - x)/c]} \quad (11)$$

$$h(x, \theta) = \begin{cases} 1: & \text{if } \theta < x \\ 0: & \text{otherwise} \end{cases} \quad (12)$$

5.1 カオス要素の構成 式(4)で定義したカオス要素 $e_i \in E$ は、5個の内部状態変数を有している。

$$X_i = \{x_i, u_i, b_i, r_i, d_i\} \quad (13)$$

x_i : 振動子状態量 u_i : 同調関係の有効度
 b_i : 振動出力荷重 r_i : 報酬分配率
 d_i : 振動場 $\bar{x}$ との相関度

各変数を、ダイナミクス D_i 、すなわち、式(15)、(23)、(24)、(27)、(28)の5つの離散時間方程式で更新する。

Fig.4 Return map : $c_0=2.5$, $c_1=3.0$, $c_2=5.5$, $c_3=0.75$, $c_4=0.05$, $\theta_x=4.0$.

また、記憶値 mem_i には次の2種類の値を保持する。

$$mem_i = \{Str_i, a_i\} \quad (14)$$

Str_i : 状態ストリング

$$(Str_i = \{0, 1, \#\}^L \quad \# : \text{Don't care symbol})$$

a_i : 動作識別コード ($a_i \in \{1, 2, \dots, 2^{N_c}-1\}$)

状態ストリング Str_i は、センサ入力ストリングに対応した長さ L のビットストリングである。5.3節で示すように、記憶値は個々の同調関係を分離するために用いられるので、同調関係ごとに異ならなければならない。ここでは、同調関係をロックした際のセンサ入力ストリングと、起動した動作の識別コードを用いる。

5.2 基本ダイナミクス カオス要素の非線形振動子部分の基本ダイナミクスを示す。

$$\left. \begin{aligned} x_i[t+1] &= c_0 f(x_i[t], c_1, \theta_x) - c_2 y_i[t] + c_3 x_i[t] \\ &\quad + \varepsilon(\bar{x} - x_i[t]) + A \sin(\bar{x} - mem_i) t + I_i S_i \\ y_i[t+1] &= g(x_i[t+1], c_4, \theta_y) \end{aligned} \right\} \quad (15)$$

y_i は、 x_i のピークに合わせて位相パルスを出力する。式(15)第1式の第4、5、最終項を、それぞれ結合項、想起入力項、活性制御項と呼ぶ。活性制御項によってカオス要素の活性・不活性状態を制御できる。図4のリターンマップからわかるように、 $A=\varepsilon=0$ としたとき、カオス要素は $I_i \cdot S_i = 0.664$ 以下で安定不動点を持つために不活性化される。活性制御項の二つの一時変数は、それぞれ次のように与えられる。

$$I_i = \begin{cases} 1: & \theta_m < m_i \\ 0: & \text{otherwise} \end{cases} \quad (16)$$

$$S_i = g(I_i \hat{b}_i, c_5, \theta_b) \quad (17)$$

式(16)の m_i は、状態ストリング $Str_i = \{str_i^1, \dots, str_i^L\}$ と、センサ入力ストリング $Sen = \{sen_1, \dots, sen_L\}$ とのマッチ率で、次式で計算する。 $\oplus$ は排他的論理和を表す。

$$m_i = \sum_{j=1}^L str_i^j \oplus sen_j / L \quad (18)$$

式(17)の $\hat{b}_i$ は、5.4節で示す競合パラメータである。ゆえに、記憶している状態ストリングがセンサ入力ストリングに一定ビット数以上マッチし、競合中の要素は活性状態を維持して振動を持続する。無限定要素は、状態ストリングの全ビットが $\#$ なので、どの状態でも

活性化する。したがって、新たな動作を発生させようとする無限定要素と、記憶している動作を想起しようとする限定要素の間に競合が生じる。

5.3 同調関係の形成と記憶 各要素は、二つの大域場を通じて相互作用する。結合項は、このうち大域振動場 $\bar{x}$ との結合量を与える。 $\bar{x}$ は、系全体での荷重平均振動量で、次式で表される。

$$\bar{x} = \frac{\sum_{e_j \in E} b_j[t]x_j[t]/n_j}{\sum_{e_j \in E} b_j[t]/n_j} \quad (19)$$

