S-モデル非定常環境中で動作する 可変階層構造学習オートマトン

最上 義夫*・馬場 則夫**・杉浦 公彦*

Variable Hierarchical Structure Learning Automata Operating in an S-model Nonstationary Environment

Yoshio MOGAMI*, Norio BABA** and Kimihiko SUGIURA*

The purpose of this paper is to construct a new learning algorithm for variable hierarchical structure learning automata operating in a nonstationary random environment by extending the relative reward strength algorithm proposed by Simha and Kurose. The learning property of our algorithm is considered theoretically, and it is proved that the optimal path probability can be approached to 1 as close as possible by using our algorithm. In numerical simulation, the number of iterations of our algorithm is compared with that of the variable hierarchical structure learning algorithm of L_{R-I} type proposed by Mogami and Baba, and it is shown that our algorithm can find the optimal path after the smaller number of iterations than that of the algorithm of L_{R-I} type.

Key Words: variable hierarchical structure learning automata, nonstationary random environment, relative reward strength algorithm, reward vector

1. ま え が き

未知周囲環境の自律的学習システムの一つとしての学習オートマトンの研究は、Tsetlin¹⁾以来、多くの研究者によって活発に進められてきた^{2)~4)}。その目的は、未知周囲環境から得る reward が、平均値の意味でできるだけ大きくなるような出力を選択するようになることである。ここで、未知周囲環境とは、学習オートマトンに対する reward の値が何らかの未知確率分布(reward 確率)に従って決定されると考えられるような環境を指す^(注1)。

このような未知周囲環境中において動作する学習オートマトンに関する研究は、多数行われており、その学習アルゴリズムとして、 L_{R-I} 強化法、 L_{R-P} 強化法および

L_{R-CP} 強化法などが提案された。さらに、近年、Thathachar and Ramakrishnan⁵⁾ が、学習オートマトンを本構造に組み合わせた階層構造学習オートマトンを提案し、P-モデル定常環境中で良好な学習特性をもつことを示した。また、Thathachar and Sastry は、P-モデル定常環境の reward 確率のサンプル平均値を用いる estimator algorithm を提案し⁶⁾、さらに、S-モデル定常環境中で動作する階層構造学習オートマトンへの拡張を行った⁷⁾。また、非定常環境に対しては、馬場、最上、三宅⁸⁾ が、P-モデル非定常複数教師環境中で動作する階層構造学習オートマトンを提案し、その学習特性を考察した。さらに、現在の reward のみならず過去に得た reward をも学習に用いる学習アルゴリズムとして、Simha and Kurose⁹⁾ は、各動作が S-モデル非定常環境から得た最新の reward を reward vector に記憶し、それを用いて学習を行う Relative Reward Strength アルゴリズムを提案した。

* 徳島大学工学部 徳島市南常三島町 2-1

** 大阪教育大学教育学部 柏原市旭ヶ丘 4

* Faculty of Engineering, The University of Tokushima, Tokushima

** Faculty of Education, Osaka Educational University, Kashihara

(Received January 7, 1994)

(Revised August 24, 1994)

(注1) reward が 0 または 1 で与えられる環境を P-モデル環境、 $[0, 1]$ 区間の有限個の有理数で与えられる環境を Q-モデル環境、 $[0, 1]$ 区間内の実数値で与えられる環境を S-モデル環境という。

ところで、ある問題に学習オートマトンを適用するときはその出力となり得るものを学習対象と呼び、また、それらの集合を対象集合と呼ぶことにすると、従来行われてきた研究はいずれも、対象集合の要素が学習開始時にすべて与えられて確定しており、その reward の確率分布のみが未知である場合、すなわち、所定の学習を遂行するためには、学習オートマトンにどのような出力を合計何個もたせればよいかが事前に明確である場合について考察している。しかし、実際の問題に学習オートマトンを適用することを考えるとき、reward の確率分布のみならず対象集合の要素数および内容もが未知であるような周囲環境中において動作する学習オートマトンについても考察することが必要である。

このような環境の例は、移動ロボットの迷路通過問題に見られる¹⁰⁾。この問題においては、迷路は未知の確率で開閉する多数のゲートからなるとし、ロボットの移動ルートを入口から出口に至るゲートの列で表わす。このとき、迷路が未知周囲環境であり、移動ルートが対象集合の要素となる。そして、移動ルート上のゲートの開閉確率の積がそのルートの通過しやすさを表わす確率、すなわち、reward 確率と考えられる。ここで、いま、迷路が大規模なものであるとしよう。このとき、迷路の構造(ゲートの数や位置)を事前に完全に明らかにすることは非常に困難なことであり、また、できたとしても大きな労力を必要とするであろう。そのため、迷路の構造は未知であるとして、迷路の通過ルートの探索を行うのが最良の方法であると考えられる。すなわち、この場合、対象集合の要素数(移動ルートの数)および内容(移動ルート上のゲートの並び)と reward 確率との両者が未知である周囲環境ということになる。このような場合に対しては、最上¹⁰⁾が、P-モデル定常環境中で動作する可変階層構造学習オートマトンを提案し、その学習特性の考察を行った。また、最上、馬場¹¹⁾が、P-モデル非定常複数教師環境に対しても可変階層構造学習オートマトンを適用した。

ところで、一般に学習を行う場合には、学習の結果獲得されるべき目標物(学習目標)が設定される。それらの集合として目標集合を定義すると、対象集合の要素数および内容が未知である場合には、目標集合は対象集合の部分集合となる(このことを迷路通過問題で考えてみると、学習目標は“迷路を通過する移動ルート”であるが、迷路の構造が未知である場合には、壁に行き当たって迷路を通過できない移動ルートや短すぎて迷路を通過できない移動ルートが対象集合には含まれる)。また、環境から成功の返答を最も得やすい学習目標を最良目標と呼ぶ。さらに、いま、何らかの方法で対象集合が事前に確

