通信機能を導入した集団学習オートマトンによる制御学習

非会員 外 山 達 斎 (明 電 舎)

正 員 小 坪 成 一 (千 葉 大)

正 員 銭 飛 (広島電機大)

正 員 平 田 廣 則 (千 葉 大)

Learning of Controlling Behavior Using Learning Automaton Networks
with Communications

Tatsushi Toyama, Non-member (Meidensha Corporation), Seiichi Koakutsu, Member (Chiba University),
Fei Qian, Member (Hiroshima-DENKI Institute of Technology), Hironori Hirata, Member (Chiba
University)

The learning automaton is one of the best known learning systems which is capable of operating in highly uncertain random environments. In this paper, we propose a new learning system which accelerates the learning speed by introducing communications between learning automaton networks. In the proposed system, learning automaton networks can quickly respond to the dynamical changes of the random environment because they can share the information about the random environment by communications. We apply the proposed system to a learning problem of controlling the behavior of artificial insects. Computational experiments show that proposed system obtains better results in terms of the convergence speed comparing conventional learning automaton networks without communications.

キーワード：学習オートマトン，ニューラルネットワーク，強化学習，制御学習，人工生命

1. ま え が き

学習者に対して未知である外的環境を同定する学習モデルの一つとして，学習オートマトン⁽¹⁾が挙げられる。学習オートマトンとは，Turing が考案した有限状態オートマトンの状態遷移と出力の決定に確率性を導入した確率オートマトンに，未知環境を組み合わせた学習モデルである。

この学習オートマトンにおける学習は，確率オートマトンのもついくつかの状態の中で，どれを出力するかを決定するための確率分布(出力選択確率分布)を修正することにより行われ，その目標は未知環境に対して最も有効な状態を出力しやすくすることである。その具体的な方法は，確率オートマトンが出力選択確率分布に従って一つの状態を未知環境に出力し，未知環境がその出力に対する評価を行い，その評価を入力として確率オートマトンに与える。確率オートマトンは，その出力がうまくいったと評価されると，その出力を選択する確率を大きくし，そのほかの出力確率を小さくすることにより，学習する。

学習オートマトンは，確率オートマトンと未知環境を一つずつ組み合わせた基本モデルにおいても十分な学習性能

を有し，学習の収束性が理論的に保証されたモデルである。しかしながら，学習オートマトンでは，未知環境からの応答に応じて自らの行動選択確率を更新する試行錯誤的な学習法を用いるため，収束までに多大な学習ステップを要する。また，学習オートマトンの入力は，未知環境からの一入力なので，未知環境以外の外部からの情報やほかのオートマトンとの相互作用に関する情報を直接入力できない。

これらの問題点を克服するために，複数入力の学習オートマトンによるネットワークを構成した集団学習オートマトン⁽²⁾およびベイズ識別法に基づいた強化法⁽³⁾が提案されている。この集団学習オートマトンでは，個々の学習オートマトンの有する自律性を活かした高度な学習が可能である。また，未知環境が緩やかに変動し，その応答に雑音を加わる場合でも，学習が可能である。しかしながら，未知環境が急激に変動する場合にはその変動に十分に追従できない場合がある。

そこで，本論文では，上述の複数入力学習オートマトンネットワークを多数用いて，その学習オートマトンネットワーク間に通信機能を導入する手法を提案する。提案手法により，未知環境に関する情報を多数の学習オートマトン

ネットワーク間で共有することで、学習の収束速度が向上し、第5章に例題として示す人工生物の制御学習問題において、餌(えさ)と虫が同じ移動速度で動くような、比較的急激に変動する未知環境に対しても十分な追従性を示すことが期待できる。

本モデルの応用例として、人工生物の制御学習を取り上げ、簡単なシミュレーションを行う。また、強化学習の代表的手法であるQ学習⁽⁴⁾を取り上げ、学習オートマトンによる手法との比較を試みる。