n_i は、要素 e_i が属している同調関係の構成要素数で、動作識別コード a_i から求める。各状態において活性化したカオス要素は、大域振動場を通じて全要素と相互作用する。無限定要素($u_i=u_0$, $n_i=1$)では、大域結合係数は $\varepsilon=0.08$ であり、図1のように、同調関係はカオス遍歴により動的に変化する。一方、動作ルールとして同調関係を記憶する場合、 ε を大きくして同調をロックする。FujisakaとYamadaの手法⁽¹⁰⁾によると、式(1)のような結合系では、 ε がカオス要素の最大リャプノフ指数⁽¹¹⁾より大きければ同調は安定となる。そこで、 $\varepsilon=0.25$ とする。なお、図4(a)のパラメータで計算した最大リャプノフ指数は0.147である。しかし、 ε を大きくしただけでは、異なる動作ルールである他の同調関係とも同調してしまう。そこで、大域結合係数に相関の強い変動を与えられたカオス要素群が同調しやすくなる機構⁽¹²⁾を応用し、同調関係を分離する。ここでは、想起入力項に変動を与えて、同様な効果を得る。

同調関係を形成して動作を発火した無限定要素群は、互いに同一の状態ストリング Str_i と、動作識別コード a_i を記憶する。 Str_i を10進数にデコードしたうえで、 a_i と加算し、式(15)第1式の想起入力項に作用させる。

$$mem'_i = Str'_i + 0.1a_i \quad (20)$$

全要素が共通して影響を受ける大域振動場 $\bar{x}$ と、同調関係ごとに異なる mem'_i の差が、想起入力項の正弦関数に入力される。等しい記憶値 $mem'_i = \{Str'_i, a_i\}$ を持つ要素群は、想起入力項に等しい変動を受けるため、同調を維持できる。一方、記憶値の差異により生じるアトラクタ間の偏差は、カオス性によって拡大されるため、同調関係を分離し得る。こうして、要素間に結合がないにもかかわらず、同調関係を記憶できる。

5.4 同調関係の競合 ある状態で活性化した要素は、自己が形成している同調関係だけを系内に存続させるべく競合する。そのため、各要素は式(21)で与えられる振動出力荷重 b_i の平均場を通して、全要素と相互作用を行う。

$$\bar{b}[t] = \frac{\sum_{e_j \in E} b_j[t]/n_j}{\sum_{e_j \in E} S_j[t]/n_j} \quad (21)$$

この値と b_i の比として、競合パラメータ $\hat{b}_i$ を定める。

$$\hat{b}_i = \frac{b_i[t]}{\bar{b}[t]} \quad (22)$$

競合パラメータは、式(17)で与えられる活性制御項に作用し、要素の活性・不活性を決める。 b_i および振動場との相関度 d_i は、それぞれ以下のように更新する。

$$b_i[t+1] = \frac{u_0 + \alpha \cdot m_i u_i[t]}{2} \cdot [\hat{b}_i(1 - \beta + y_i[t+1]d_i[t]) + \beta \cdot d_i[t]] L_i S_i \quad (23)$$

$$d_i[t+1] = f(\bar{x} - x_i[t], c_0, 0) \lambda^{1-\hat{b}_i^2} \cdot d_i[t](1 - y_i[t]) + y_i[t] \quad (24)$$

式(23)によれば、有効度 u_i 、マッチ率 m_i 、および大域振動場との相関度 d_i が大きい要素の振動出力荷重 b_i は増加する。すると、式(19)により、この要素の振動が大域振動場で占める比率が高まるので、式(24)から相関度はさらに増加し、この要素の競合パラメータは一層増していく。逆に、他の要素は競合パラメータが小さくなるため不活性化していく。最終的には一つの同調関係だけが振動を持続する。このとき、勝者の振動がそのまま大域振動場の振動となるので、カオス要素はこの状態を次式で個別に評価する。

$$F_i = h(\lambda^{[\bar{b}[t+1]-2\bar{b}[t]]}, \theta_f) \quad (25)$$

$F_i=1$ をとった個々の要素は、競合の勝者として属しているコラムのエフェクタ操作を別個に発火する。複数の要素が同調している場合には、複数の操作が同時に発火し、ロボットが動作する。

ところで、式(23)には b_i の発展を平均値 $\bar{b}$ の近傍に抑える性質がある。また、同調アトラクタどうしもカオス性により周期を変えながら引き込みあっており、有効度の小さな同調アトラクタでも、複数が位相を合わせて協調すれば大域振動場の支配率が増える。したがって、有効度の差が大きい場合には、有効度の大きい同調関係が競合に勝つとは限らない。このように、競合の勝者を決定するために要素間で有効度を直接比較したり、ゆらぎを与えるために改めてランダム性を加える必要はない。これらの2点は、二つの大域平均場を通じた相互作用によっておのずと実現される。