定できたとすると、全ての学習対象にそれぞれパス^(注2)を割り当てた完全な形の階層構造が構成できる。本論文では、このような階層構造を完全構造と呼ぶ。また、完全構造から1個以上の学習オートマトンが欠けた階層構造を部分構造と呼ぶ。さらに、完全構造上のパスを完全パス、部分構造上のパスで完全パスではないものを部分パス、学習目標を出力とするパスを目標パス、最良目標を出力とするパスを最良目標パスと呼ぶ(目標パスは、全て完全パスである)。そして、つぎの条件が成立するものとする。

(C1) 対象集合は、学習開始時には確定していない。

(C2) 最良目標は唯一に存在し、最良目標パスは完全構造の最適パスである。

(C3) 可変階層構造学習オートマトンは、収束したパスが目標パスであるかどうかは認識できる。

本論文では、条件(C1)~(C3)が成り立つ場合において、対象集合の要素数および内容とそれに対する reward の確率分布との両者が未知であるような S-モデル非定常環境中で動作する可変階層構造学習オートマトンについて考察する。可変階層構造学習オートマトンについては、文献10)および11)において報告されているが、これらは現在の入出力のみに基づいて学習を進めていくものである。ところで、学習オートマトンを実問題に適用するときには、出力数の増加につれてその学習回数が飛躍的に増加するという問題点が存在する。この問題点を克服するために、現在の入力のみならず過去に得た入力をも学習に利用することによって、学習回数の減少を計ることが考えられる。このような新しい可変階層構造学習アルゴリズムが、Simha and Kurose によって提案された Relative Reward Strength アルゴリズムを拡張することによって得られることを示す。さらに、その学習特性についても考察する。

2. s-モデル非定常環境中で動作する可変階層構造学習オートマトン

可変階層構造学習オートマトンの概略を Fig. 1 に示す。まず、その構成であるが、各学習オートマトンはそれぞれ r 個の動作をもち、レベル1には1個の学習オートマトン Z が存在する。そして、その動作 $a_{i1}(i_1=1, \dots, r)$ の下には、それぞれ1個の学習オートマトン Z_{i1} が結合可能であり、よって、レベル2には最大 r 個の学習オートマトンが存在可能である。同様にして、レベル2の各動作 $a_{i_1 i_2}(i_1, i_2=1, \dots, r)$ にも、それぞれ1個の学習

(注2) 階層構造のレベル1の学習オートマトンから最下レベルの学習オートマトンに至る動作の列をパスと呼ぶ。

Fig. 1 Variable hierarchical structure learning automata

オートマトン $Z_{i_1 i_2}$ が結合可能であり、その結果、レベル 3 には最大 r^2 個の学習オートマトンが存在可能となる。このようにして、一般に、レベル s には r^{s-1} 個の学習オートマトンが存在可能であり、レベル s の学習オートマトンを $Z_{i_1 \dots i_{s-1}}$ 、その動作を $\alpha_{i_1 \dots i_{s-1} i_s}$ ($i_s = 1, \dots, r$) と表記する。また、レベル 1 から最下レベル N に至る動作の列 $\{a_{i_1}, a_{i_1 i_2}, \dots, a_{i_1 \dots i_N}\}$ をパスと呼んで $\phi_{i_1 i_2 \dots i_N}$ と記述する。

つぎに、その動作について述べる。(a) 初期構造 Σ^0 として、 n_0 レベル階層構造学習オートマトンを構成し、学習を開始する (Fig. 1 では、 $n_0=1$)。初期時刻 $t=0$ における動作確率は、すべてのレベルの学習オートマトンについて等しいとする (つまり、 $1/r$ となる)。さて、今、レベル 1 の学習オートマトン Z が時刻 t において、動作確率分布 $(p_1(t), \dots, p_r(t))$ に基づいて動作 α_{j_1} を選択したとする (ここで、 $p_i(t)$ ($i=1, \dots, r$) は時刻 t において Z が i 番目の動作 α_i を選択する確率を示す)。すると、 $\phi(t) = \phi_{j_1} (= \alpha_{j_1})$ が環境に出力される。この出力に対して、未知周囲環境は reward $\beta(t) = \beta_{j_1}(t)$ を学習オートマトンに与える。そして、この返答を用いて Z は学習アルゴリ

ムに従って動作確率分布を変更する。このような学習を繰り返した結果、 Σ^0 の最適パスとして ϕ_2 が選択されたとする。(b) ここで、 ϕ_2 が目標パスではないとすると、可変階層構造学習オートマトンは、 α_2 の下に新たに 1 個の学習オートマトン Z_2 を追加して構造 Σ^1 を構成し、学習を継続する。このとき、レベル 2 の学習オートマトン Z_{j_1} (ここでは Z_2) は、レベル 1 で動作 α_{j_1} (ここでは α_2) が選択されたときのみ動作し、 Z と同様、時刻 t における動作確率分布 $(p_{j_1 1}(t), \dots, p_{j_1 r}(t))$ に基づいて動作 $\alpha_{j_1 j_2}$ を選択する。ここで、レベル 1 から最下レベルに至る動作の列 $\{a_{j_1}, a_{j_1 j_2}\}$ 、すなわち、パス $\phi_{j_1 j_2}$ が選択されたとする。すると、このパスに対応する出力が環境に与えられ、それに対して、環境は reward $\beta(t) = \beta_{j_1 j_2}(t)$ を学習オートマトンに与える。そして、この返答に基づいて学習オートマトンはその動作確率分布を変更する。ここでは、学習の結果、 Σ^1 の最適パスとして ϕ_{21} が選択されたとする。(c) ϕ_{21} も目標パスではないとすると、可変階層構造学習オートマトンは α_{21} の下に新たに 1 個の学習オートマトン Z_{21} を追加して構造 Σ^2 を構成し、学習を継続する。