2. 複数入力学習オートマトンネットワーク

学習オートマトンを複数入力へと拡張し、外部からの複数の入力を可能とするとともに、学習オートマトン間の相互作用を有するような、複数入力学習オートマトンネットワーク⁽²⁾が提案されている。以下に、 N 個の複数入力学習オートマトンから構成されるネットワークを定義し、その構成例を図1に示す。

(定義1) N 個の複数入力学習オートマトンで構成されるネットワークは、以下の形式で定義される。

SA^i : i 番目の複数入力学習オートマトン

$N(i)$: SA^i の隣接オートマトンの集合

Y^i : SA^i の出力集合

$Y^i = \{-1, +1\}$

x^i : SA^i への入力ベクトル

$x^i = (y^j | y^j \in Y^j, j \in N(i))^T$

w^i : SA^i の入力重みベクトル

$w^i = (w_j^i | j \in N(i))^T$

Φ^i : SA^i の状態集合

$\Phi^i = \{\phi_1^i, \phi_2^i\}$

p^i : SA^i の出力選択確率分布

$p^i = (p_1^i, p_2^i)^T$

T^i : SA^i の強化法

$p^i(t+1) = T^i(w^i(t), x^i(t), y(t), r(t))$

o^i : SA^i の出力関数

$o^i : \phi \mapsto y^i$

r : ネットワークの強化信号

$r \in R = \{-1, +1\}$

RE : 未知環境 (ペナルティ確率分布 C をもつ)

図1に示すように、複数入力学習オートマトンは外部からの入力ベクトル x とその重みベクトル w をもち、出力

図1 学習オートマトンネットワークの構成例
Fig.1. An example of learning automaton network.

選択確率分布 $p^i = (p_1^i, p_2^i)$ は、次式のように x と w により表現する。

$$p_1^i = \Phi\left(-\sum_{j \in N(i)} w_j^i x_j^i\right) = \Phi(-w^{iT} x^i) \quad \dots\dots\dots (1)$$

$$p_2^i = 1 - p_1^i \quad \dots\dots\dots (2)$$

ここで、 $\Phi(\cdot)$ は $[0, 1]$ の単調増加関数であり、本モデルではシグモイド関数を用いる。また、未知環境からの応答は各オートマトンに返送される。

各複数入力学習オートマトン SA^i の目標は、入力パターン x に応じて、未知環境 RE からできるだけペナルティ応答を受けないような出力 $y^i \in Y^i$ を選択することである。しかし、未知環境のペナルティ確率分布は事前にわからないため、各 SA^i は自分が出した出力およびそれに対する未知環境からの応答に基づき、強化法 T^i を用いて、重み w^i すなわち出力選択確率 p^i を修正する。この場合、強化法 T^i は入力の重みベクトル w^i の関数でなければならない。本研究では、パターン認識理論でよく用いられるベイズ識別法に基づいた以下のような強化法⁽³⁾を利用する。

強化法 T^i

$$w(t+1) = w(t) + \Delta w$$

if $r(t) = 1$ (リワード強化)

$$\Delta w = -\theta(t) \{ \tau(w^T(t)x) - r(t)y(t) \} x \quad \dots\dots\dots (3)$$

$$p_1(t+1) = \Psi(-w^T(t)x + \theta(t) \|x\|^2 (w^T(t)x - 1)) \quad \dots\dots\dots (4)$$

$$p_2(t+1) = 1 - p_1(t+1) \quad \dots\dots\dots (5)$$

if $r(t) = -1$ (ペナルティ強化)

$$\Delta w = -\lambda \theta(t) \{ \tau(w^T(t)x) - r(t)y(t) \} x \quad \dots\dots\dots (6)$$

$$p_1(t+1) = \Psi(-w^T(t)x + \lambda \theta(t) \|x\|^2 (w^T(t)x - 1)) \quad \dots\dots\dots (7)$$

$$p_2(t+1) = 1 - p_1(t+1) \quad \dots\dots\dots (8)$$

ただし、 λ および $\theta(t)$ は学習パラメータであり、 $\tau(\cdot)$ は、以下のような関数とする。

$$\tau(\xi) = \begin{cases} -1, & \xi \leq -1 \\ \xi, & -1 < \xi < 1 \\ 1, & \xi \geq 1 \end{cases} \quad \dots\dots\dots (9)$$