5.5 カオス要素の再生 カオス要素の再生には、二つのタイプがある。まず、無限定要素が記憶を行い関係を限定する際に、新たな無限定要素をコラム内に再生する。一方、センサ入力ストリングに状態ストリングが一致していなくても($\theta_m < m_i < 1.0$)、有効度が大きければ、要素は競合に勝ち動作を発火する場合がある。その場合、要素は汎化要素を再生する。親要素は、自己の状態ストリングのうち、センサ入力ストリングにマッチしていないビットを#に置き換え別の状態ストリング $Str_{Rep} = \{str_R^1, \dots, str_R^L\}$ を作り、汎化要素に与える。

$$\text{for } j \in \{1, \dots, L\} \\ str_R^j = \begin{cases} \# & : str_i^j \neq sen_j \\ str_i^j & : str_i^j = sen_j \end{cases} \dots \dots \dots (26)$$

これにより、汎化要素は再生された状態と、親要素が記憶している状態を含む、複数の状態でマッチ率1.0をとり活性化できる。なお、汎化要素の有効度は、親要素の有効度とマッチ率から、 $u_i \cdot m_i$ とする。

コラム内にすでに同じ要素がある場合と、動作に対して罰が与えられた場合を除き、再生された要素は親と同じ動作コラムに記憶される。

5.6 動作ルールへの信頼度割り当て 動作に対する評価として、要素にはペイオフ P が与えられる。ペイオフには、報酬と罰がある。報酬は、系が目標状態に到達した際にのみ与えられ、要素は有効度を増す。一方、罰を受けると有効度を減らす。ペイオフは、個々の動作に対して即座に与えられるとは限らず、動作列に対する評価として遅れて与えられるので、動作ルールに対して時間的に信頼度を割り当てる機構⁽¹³⁾が必要となる。要素は、有効度 u_i を次式で調整する。

$$u_i[t+1] = u_i[t] + (\bar{b}[t] - u_0)F_i \\ \cdot \{h(r[t], \theta_f) - I_i\} + r_i[t] \cdot P \dots \dots \dots (27)$$

$$r_i[t+1] = \gamma^{F_i} \cdot S_i \cdot r_i[t](1 - I_i) + I_i \cdot F_i \dots \dots \dots (28)$$

ペイオフは、式(27)にしたがい、過去複数の状態での競合の勝者($\theta_b < r_i$)に分配率 r_i で分配される。分配率は、式(28)によって状態遷移ごとに割引率 γ で削減される。また、動作発火時には大域場 $\bar{b}$ に競合の勝者の有効度の一部があるが、式(27)にしたがい、直前の状態での競合の勝者がこの量を得る。このように、カオス要素は、直接的な報酬獲得とBucket Brigade法⁽¹⁴⁾的な有効度伝播により、信頼度を割り当てられる。なお、限定要素は、系が目標状態に到達した場合にも一定の割合 δ で有効度を減らす。有効度がしきい値以下($u_i < u_{min}$)になった要素は死滅するので、報酬の獲得に寄与しない要素は除去される。系は、再生と死滅により要素の構成を自律的に変化させながら、報酬獲得に至る動作ルール列を同調関係列として自己組織していく。有効度の伝播も大域場を介して行われ、要素間どうしでの直接的な授受はない。要素間に結合は全くないため、再生と死滅による要素の追加・削除は容易である。

6. 計算機実験

6.1 タスクの記述 第3章で示した自律移動ロボットを用い、ナビゲーション・タスクについて計算機実験を行った。試行では、スタートからゴールである光源到達までを1エピソードとし、ロボットが光源に到達した場合と、500回動作を発火してもゴールに到達しない場合に、ロボットをスタート地点に戻してエ

Fig.5 Number of collisions

Fig.6 Number of generated rules and chaotic elements

ピソードを更新する。スタート地点は各エピソードを通じて不動であるが、ロボットはスタート地点に方向をランダムにして置かれる。また、ゴールの光源は、図7(a)に示した円内でのみ感知可能とする。なお、ゴールの位置や方向などの環境に関する予備知識や、センサに入力がない場合の対処のしかたといった予備的な動作ルールは、ロボットには全く与えない。したがって、試行の初期段階においては、たとえロボットが光源を感知したとしても、光源に向かうとは限らない。

ペイオフは次のように設定した。ロボットが光源に到達した場合の報酬は50.0とし、壁や障害物に衝突した場合の罰は、各要素の有効度の4%とした。また、デッドロックやループを避けるため、動作を発火させた要素は、発火コストとして有効度を1%減らす。

なお、カオス要素は総数100まで増殖可能とした。第5章で記した以外のダイナミクスのパラメータは、次のとおりである：

$$c_5=0.02, c_6=1.0, A=0.2, \alpha=0.1, \beta=0.35, \theta_f=0.999, \\ \theta_m=0.66, \theta_b=0.3, \lambda=\gamma=0.8, \delta=0.96, u_0=40.0, u_{min}=19.0.$$