以上のように学習が進められて行き、最適パスとして目標パスが得られると、可変階層構造学習オートマトンは学習を終了する。ここで、初期構造から m 回の変更を経た構造を Σ^m ($m=0, 1, \dots$) と記述し、構造 Σ^m のパスの総数を Ψ^m 、Fig. 1 のように階層構造を図示したときに左から数えて k 番目のパスを $\phi_{i_1 i_2 \dots i_{N_k m}}$ ($k=1, \dots, \Psi^m$) と記述する (ただし、 N_k^m はこのパスのレベル数である)。そして、これに対応するパス確率は $\pi_{i_1 i_2 \dots i_{N_k m}}(t) = p_{i_1}(t) \times p_{i_1 i_2}(t) \times \dots \times p_{i_1 i_2 \dots i_{N_k m}}(t)$ である。また、構造 Σ^m 上のパス確率の最大値を Σ^m 上の最大パス確率と呼んで $\bar{\pi}_{i_1 i_2 \dots i_{N_k m}}(t)$ と記述し、これに対応するパスを Σ^m 上の最大パス $\bar{\phi}_{i_1 i_2 \dots i_{N_k m}}$ と呼ぶ。

本論文では、対象集合の要素数および内容と reward の確率分布との両者が未知であるような S-モデル非定常環境について考える。S-モデル非定常環境とは、可変階層構造学習オートマトンの出力 $\phi(t) = \phi_{i_1 i_2 \dots i_{N_k m}}$ に対する reward $\beta_{i_1 i_2 \dots i_{N_k m}}(t)$ が、平均 $E\{\beta_{i_1 i_2 \dots i_{N_k m}}(t)\}$ をもつ分布関数 $F_{i_1 i_2 \dots i_{N_k m}}(t)$ に従って与えられるような環境である (ただし、reward の値は閉区間 $[0, 1]$ 内の実数値)。そして、構造 Σ^m において、つぎの関係を満たすパスを構造 Σ^m における最適パス $\phi_{j_1^* j_2^* \dots j_{N_k m}^*}$ と呼び、唯一に存在するものとする：

$$\min_t E\{\beta_{j_1^* j_2^* \dots j_{N_k m}^*}(t)\} > \max_t E\{\beta_{i_1 i_2 \dots i_{N_k m}}(t)\} \quad (1)$$

ただし、 $k=1, \dots, \phi^m$ ($k \neq a$) である。

ここで、完全構造を Σ^p とし最良目標パスを $\phi_{h_1^* h_2^* \dots h_{N_k p}^*}$

とすると,

$$\min_t E\{\beta_{k_1 k_2 \dots k_{N_k}}(t)\} > \max_t E\{\beta_{i_1 i_2 \dots i_{N_k}}(t)\} \quad (2)$$

$$k=1, \dots, \psi^P (k \neq a)$$

が成り立つ。また、可変階層構造学習オートマトンの学習目的は、最良目標パスを選択する確率をできるだけ大きくすることである。

3. 可変階層構造学習アルゴリズムとその学習特性

3.1 可変階層構造学習アルゴリズム

いま、構造 Σ^m において、任意のパス $\phi_{i_1 \dots i_{N_k}} (k=1, \dots, \Psi^m)$ が時刻 t までに得た reward の最新の値を $\bar{\beta}_{i_1 \dots i_{N_k}}(t)$ とする(ここでは、これをパス最新 reward と呼ぶことにする)。すなわち、

$$\bar{\beta}_{i_1 \dots i_{N_k}}(t) = \beta_{i_1 \dots i_{N_k}}(\tau_{i_1 \dots i_{N_k}}(t)) \quad (3)$$

ただし、 $\tau_{i_1 \dots i_{N_k}}(t)$ は、時刻 t までにパス $\phi_{i_1 \dots i_{N_k}}$ が選択された最新の時刻である ($\tau_{i_1 \dots i_{N_k}} \leq t$)。すると、

$$E\{\bar{\beta}_{j_1^* \dots j_{N_k}^*}(t)\} = E\{\beta_{j_1^* \dots j_{N_k}^*}(\tau_{j_1^* \dots j_{N_k}^*}(t))\} \quad (4)$$

$$E\{\bar{\beta}_{i_1^* \dots i_{N_k}^*}(t)\} = E\{\beta_{i_1^* \dots i_{N_k}^*}(\tau_{i_1^* \dots i_{N_k}^*}(t))\} \quad (k \neq a) \quad (5)$$

であることと(1)式とから、次式が成立する。

$$E\{\bar{\beta}_{j_1^* \dots j_{N_k}^*}(t)\} > E\{\bar{\beta}_{i_1^* \dots i_{N_k}^*}(t)\} \text{ for } \forall t \quad (6)$$

ただし、 $k=1, 2, \dots, \Psi^m$, かつ、 $k \neq a$ 。

このとき、つぎのことに注意しよう。すなわち、Relative Reward Strength アルゴリズムにおいては、学習オートマトンの各動作にそれぞれ最新 reward が対応していることが必要であるので、それを可変階層構造学習オートマトンに拡張する場合にも、階層構造上の全ての動作にそれぞれ最新 reward が対応していることが要求される。しかし、本論文において考察する可変階層構造学習オートマトンにおいて与えられるのは、パス最新 reward のみである。そこで、本論文では、このパス最新 reward を用いて、そのパス上の各動作に対応する最新 reward v をつぎのように構成する。