以下に、ネットワークの動作とそれに伴う各オートマトンの一般的な学習アルゴリズム⁽²⁾⁽³⁾を示す。

アルゴリズム

- (1) ネットワークの構造 $N(i)$ を与える。
- (2) 時刻 $t=0$ ですべての学習オートマトン $SA^i (i=1, 2, \dots, N)$ に対する重み w_j^i を一様乱数で初期化する。
- (3) 入力ベクトル x を入力層の各学習オートマトン SA^i に与える。
- (4) ネットワークの最下層(パターン入力層)から順に、確率分布 $p^i(t) = (p_1^i, p_2^i)$ に従って、出力 $y^i(t) \in \{-1, +1\}$ を決定する。
- (5) 環境からの応答 $r(t)$ を受け、(1)式および(3)式の強化法 T^i に基づいて、 $w^i(t)$ を修正する。

(6)あらかじめ設定された終了条件を満たすまで(3)～(5)を繰り返す。

3. 通信機能の導入

本論文では、学習オートマトンネットワークにおける学習の収束速度の向上および急激に変化する動的環境に対する追従性の向上のために、学習オートマトンネットワークを多数用い、その学習オートマトンネットワーク間に通信機能を導入する手法を提案する。本章では、前述した単体の学習オートマトンネットワークを一匹の虫に割り当てることを考える。ここでは、簡単のため虫の数が二匹の場合(A₁, A₂)を考える。虫1は、未知環境に関する情報を虫2に通信する。前述のように、複数入力学習オートマトンネットワークにおける各学習オートマトンは、入力加重ベクトルを修正することで学習する。この学習法に基づいて、虫2は、虫1からの通信信号により、自身の加重ベクトルを修正する。ここで、強化法 Tⁱにおいて、(3)式および(6)式を以下のように修正する。

$$w_{A_2}(t+1) = w_{A_2}(t) + \Delta w_{A_2}$$
$$\text{if } r(t) = +1 (\text{リワード強化})$$
$$\Delta w_{A_2} = -\theta(t) \{ \tau(w_{A_2}^T(t) x_{A_2}(t)) - r(t) y_{A_2}(t) \} x_{A_2}(t) + a(CS - w_{A_2}(t)) \dots\dots\dots (10)$$

$$\text{if } r(t) = -1 (\text{ペナルティ強化})$$
$$\Delta w_{A_2} = -\lambda\theta(t) \{ \tau(w_{A_2}^T(t) x_{A_2}(t)) - r(t) y_{A_2}(t) \} x_{A_2}(t) + a(CS - w_{A_2}(t)) \dots\dots\dots (11)$$

ここで、
 $CS = w_{A_1}(t) \dots\dots\dots (12)$
は通信信号を表し、aは通信信号の加重パラメータである。なお、虫数が三匹以上で通信を行う場合、たとえば虫2がほかの複数の虫から通信信号を受信する場合は、(10)式および(11)式の(CS - w_{A₂}(t))をΣ_{i=A₂}(CS_i - w_{A₂}(t))として加重ベクトルを修正すればよい。ここで、CS_iは虫iからの通信信号を表す。

このように、複数の複数入力学習オートマトンネットワーク間に通信を形成させると、未知環境に関する情報を共有することができるため、学習速度の向上が期待できる。また、学習速度が向上するため、第5章に例題として示す人工生物の制御学習問題において、餌と虫が同じ移動速度で動くような、比較的急激に変動する未知環境に対する追従性の向上も期待できる。

4. 人工生物の制御学習への応用

ここで、学習オートマトンネットワークおよび通信機能の有効性を検証するために、制御学習の一例⁹⁾について考える。本研究で扱う制御学習は、人工生物(虫)の空間移動に関する簡単なモデルである。ここで、制御の対象となるものが虫であり、学習オートマトンネットワークの目的は、虫のおかれている環境(位置)に合わせて虫を移動させ、餌を探索することである。

〈4・1〉 未知環境の設定 図2は、虫の移動平面を真

図 2 虫の移動空間
Fig. 2. The movement space.