6.2 実験結果 各エピソードにおける衝突回数の変化を図5に示し、カオス要素の総数と、各エピソードで発生された動作ルール数の変化を図6に示す。また、図7に、試行中の四つのエピソードでのロボットの軌跡を示す。これらの図から、各エピソードにおける衝突回数、動作ルール発生数ともに、試行を通じ

大勢として減少している傾向がわかる。

1回目のエピソードでは、動作発火ごとに観測される状態のほとんどが未知の状態である。これに対し、無限定要素はエフェクタ駆動操作間に関係を形成することによって、即応的に動作を発生する。その結果、計224個の動作を新規に発生して、要素数は試行開始時の3個から90個に増殖した。一方、ロボットは作業環境内をさまようような行動を見せ、左下隅で壁と何度も衝突している[図7(a)]。罰は、計158回受けた。このように、無限定要素が即応的に発生する動作は、その局面での環境状態と整合性があるとは限らない。

エピソード36[図7(b)]では、光源を感知した3状態で、3個の動作(000000101→右旋回, 00000#110→左旋回, 000#00#11→前進)が発火し、光源にロボットを導いている。それらの動作を発火させたカオス要素の有効度は、それぞれ、204.7, 250.1, 359.5と初期値 $u_0=40.0$ に比べて大きく、光源への接近行動は十分に強化されている。さらに、衝突回避行動も獲得されてきており、総動作数256回に対し、衝突回数は30回と少なくなってきた。このエピソードでは、左側あるいは左前方に壁を感知した場合に右旋回するという動作(たとえば11##00000→右旋回)が強い。また、図7(b)からは、何らかの形で壁沿いに移動できなければ、回廊部(a-b-c-d)や開口部(e-f)を通り抜けて光源に近づくのは難しいことがわかる。

これに対し図7(c)では、ロボットは回廊部において100#0#100→前進という動作ルールを4回発火させ、両側に壁を感知した状態でうまく直進して光源に到達した。その結果、この動作ルールを記憶しているカオス要素が報酬16.5を得て強化された。計68回の動作発火に対して、衝突も1回のみである。図7(b)(c)に示したロボットの軌跡からは、このような壁沿いに直進する行動が獲得されている様子が確認できる。(b), (c)の時点で観察できた壁沿い直進行動は安定せず、衝突回数もエピソード50から60にかけて再び増加した。しかし、ゴールへの接近行動は十分強化されているので報酬は比較的得られやすく、エピソード65から75にかけて再び壁に沿った直進行動が見られた[図7(d)]。エピソード74では、たとえば左側と前方に壁を感知した場合に右回転(#1101#000→右回転)といった衝突回避行動がなされており、衝突回数は0である。

6・3 考察 大域結合カオス系を構成するカオス要素は、同調している相手さえ認識しない。各要素は、内在的なダイナミクスに基づいて、自己の内部状態と二つの大域場の状態から、次の時点での内部状態を決定しているにすぎない。こうした個々のカオス要素のボトムアップ的な挙動から、ある状態でのロボットの

(a) Episode1

(b) Episode36

(c) Episode41

(d) Episode74

Fig.7 Behavior of robot

動作が定まり、ロボットと外部環境との相互作用が生まれる。さらに、この相互作用の結果が、時間遅れを伴うペイオフとしてカオス要素に還元され、要素の存続や死滅が決まる。こうした一種の創発的機能形成の循環⁽¹⁵⁾を通じ、より高い有効度を持つカオス要素群が自己組織する結果、ゴールへの接近、壁沿い移動、障害物回避などの行動を、ロボットが獲得したといえる。

次に、図6の結果を検討する。センサ数から想定し得る最大状態数512に対して、この試行でロボットが実際に感知した状態数は、142状態であった。観測した状態に対して動作を決める本手法では、142状態×7動作の範囲でのみ動作ルールを発生したことになる。しかし、試行を通じた動作ルール発生数は512より多いので、同じ動作ルールを複数回発生したことになる。一方、最終エピソードでは、動作ルール54個のうち、#記号を含むルールは28個あり、状態空間の汎化が進んでいる。しかし、数は少なくなっているが、試行の後半においても新しいルールをなお発生しており、全要素数も限度である100に近い。以上の結果から、動作列が定まらない場合には、ルールの記憶容量(要素の増殖限度数)が少ないと考えられる。