(i) 最下レベル N_k^m の動作に対する最新 reward :

$$v_{i_1 \dots i_{N_k}}(t) = \bar{\beta}_{i_1 \dots i_{N_k}}(t) \quad (7)$$

(ii) レベル s の動作に対する最新 reward

$$(1 \leq s \leq N_k^m - 1):$$

$$v_{i_1 \dots i_s}(t) = \max \{v_{i_1 \dots i_{s-1} 1}(t), \dots, v_{i_1 \dots i_{s-1} r}(t)\} \quad (8)$$

また、動作確率に対してつぎの特性が成り立つものとする。

(P1) すべての時刻 t およびすべての $i_1, i_2, \dots, i_s (i_q = 1, 2, \dots, r; q=1, 2, \dots, s (s=1, 2, \dots, N_k^m))$ において、動作確率 $p_{i_1 i_2 \dots i_s}(t)$ はつぎの不等式を満たす。

$$q_{\min} \leq p_{i_1 i_2 \dots i_s}(t) \leq q_{\max} \quad (9)$$

ただし、 $q_{\min}$ および $q_{\max}$ は、 $0 < q_{\min} < q_{\max} < 1$ かつ $q_{\max}$

$= 1 - (r-1)q_{\min}$ を満たす実数である。

このとき、Relative Reward Strength アルゴリズムを拡張した新しい可変階層構造学習アルゴリズムはつぎのように構成される。

《可変階層構造学習アルゴリズム》

Step 1. 初期時刻 $t=0$ として、初期構造 Σ^0 (n_0 レベル階層構造学習オートマトン)を構成する ($m=0$, $\Psi^0 = r^{n_0}$, $p_{i_1 i_2 \dots i_s}(0) = 1/r (s=1, \dots, n_0)$)。また、収束判定値 $q_{\max}$ ($q_{\max} > 1/r$ を満たし、かつ、1に近い小数)を与える。

Step 2. 最大パス $\bar{\phi}_{i_1 i_2 \dots i_{N_k}}(t)$ 上の各動作の動作確率を $\bar{p}_{i_1 i_2 \dots i_s}(t) (s=1, \dots, N_k^m)$ とする。

(i) $\bar{p}_{i_1 i_2 \dots i_s}(t) \geq q_{\max}$ for $\forall s$ のとき：最大パスが目標パスであるならば、本アルゴリズムを終了する。それ以外のときは、Step 4 へ。

(ii) (i) 以外のとき：Step 3 へ。

Step 3. 構造 Σ^m のもとで、下記の強化法を用いて学習を行う。

[強化法] 時刻 t において、パス $\phi_{j_1 j_2 \dots j_{N_k}} (= \{\alpha_{j_1}, \alpha_{j_2}, \dots, \alpha_{j_{N_k}}\})$ が選択され、これに対して未知周囲環境から $\beta(t)$ なる reward が与えられたとする。

(i) $\beta(t)$ を用いてパス $\phi_{j_1 j_2 \dots j_{N_k}}$ 上の各動作の最新 reward を更新する。

(ii) パス $\phi_{j_1 j_2 \dots j_{N_k}}$ 上の各学習オートマトン $Z_{j_1 \dots j_{s-1}} (s=1, 2, \dots, N_k^m)$ の動作確率分布を次式によって変更する。

$$p_{j_1 \dots j_{s-1} i_s}(t+1) = p_{j_1 \dots j_{s-1} i_s}(t) + \lambda_{j_1 \dots j_{s-1}}(t) \Delta p_{j_1 \dots j_{s-1} i_s}(t) \quad (10)$$

このとき、 $\Delta p_{j_1 \dots j_{s-1} i_s}(t)$ はつぎのように計算される。

$$\Delta p_{j_1 \dots j_{s-1} i_s}(t) = \begin{cases} \frac{v_{j_1 \dots j_{s-1} i_s}(t) - \frac{1}{|A(t)_{j_1 \dots j_{s-1}}|}}{\sum_{i_s \in A_{j_1 \dots j_{s-1}}(t)} v_{j_1 \dots j_{s-1} i_s}(t)} & \text{for } i_s \in A_{j_1 \dots j_{s-1}}(t) \\ 0 & \text{for } i_s \notin A_{j_1 \dots j_{s-1}}(t) \end{cases} \quad (11)$$

ただし、 $\lambda_{j_1 \dots j_{s-1}}(t)$ はステップサイズパラメータである。また、選ばれたパス上にない学習オートマトンの動作確率分布はまったく変更されない。

(iii) $t=t+1$ として、Step 2 へ。

Step 4. Σ^m 上の最大パスの下に学習オートマトンを1個追加することによって、構造を Σ^m から Σ^{m+1} に更新する。

Step 5. 新構造 Σ^{m+1} 上の全てのパスのパス確率を $1/\Psi^{m+1}$ と置く。

Step 6. パス確率をもとに、パス上の各学習オートマトンの動作確率を計算し、この値を新構造での初期動作確率とする。 $m=m+1$ とし、Step 2 へ。

ここで、強化法に含まれる集合 $A_{j_1 \dots j_{s-1}}(t)$ は、つぎのよう
に与えられる。

$A_{j_1 \dots j_{s-1}}(t)$ の計算

a) 学習オートマトン $Z_{j_1 \dots j_{s-1}}$ において、 i_s に関して、 $v_{j_1 \dots j_{s-1} i_s}(t)$ を降順に並べる。

b) $D_{j_1 \dots j_{s-1}} = \{k_s | v_{j_1 \dots j_{s-1} k_s}(t) = \max_{i_s} \{v_{j_1 \dots j_{s-1} i_s}(t)\}\}$ と
おく。

c) $v_{j_1 \dots j_{s-1} i_s}(t)$ の大きい順に以下を繰り返す(ただし、 $i_s \notin D_{j_1 \dots j_{s-1}}$) : $A_{j_1 \dots j_{s-1}}(t) = D_{j_1 \dots j_{s-1}} \cup \{i_s\}$ として (10)、
(11) 式を実行したとき、 $p_{j_1 \dots j_{s-1} i_s}(t+1) \geq q_{\min}$ となるなら
ば、 $D_{j_1 \dots j_{s-1}} = D_{j_1 \dots j_{s-1}} \cup \{i_s\}$ とおく。

d) すべての $i_s (=1, \dots, r)$ に対して c) が終了した
ら、 $A_{j_1 \dots j_{s-1}}(t) = D_{j_1 \dots j_{s-1}}$ とする。