上から見た図である。虫の移動平面は500×500=250,000個の格子点で構成され、二種類の香りで満たされているとする。まず、一つ目の香りは、虫を引きつける餌の香りである。この香りは、餌を中心に同心円状に広がり、餌から離れるに従って一次関数的に減衰する。餌を中心においた場合、香りは中心で最大値1、中心からもっとも離れた四隅で最小値0になる(図2の実線)。

もう一つの香りは、虫の位置を示すために各方角から発する香りである。これは、東西南北の4方向から発する香りで、それぞれ中心から周囲に広がり、餌の香りと同様に中心が1、最も遠い二つの角で0になるように一次関的に減衰する(図2の点線は、西を例にした方角の香りの概念を示す)。

虫は、自分の経験を他人に伝えるために、通信信号として声を発する。虫は、この声を聞く聴覚をもっている。通信信号は、上述の香りと同様に自分自身を中心とした同心円状に広がり、自分自身から離れるに従って一次関的に減衰する。

〈4・2〉 学習オートマトンネットワークの応用 これら二種類の香りおよび声のうち、虫を引きつける「餌の香り」は、学習オートマトンネットワークにおいては、未知環境が虫の移動結果に対する評価を行うときのみ用いられる。また、「各方角からの香り」は、ネットワークへの入力パターンxとして使用される。なお、虫(または各学習オートマトン)には餌の香りの分布がわからず、環境からの強化信号を受け取るまで、移動方向が適切であるか否かが判断できない。

シミュレーションにおけるネットワークの構成を図3に示す。各構成要素は以下になる。

- x^i : 現在の虫の位置における各方角iからの香りの強度
- w^i : 複数入力学習オートマトンへの方角iについての加重ベクトル
- SA^i : 方角iに対応する複数入力学習オートマトン発火($y^i = +1$)することにより、方角iに向かって移動するよう指示する
- y^i : 方角iに対応するSAⁱの出力

図 3 制御学習ネットワーク
Fig. 3. The network for learning of controlling behavior.

RE : 未知環境

リワード確率分布を香りの強度に比例して設定する

r : 強化信号

CS : 通信信号

なお、図 3 における $i=N, S, E, W$ はそれぞれ北、南、東、西を表す。

四つの学習オートマトン SA^i は、時刻 t で、現在の虫の位置における各方角からの香りの強度を入力ベクトルとして受け取り、(1)式および(2)式により、出力選択確率分布 p^i を計算し、これに従って出力 y^i を決定する。各 SA^i は、 $y^i(t)=1$ (発火)を出力することにより、東西南北への移動を指示し、虫は発火した SA^i の示す方角へ移動する。なお、複数の学習オートマトンが同時に発火した場合、虫はそれらを合成させた方向へ移動する。例えば、北と東が同時に発火した場合は北東へ移動し、北、東、南が同時に発火した場合は東へ移動する。

虫が移動を完了すると、未知環境は虫の移動先の餌の香りの強度が元の位置よりも強くなっていればリワード、そうでなければペナルティの強化信号 r を各 SA^i に送る。この強化信号に基づいて、各 SA^i は現在の重み w^i を(10)式および(11)式の強化法 T^i により修正する。重みの修正後、虫は自身のネットワークの重みを通信信号 CS として発声する。自分以外の通信信号により、各 SA^i は現在の重みを修正し、再び入力ベクトルと重みベクトルから出力選択確率分布を計算し、虫を移動させる。以上の動作を、虫が餌にたどりつくまで繰り返す。

なお、餌にたどり着いた虫は、餌のある方角をほかの虫に通信する。

5. シミュレーション実験

餌が1個および2個、虫が1匹〜4匹について、餌が静止している場合および不規則に移動する場合の実験を行う。それぞれについて重み w^i の初期値を変えて、100 回実験を行う。表 1、表 2 に、環境、通信機能と平均収束ス

表 1 平均収束ステップ数(餌 1 個)
Table 1. A mean value of convergence step (1 prey).