今回の実験では、ロボットは各状態における動作のしかたを強化学習する。そして、ロボットは状態に応じて適切な動作発火を繰り返す結果、ゴールに到達する。その一方、学習システムは特定の動作列を学習できるため、最短ではなくとも特定の移動経路でゴールに到達するようになることも想定された。しかし、今回の実験では確認されなかった。その理由として、スタート地点へのロボットの配置角がランダムであり、ゴールへの接近経路も多数存在することが挙げられる。

もう一つの理由として、両側に壁を感知しながら直進するといった、支配的な行動の持続期間が短いことがある。スタートからゴールまでの状態-動作列上で、センサ入力の状態ストリングに一定値以上マッチし、競合に勝ちさえすれば、各動作ルールは、どの状態でも発火し得る。動作発火の際、カオス要素は有効度の一部を大域場に放出するので、状態-動作列上で幾度も発火しながら報酬の伝播を受けられなかった動作ルールは、有効度を減らす。その結果、こうした動作ルールから導かれる行動も、不安定化しやすくなる。ただし、特定の動作列でゴールに到達するようになれば、行動は安定化していくので、両者は密接に関係している。また、大域結合カオス系では、より多く報酬を分配された要素が自己組織していく。したがって、実験でも確認されたように、一旦は行動が不安定化しても、ロボットはいずれ何らかの行動を獲得する。

7. おわりに

自律移動ロボットを対象に、大域結合カオス系を応用した新しい制御アーキテクチャを構築した。系を構成する要素のダイナミクスには、カオス振動子の性質に加えて、系が全体として強化学習をするように、競合と信頼度割り当ての機構を新たに組み込んだ。また、学習の進展に応じて系の構成を自律的に変えられるように、要素には再生と死滅の機能を与えた。その結果、ナビゲーション・タスクを扱った計算機実験では、報酬獲得に寄与する動作を記憶したカオス要素が、自己組織していくことが確認できた。さらに、このカオス要素群の自己組織的な挙動の結果、環境に関する先見的な知識や動作ルールが全くない状態から、ロボットがゴールへの接近、壁沿い移動、障害物回避といった行動を獲得していく様子が観察できた。今後は、行動がより安定するよう信頼度割り当てのダイナミクスを改良し、動的環境での行動学習に取り組む。

文 献

- (1) R.A. Brooks, A Robust Layered Control System for a Mobile Robot, *IEEE Robotics and Automation*, 2-1(1986), 14-23.
- (2) Long-Ji Lin, Scaling Up Reinforcement Learning for Robot Control, *Proc. of the 10th Inter. Conf. on Machine Learning*, (1993), 182-189.
- (3) D. Floreano, F. Mondada, Evolution of Homing Navigation in a Real Mobile Robot, *IEEE Trans. on Systems, Man, and Cybernetics*, B26-3(1996), 396-407.
- (4) B.J.A. Verschure, B.J.A. Kröse and R. Pfeifer, Distributed adaptive control: The self-organization of structured behavior, *Robotics and Autonomous Systems*, 9(1992), 181-196.
- (5) 金子・津田, 複雑系のカオス的シナリオ, (1996), 朝倉書店.
- (6) 山口・吉原・若松, カオスの想起に基づく知的ロボットの試作, 第5回インテリジェント・システム・シンポジウム講演論文集, (1995), 307-312.
- (7) 清水, 生命科学から見た生命, 清水ほか編, 生命と科学, 宗教と科学, 6(1993), 1-54, 岩波書店.
- (8) 湯浅・伊藤, 自律分散システム, 日本ロボット学会誌, 10-4(1992), 464-467.
- (9) Kaneko K., Clustering, coding, switching, hierarchical ordering, and control in a network of chaotic elements, *Physica D*, 41(1990), 137-172.
- (10) Fujisaka H. and Yamada T., Stability Theory of Synchronized Motion in Coupled-Oscillator Systems, *Prog. Theore. Phys.* 69-1, (1983), 32-47.
- (11) T. S. Parker and L.O. Chua, *Practical Numerical Algorithms for Chaotic Systems*, (1989), 66-82, Springer Verlag.
- (12) Kaneko K., Relevance of dynamic clustering to biological networks, *Physica D*, 75(1994), 55-73.
- (13) J.H. Connell and S. Mahadevan eds., *Robot Learning*, (1993), Kluwer Academic Publishers.
- (14) J.H. Holland, Properties of the bucket brigade algorithm, J.J. Grefenstette ed., *Proc. of the 1st Inter. Conf. on Genetic Algorithms and Their Applications*, (1985), 1-7.
- (15) 伊藤, 自律ロボット・進化ロボット, 電子情報通信学会誌, 77-2(1994), 155-160.