(注意) 集合 $A_{j_1 \dots j_{s-1}}(t)$ は、学習オートマトン $Z_{j_1 \dots j_{s-1}}$ において (10)、(11) 式を実行したときに、時刻 $t+1$ の動作確率 $p_{j_1 \dots j_{s-1} i_s}(t+1)$ が特性 (P1) を満たす場合の番号 $\{i_s\}$ の集合である。そして、集合 $A_{j_1 \dots j_{s-1}}(t)$ に含まれる番号に対応する動作に対しては特性 (P1) を満たすように動作確率の更新が行われ、集合 $A_{j_1 \dots j_{s-1}}(t)$ に含まれない番号に対応する動作については更新は行われない。このことから、初期動作確率を特性 (P1) を満たすように取ると、本学習アルゴリズムによる動作確率の更新は、常に特性 (P1) を満たすように行われるということになる。ここで、初期動作確率 $1/r$ と $q_{\max}$ とが $q_{\max} > 1/r$ なる関係を満たすことに注意すると、 $q_{\max} = 1 - (r-1)q_{\min} > 1/r$ より $1/r > q_{\min}$ が成り立つため、初期動作確率は特性 (P1) を満たすことになる。以上のことから、本学習アルゴリズムは常に特性 (P1) を満たすことになる。

3.2 学習特性

構造 Σ^m において、最適パス $\phi_{j_1^* j_2^* \dots j_{N^m}^*}$ 上の各学習オートマトン $Z_{j_1^* \dots j_{s-1}^*} (s=1, \dots, N^m)$ における最新 reward vector $v_{j_1^* \dots j_{s-1}^* i_s}(t) = (v_{j_1^* \dots j_{s-1}^* i_s}(t), \dots, v_{j_1^* \dots j_{s-1}^* r}(t))'$ に着目すると、つぎの定理を得る。

《定理1》 任意の時刻 t における最新 reward が (7)、(8) 式で与えられるものとする。すると、構造 Σ^m の最適パス上の各学習オートマトン $Z_{j_1^* \dots j_{s-1}^*} (s=1, \dots, N^m)$ において、つぎの関係式が成立する。

$$E\{v_{j_1^* \dots j_{s-1}^* i_s}(t)\} > E\{v_{j_1^* \dots j_{s-1}^* i_s}(t)\} \text{ for } \forall t \quad (12)$$

ただし、 $i_s = 1, \dots, r$, $i_s \neq j_s^*$ 。

(証明) 付録に示す。

定理1に注意すると、つぎの定理が成立する。

《定理2》 構造 Σ^m において、定理1の条件が成り立つものとする。また、各学習オートマトン $Z_{i_1 \dots i_{s-1}} (s=1, \dots, N^m, k=1, \dots, \Psi^m)$ において、ステップサイズパラメータ $\lambda_{i_1 \dots i_{s-1}}(t)$ を

$$\lambda_{i_1 \dots i_{s-1}}(t) > 0, \quad \sum_{t=1}^{\infty} \lambda_{i_1 \dots i_{s-1}}(t) = \infty,$$

$$\sum_{t=1}^{\infty} \lambda_{i_1 \dots i_{s-1}}^2(t) < \infty \quad (13)$$

を満たす実数列とする。このとき、可変階層構造学習アルゴリズムのもとで、構造 Σ^m の最適パス $\phi_{j_1^* j_2^* \dots j_{N^m}^*}$ に対して次式が成立する。

$$\lim_{t \rightarrow \infty} \pi_{j_1^* j_2^* \dots j_{N^m}^*}(t) = \bar{q} \quad \text{w. p. 1} \quad (14)$$

ただし、 $\bar{q} = (q_{\max})^{N^m}$ 。

(証明) 定理1より、最適パス上の各学習オートマトン $Z_{j_1^* \dots j_{s-1}^*} (s=1, \dots, N^m)$ において、最適パス上の動作 $\alpha_{j_1^* \dots j_{s-1}^* j_s^*}$ に対する $E\{v_{j_1^* \dots j_{s-1}^* j_s^*}(t)\}$ の値がそのほかの動作 $\alpha_{j_1^* \dots j_{s-1}^* i_s} (i_s \neq j_s^*)$ に対する $E\{v_{j_1^* \dots j_{s-1}^* i_s}(t)\}$ の値よりも常に大きいことがいえる。すなわち、文献9) Theorem 1 の reward の平均値に対する条件が成立することになる。このことから、最適パス上の各確率オートマトン $Z_{j_1^* \dots j_{s-1}^*} (s=1, \dots, N^m)$ において、

$$\lim_{t \rightarrow \infty} p_{j_1^* \dots j_{s-1}^* j_s^*}(t) = q_{\max} \quad \text{w. p. 1} \quad (15)$$