餌	通信 (虫数)	平均収束ステップ数(成功率[%])			
		スタート位置			
		(50,50)	(450,450)	(450,50)	(50,450)
静止	なし	960	1003	953	982
	(一)	(100)	(100)	(100)	(100)
	あり	903	910		
	(2匹)	(100)	(100)		
	あり	859	882	894	
	(3匹)	(100)	(100)	(100)	
	あり	740	767	772	757
	(4匹)	(100)	(100)	(100)	(100)
移動	なし	1009	1040	1025	1031
	(一)	(100)	(100)	(100)	(100)
	あり	986	987		
	(2匹)	(100)	(100)		
	あり	878	905	899	
	(3匹)	(100)	(100)	(100)	
	あり	781	811	781	793
	(4匹)	(100)	(100)	(100)	(100)

表 2 平均収束ステップ数(餌 2 個)
Table 2. A mean value of convergence step (2 preys).

餌	通信 (虫数)	平均収束ステップ数(成功率[%])			
		スタート位置			
		(50,50)	(450,450)	(450,50)	(50,450)
静止	なし	3211	3603	3105	3439
	(一)	(90)	(92)	(91)	(94)
	あり	2551	2507		
	(2匹)	(99)	(99)		
	あり	1983	1927	1954	
	(3匹)	(100)	(100)	(100)	
	あり	1488	1548	1484	1584
	(4匹)	(100)	(100)	(100)	(100)
移動	なし	3464	3441	3528	3493
	(一)	(73)	(73)	(70)	(65)
	あり	3122	3161		
	(2匹)	(82)	(82)		
	あり	2773	2711	2781	
	(3匹)	(85)	(85)	(85)	
	あり	2391	2370	2297	2268
	(4匹)	(94)	(92)	(94)	(90)

テップ数との関係を示す。ここで、表 1 は餌 1 個、表 2 は餌 2 個の場合の結果である。ステップ数は、図 2 の角 a を座標 (0, 0) とし、虫を (50, 50), (450, 450), (450, 50), (50, 450) の位置からスタートさせたときの餌への到達ステップ数の平均値である。成功率は、各試行において最大ステップ数 10,000 ステップ以内にすべての虫が餌に到達した場合を成功とし、最大ステップ数経過後にいずれかの虫が餌に到達していない場合を失敗とみなして算出した値である。表 1、表 2 より、通信機能を導入することにより収束速度が速くなり、餌 2 個の場合は餌への到達の成功率が高くなることがわかる。また、通信機能ありの場合は、通信を行う虫の数が増えるほど収束速度が速くなり、餌 2 個の場合の餌への到達の成功率が高くなることがわかる。図 4 に、餌が不規則に移動する場合の通信機能を導入した虫の移動軌跡を示す。また、学習結果の有効性を検証するための実験を行う。スタート位置 (50, 50) において、1 回の学習が終了後のネットワークの重みを使用し、再学習を行う。

表 3 に、学習と平均ステップ数との関係を示す。表 3 よ

図 4 虫の移動の軌跡(餌 2 個, 虫 4 匹)
Fig. 4. The behavior of the insects seeking moving preys.

表 3 学習の効果
Table 3. The effect of learning.

学習	通信	平均収束ステップ数			
		スタート位置			
前	なし	(50,50)	(450,450)	(450,50)	(50,450)
	あり	1019	1032	1019	1023
後	なし	860	917	811	776
	あり	1005	994	980	994
	なし	817	799	780	798
	あり				

り、スタート位置 (50, 50) における学習が、ほかのスタート位置においても有効であることがわかる。

6. Q 学 習

比較対象として、強化学習の代表的手法である Q 学習⁽⁴⁾を用いる。

虫の移動空間は格子で構成され、各格子の東西南北の 4 方向に Q 値を割り当てる。餌は中央に置かれている。虫は最大 Q 値の方向へ移動する。未知環境は、虫の移動結果に対する評価を行い、次式に従って Q 値を更新⁽⁴⁾⁻⁽⁶⁾する。

$$Q(x, t+1) = (1-\alpha)Q(x, t) + \alpha\{r(t) + \gamma \max_y Q(y, t)\} \dots\dots\dots (13)$$

ここで、 x は現在の状態、 y は次の状態、 γ は減衰率、 r は強化信号、 α は学習定数である。 γ および α は予備実験に基づき各環境で適切と考えられる値に設定する。図 5 に、移動空間の大きさと学習回数との関係を示す。ここでは、図 2 の角 a を座標 (0, 0) とし、餌は位置固定、虫のスタート位置を (1, 1) とする。また、学習回数(重みまたは Q 値の更新回数)は、学習オートマトンネットワークでは、スタートしてから餌に到達するまでのステップ数であり、Q 学習では、スタートしてから餌に到達するまでを 1 試行とし、最適戦略が得られるまでの試行回数とする。ここで、学習後に生成される経路は、Q 学習では最適戦略、すなわち最短経路であるのに対し、学習オートマトン

図 5 移動空間の大きさと学習回数の関係
Fig. 5. The movement space vs. a mean value of convergence step.

ネットワークでは出力選択の確率性のため準最適戦略となる。

餌が動かない静的未知環境においては、途中に侵入禁止区域を設けても、最適戦略が得られる。しかし、Q 学習は、餌にたどり着いたときだけ、環境から大きなリワードが得られ、それを次々と周囲に伝搬させていく手法である。つまり、リワードやペナルティはある長さの行動系列の実行結果として得られるのであって、学習者の意図決定の直後に(一回行動するごとに)評価が得られるわけではない。また、Q 値を収束させるためには、全空間を探索する必要がある。このため、餌が不規則に動く動的未知環境に対しては最適戦略が得られない。また、図 5 に示すように、移動空間が大きくなると、状態数が指数関数的に増加するため、試行回数(学習回数)が膨大になってしまう。一方、学習オートマトンでは、学習者の意思決定の直後に評価が得られるため、虫はスタート地点から餌に到達するまでにうまく学習が行われ、Q 学習と同様に準最適戦略で餌にたどり着くことができる。また、状態数は移動空間に依存しないため、学習回数が指数関数的に増加することはない。

7. む す び

本論文では、学習オートマトンネットワーク間に通信機能を導入する手法を提案した。本モデルの応用例として、人工生物の制御学習を取り上げ、従来の手法および強化学習の代表的手法である Q 学習と比較した。

その結果をまとめると、以下ようになる。

通信機能と収束速度 学習オートマトンネットワーク間に通信を形成し未知環境に関する情報を共有させると、収束速度が向上する。

Q 学習と学習オートマトン Q 学習は、環境が静的であれば、最適戦略に収束する。しかし、空間の大きさの増加に伴って状態数が指数関数的に増加するため、収束するまでに多大な学習時間が必要である。また、動的な環境に

対しては追従しにくいという欠点をもつ。一方、学習オートマトンは、環境の静的、動的にかかわらず最適戦略には収束しないものの、それに近い準最適戦略に収束する。また、単位行動ごとに評価を得られるので、学習時間が空間の大きさに依存して指数関数的に増加することはない。

以上のことから、通信機能を導入した学習オートマトンネットワークでは、通信内容を適切に選ぶことにより、どのような環境においても、従来の手法およびQ学習よりも良好な結果が得られるといえる。

しかし、現状ではいくつかの問題点がある。通信機能を導入して多数の虫間で環境に関する情報を共有すると、(10)式および(11)式に基づく重みの更新量 Δw が増加して重みの収束が速くなり、結果的に虫が餌の方向により速く、すなわち直線的に最短距離で移動するようになる。このとき、途中で障害物を設けるとそれを回避できず、通信機能がかえって虫の餌への到達を妨げることがある。また、未知環境においては「餌の香り」あるいは「各方角からの香り」なるものが存在しないことも十分考えられる。今後の課題として、これらを考慮した通信機能を構築することが挙げられる。