が成り立つことが、文献9) の Theorem 1 と同様にして示せる。よって、パス確率の定義に注意すると、(15) 式より次式が成立する。

$$\begin{aligned} \lim_{t \rightarrow \infty} \pi_{j_1^* j_2^* \dots j_{N^m}^*}(t) &= \lim_{t \rightarrow \infty} \{p_{j_1^*}(t) p_{j_2^*}(t) \dots p_{j_{N^m}^*}(t)\} \\ &= \lim_{t \rightarrow \infty} p_{j_1^*}(t) \lim_{t \rightarrow \infty} p_{j_2^*}(t) \\ &\quad \dots \lim_{t \rightarrow \infty} p_{j_{N^m}^*}(t) \\ &= q_{\max} q_{\max} \dots q_{\max} \\ &= (q_{\max})^{N^m} \\ &= \bar{q} \quad \text{w. p. 1} \end{aligned} \quad (16)$$

Q. E. D.

また、任意の部分構造 Σ^m における最適パスに関して、つぎの定理が成立する。

《定理3》 構造 Σ^m のパス $\phi_{i_1 \dots i_{N^m}}$ に学習オートマトンを追加するとする。このとき、つぎの条件が成り立つものとする。

$$E\{\beta_{i_1 \dots i_{N^m}}(t)\} \geq E\{\beta_{i_1 \dots i_{N^m} \min_{s=1}^{N^m}}(t)\} \text{ for } \forall t \quad (17)$$

このとき、任意の部分構造の最適パスは、部分パスかあるいは最良目標パスのいずれかである。

(証明) 部分構造 Σ^m 上に存在するパスは、部分パス、目標パス(完全パス)、および、目標パス以外の完全パスのいずれかに分類される。そこで、個々のパスについて、それが部分構造 Σ^m の最適パスになるか否かについて考察する。(i) まず、部分パスが最適パスとなり得ることは明らかである。(ii) つぎに、目標パスが最適パスとなる場合について考える。いま、最良目標パスは $\phi_{j_1^* j_2^* \dots j_{N^m}^*}$ であり、また、構造 Σ^m の最適パスが目標パス $\phi_{j_1^* j_2^* \dots j_{N^m}^*}$ であるとすると、 $\phi_{j_1^* j_2^* \dots j_{N^m}^*} \neq \phi_{j_1^* j_2^* \dots j_{N^m}^*}$ であると仮定す

る.すると,(1)式より,次式が成立する.すなわち,

$$\min_t E\{\beta_{j_1^*j_2^*\dots j_{N_m}^*}(t)\} > \max_t E\{\beta_{j_1^*j_2^*\dots j_{N_m}^*}(t)\} \quad (18)$$

ただし, N_m^* は最良目標パスの構造 Σ^m におけるレベル数である ($n_0 \leq N_m^* \leq N$). ところで, (17)式より,

$$\max_t E\{\beta_{j_1^*j_2^*\dots j_{N_m}^*}(t)\} \geq \max_t E\{\beta_{j_1^*j_2^*\dots j_{N_m}^*}(t)\} \quad (19)$$

であることに注意すると, (18), (19)式より,

$$\min_t E\{\beta_{j_1^*j_2^*\dots j_{N_m}^*}(t)\} > \max_t E\{\beta_{j_1^*j_2^*\dots j_{N_m}^*}(t)\} \quad (20)$$

となる.すなわち,

$$\max_t E\{\beta_{j_1^*j_2^*\dots j_{N_m}^*}(t)\} > \min_t E\{\beta_{j_1^*j_2^*\dots j_{N_m}^*}(t)\} \quad (21)$$

となる.これは, $\phi_{j_1^*j_2^*\dots j_{N_m}^*}$ が最良目標パスであることに反する.よって, $\phi_{j_1^*j_2^*\dots j_{N_m}^*} = \phi_{j_1^*j_2^*\dots j_{N_m}^*}$ となる.すなわち,目標パスの中では最良目標パスのみが最適パスとなり得る.(iii)最後に,目標パス以外の完全パスが最適パスになる場合について考える.いま,目標パス以外の完全パス $\phi_{j_1^*j_2^*\dots j_{N_m}^*}$ が最適パスであると仮定する.すると,(ii)の場合と同様にして,(1)式から(18)式を,また(17)式から(19)式を得ることから,(21)式の関係が得られる.これは,条件(C2)に反する.ゆえに,目標パス以外の完全パスは最適パスにはなり得ない.

よって,(i),(ii),(iii)より,任意の構造において最適パスとなるのは,部分パスか,あるいは,最良目標パスのいずれかとなる. Q.E.D.

定理2と定理3とから,つぎの定理が得られる.

《定理4》 定理2および定理3の条件が成り立つものとする.このとき,可変階層構造学習アルゴリズムによって最良目標パスが選択される確率は,いくらかでも1に近くすることができる.

(証明) 定理2および定理3より,可変階層構造学習オートマトンによる最良目標パスの選択は,構造 Σ^m の最適パスとしての部分パスの選択と構造 Σ^{m+1} への構造更新を繰り返しながら進められ,ついには,ある構造 Σ^M の最適パスとして最良目標パスを選択することによって完了する.そして,このとき,次式が成立する.

$$\lim_{t \rightarrow \infty} \pi_{j_1^*j_2^*\dots j_{N_m}^*}(t) = \bar{q} \quad \text{w. p. 1}, \quad (22)$$

ここで, $q_{\max}$ がいくらかでも1に近い値にとれることから $\bar{q}$ もいくらかでも1に近い値にとれることに注意すると,可変階層構造学習アルゴリズムによって,最適パスのパス確率はいくらかでも1に近い値にすることができる.

Q.E.D.

4. 数値シミュレーション

各学習オートマトンがそれぞれ2個($r=2$)の動作をもち,完全構造の最大レベル数が11である場合に本学習アルゴリズムを適用し,その学習の様子を見る(最適パス

は $\phi_{11111111222}$).

いま,構造 Σ^m の時刻 t における出力 $\phi(t) = \phi_{i_1 \dots i_{N_m}}$ に対応する非定常環境からの reward $\beta_{i_1 \dots i_{N_m}}(t)$ を次式で与えることにする:

$$\beta_{i_1 \dots i_{N_m}}(t) = \gamma_{i_1 \dots i_{N_m}}(t) + e_{i_1 \dots i_{N_m}} u \quad (23)$$

ここで, u は $[-0.5, 0.5]$ の一様乱数であり, また,

$$\gamma_{i_1 \dots i_{N_m}}(t) = a_{i_1 \dots i_{N_m}} + b_{i_1 \dots i_{N_m}} \sin(c_{i_1 \dots i_{N_m}} \pi t + d_{i_1 \dots i_{N_m}}) \quad (24)$$

である.ただし, $a_{i_1 \dots i_{N_m}}, b_{i_1 \dots i_{N_m}}, c_{i_1 \dots i_{N_m}}, d_{i_1 \dots i_{N_m}}$ および $e_{i_1 \dots i_{N_m}}$ は正の実数であり,(1)式およびあらゆる時刻 t において $0 \leq \beta_{i_1 \dots i_{N_m}}(t) \leq 1$ が成り立つように与える.

この問題に対して本可変階層構造学習アルゴリズムを適用するわけであるが,本シミュレーションでは $q_{\min} = 0.005$ ($q_{\max} = 0.995$) とし,また,ステップサイズパラメータ $\lambda_{i_1 \dots i_{s-1}} (s=1, \dots, N_m^*)$ は各レベルで同一の値として, 0.15 ($t \leq 30000$), $1/t$ ($t > 30000$) と与えた.そして,30回の実験を行いその学習回数の平均を Table 1 に示す(30回とも最良目標パスに収束した場合の平均である).ここでは比較のために,本学習アルゴリズムの代わりに,最上¹⁾によって提案された L_{R-1} タイプの可変階層構造学習アルゴリズムを動作させた場合の結果もあわせて示している(完全構造および非定常環境の特性は同一である).Table 1 より,本可変階層構造学習アルゴリズムが, L_{R-1} タイプの可変階層構造学習アルゴリズムに比べて約21.6倍の速さで収束することがわかる.また,ステップサイズパラメータと学習性能との関係を見るために,ステップサイズパラメータを種々変更した場合の平均学習回数と正解率(30回の実験中最良目標パスに収束した割合)を Table 2 に示す.これより明らかのように,ステップサイズパラメータを大きく取ると学習回数は少なくなるが正解率が下がり,小さく取ると学習回数は多くなるが正解率が100%となる.このことから,最適なステップサイズパラメータが存在すると考えられるが,その決定法の確立は今後の課題である(Table 1 の Our algorithm に対する Average number 21019 は, Table 2 から明らかのように,ステップサイズパラメータを0.01刻みで大きくしていったときに正解率100%が途切れる

Table 1 Simulation Results

Algorithm	Average number of iterations	Ratio
Our algorithm	21019	1
L_{R-1} algorithm	453475	21.6

Table 2 Learning Performance for Various Step Size Parameters

Step size parameter	Average number of iterations	Percentage of correct convergences
0.10	28077	100 %
0.11	24858	100 %
0.12	22458	100 %
0.13	23007	100 %
0.14	21351	100 %
0.15	21019	100 %
0.16	18104	96.7 %
0.17	19564	100 %
0.18	17593	93.3 %
0.19	16199	86.7 %
0.20	16075	96.7 %

直前のものである)。

5. む す び

本論文では、対象集合の要素数および内容とそれに対する reward の確率分布との両者が未知であるような S-モデル非定常環境中で動作する可変階層構造学習オートマトンについて考察し、Simha and Kurose によって提案された Relative Reward Strength Algorithm を拡張することによって、新しい可変階層構造学習アルゴリズムを構築した。そして、各レベルの学習オートマトンに対する最新 reward の値が(7)、(8)式のように与えられるならば、条件(C1)～(C3)のもとで、ステップサイズパラメータを適切にとると、本可変階層構造学習アルゴリズムによって、最良目標パスが選択される確率をいくらでも1に近い値にすることができることを示した。

また、数値シミュレーションにおいては、各学習オートマトンがそれぞれ2個の動作をもち、完全構造の最大レベル数が11である場合に、本学習アルゴリズムと L_{R-1} タイプの可変階層構造学習アルゴリズムの両者を適用することによって、本可変階層構造学習アルゴリズムが、 L_{R-1} タイプの可変階層構造学習アルゴリズムに比べて少ない学習回数で収束することを確認した。これは、 L_{R-1} タイプの可変階層構造学習アルゴリズムが現在の入出力のみによって学習を進めるのに対して、本学習アルゴリズムでは過去に得た入力をも学習に用いることによると考えられる。

今後は、本学習アルゴリズムの種々の問題への適用を計ることが望まれる。

参 考 文 献

- 1) M. L. Testlin: On the Behavior of Finite Automata in Random Media, *Avtomatika i Telemekhanika*, **22**-10, 1345/1354 (1961)
- 2) S. Lakshmivarahan: Learning Algorithms Theory

and Applications, Springer-Verlag (1981)