また、本論文で取り上げたような移動体の経路生成に対し、Q学習とボルツマン分布による意思決定を組み合わせた強化学習法⁽⁷⁾が提案されている。単なるQ学習との比較のみでなく、このような強化学習法と比較することで学習オートマトンを用いた手法の有効性をより明確にすることも、今後の課題の一つと考えられる。

(平成9年5月20日受付, 同9年12月5日再受付)

文 献

- (1) K. S. Narendra & M. A. L. Thathachar: "Learning Automata an Introduction," *Prentice-Hall International* (1989)
- (2) 銭 飛・海野富士也・平田廣則: 「学習オートマトンネットワークとそのダイナミクス」, 電学論C, **111**, No. 10, 490~499 (平3)
- (3) 銭 飛・海野富士也・平田廣則: 「動的な未知環境における学習オートマトンネットワーク」, 同上C, **115**, No. 8, 953~960 (平7)
- (4) C. J. C. H. Watkins & P. Dayan: "Q-Learning," *Machine Learning*, **8**, 279~292 (1992)
- (5) N. Ono, T. Ohira, & A. T. Rahmani: "Emergent Organization of Interspecies Communication in Q-Learning Artificial Organisms," *Proc. Third European Conference on Artificial Life*, 396~405 (1995)
- (6) N. Ono, O. Ikeda, & A. T. Rahmani: "Synthesis of Organizational Behavior by Modular Q-learning Agents," *Proc. 1995 IEEE/Nagoya Univ. World Wisepersons Workshop*

(WWW'95) on Fuzzy Logic and Neural Networks/Evolutionary Computation, 79~80 (1995)

- (7) L. Lin: "Self-improving Reactive Agents based on Reinforcement Learning, Planning and Teaching," *Machine Learning*, **8**, 293~321 (1992)

外 山 達 斎 (非会員) 1973年9月2日生。97年千葉大学大学院工学研究科電気電子工学専攻修士課程修了。同年(株)明電舎入社, 現在に至る。在学中, 集団学習オートマトン, 強化学習の研究に従事。現在, オブジェクト指向ソフトウェア利用技術の開発に従事。

小 坏 成 一 (正員) 1964年7月7日生。92年千葉大学大学院自然科学研究科生産科学専攻博士課程修了。同年同大工学部助手, 97年同助教授, 同年同大学院自然科学研究科助教授となり現在に至る。1994~95年, カリフォルニア大学サンタクルズ校客員研究員。VLSIレイアウト設計, 確率的最適化手法, ニューラルネットワーク, 学習オートマトンの研究に従事。博士(工学)。IEEE, INNS, 電子情報通信学会, 計測自動制御学会会員。

銭 飛 (正員) 1957年2月15日生。87年東京工業大学大学院修士課程修了。91年千葉大学大学院自然科学研究科博士課程修了。同年広島電機大学電子工学科講師, 94年同情報工学科助教授となり, 現在に至る。大規模システムのモデリング, 解析, およびコンピュータネットワークなどの分散システムの基礎的理論の研究に従事。Ph. D., IEEE, 計測自動制御学会, 電子情報通信学会会員。

平 田 廣 則 (正員) 1948年6月2日生。76年東京工業大学大学院電気工学専攻博士課程修了。同年千葉大学工学部助手, 81年同助教授, 94年同教授, 97年同大学院自然科学研究科教授となり現在に至る。大規模システムのモデリング, 解析と設計, 特に生態システム, VLSIレイアウト, 並びに分散型コンピュータの基礎的理論に興味をもつ。工学博士。IEEE, ISEM, INNS, 電子情報通信学会, 情報処理学会, 計測自動制御学会, システム制御情報学会, 人工知能学会会員。