- 3) N. Baba: New Topics in Learning Automata Theory and Applications, Lecture Note in Control and Information Sciences, Springer-Verlag (1985)
- 4) K. S. Narendra and M. A. L. Thathachar: Learning Automata An Introduction, Prentice-Hall, Inc., (1989)
- 5) M. A. L. Thathachar and K. R. Ramakrishnan: A Hierarchical System of Learning Automata. *IEEE Trans.*, **SMC-11**-3, 236/241 (1981)
- 6) M. A. L. Thathachar and P. S. Sastry: A Class of Rapidly Converging Algorithms for Learning Automata. *IEEE International Conference on Cybernetics and Society*, 602/606 (1984)
- 7) M. A. L. Thathachar and P. S. Sastry: A New Approach to the Design of Reinforcement Schemes for Learning Automata. *IEEE Trans.*, **SMC-15**-1, 168/175 (1985)
- 8) 馬場, 最上, 三宅: 非定常複数教師環境における階層構造学習オートマトンの学習特性. *電子通信学会論文誌*, **J69-D-11**, 1724/1731 (1986)
- 9) Rahul Simha and James F. Kurose: Relative Reward Strength Algorithms for Learning Automata. *IEEE Trans.*, **SMC-19**-2, 388/398 (1989)
- 10) 最上義夫: 一人教師定常環境中で動作する可変階層構造学習オートマトン, *計測自動制御学会論文集*, **25**-9, 1108/1025 (1989)
- 11) 最上, 馬場: 非定常複数教師環境中において動作する可変階層構造学習オートマトン, *電子情報通信学会論文誌*, **J74-D-1**-7, 401/408 (1991)

《付録 定理1の証明》

最適パス $\phi_{j_1^* \dots j_{R_m}^*}$ 上の学習オートマトンに着目する。まず、最下レベル N_a^m の学習オートマトン $Z_{j_1^* \dots j_{R_m-1}^*}$ においては、(7)式より、その動作 $a_{j_1^* \dots j_{R_m-1}^* l_1}(l_1=1, \dots, r)$ に対する最新 reward $v_{j_1^* \dots j_{R_m-1}^* l_1}(t)$ として、その動作を通るパス $\phi_{j_1^* \dots j_{R_m-1}^* l_1}$ が時刻 t までに得たパス最新 reward が格納されるため、

$$E\{v_{j_1^* \dots j_{R_m-1}^* l_1}(t)\} = E\{\bar{\beta}_{j_1^* \dots j_{R_m-1}^* l_1}(t)\} \quad (A.1)$$

$$(l_1=1, \dots, r)$$

が成立する。よって、(6)、(A.1)式に注意すると、つぎの関係が成立する。

$$\begin{aligned} E\{v_{j_1^* \dots j_{R_m-1}^* j_{R_m}^*}(t)\} &= E\{\bar{\beta}_{j_1^* \dots j_{R_m-1}^* j_{R_m}^*}(t)\} \\ &> E\{\bar{\beta}_{j_1^* \dots j_{R_m-1}^* l_1}(t)\} \\ &= E\{v_{j_1^* \dots j_{R_m-1}^* l_1}(t)\} \quad (A.2) \\ (l_1=1, \dots, r; l_1 \neq j_{R_m}^*) \end{aligned}$$

つぎに、学習オートマトン $Z_{j_1^* \dots j_{R_m-1}^*}$ が結合されている動作 $a_{j_1^* \dots j_{R_m-1}^*}$ をもつレベル (N_a^m-1) の学習オートマトン $Z_{j_1^* \dots j_{R_m-2}^*}$ について考える。(7)、(8)、(A.2)式より、動作 $a_{j_1^* \dots j_{R_m-1}^*}$ に対して、つぎの関係が成立する。

$$\begin{aligned} E\{v_{j_1^* \dots j_{R_m-2}^* j_{R_m-1}^*}(t)\} &= \max \{E\{v_{j_1^* \dots j_{R_m-1}^* l_1}(t)\}, \\ &\quad \dots, E\{v_{j_1^* \dots j_{R_m-1}^* l_r}(t)\}\} \\ &= E\{v_{j_1^* \dots j_{R_m-1}^* j_{R_m}^*}(t)\} \\ &= E\{\bar{\beta}_{j_1^* \dots j_{R_m-2}^* j_{R_m-1}^* j_{R_m}^*}(t)\} \quad (A.3) \end{aligned}$$

また、 $Z_{j_1^* \dots j_{R_m-2}^*}$ の $a_{j_1^* \dots j_{R_m-1}^*}$ 以外の動作 $a_{j_1^* \dots j_{R_m-2}^* l_2}(l_2=1, \dots,$

$r; l_2 \neq j_{a^{m-1}}^*$ に対しても, (7), (8)式より, 次式が成立する.

$\alpha_{j_1^* \dots j_{a^{m-2}}^* l_2}$ が最下レベルの動作のとき:

$$E\{v_{j_1^* \dots j_{a^{m-2}}^* l_2}(t)\} = E\{\bar{\beta}_{j_1^* \dots j_{a^{m-2}}^* l_2}(t)\} \quad (A.4)$$

$\alpha_{j_1^* \dots j_{a^{m-2}}^* l_2}$ が最下レベルの動作でないとき:

$$\begin{aligned} E\{v_{j_1^* \dots j_{a^{m-2}}^* j_{a^{m-1}}^* l_2}(t)\} &= \max \{E\{v_{j_1^* \dots j_{a^{m-2}}^* j_{a^{m-1}}^* l_2 1}(t)\}, \\ &\quad \dots, E\{v_{j_1^* \dots j_{a^{m-2}}^* j_{a^{m-1}}^* l_2 r}(t)\}\} \\ &= \max \{E\{\bar{\beta}_{j_1^* \dots j_{a^{m-2}}^* j_{a^{m-1}}^* l_2 1}(t)\}, \\ &\quad \dots, E\{\bar{\beta}_{j_1^* \dots j_{a^{m-2}}^* j_{a^{m-1}}^* l_2 r}(t)\}\} \end{aligned} \quad (A.5)$$

よって, (6)式に注意すると, (A.3), (A.4), (A.5)式より, レベル $(N_a^m - 1)$ の学習オートマトン $Z_{j_1^* \dots j_{a^{m-2}}^*}$ についても, つぎの関係が成立する.

$$E\{v_{j_1^* \dots j_{a^{m-2}}^* j_{a^{m-1}}^* l_2}(t)\} > E\{v_{j_1^* \dots j_{a^{m-2}}^* l_2}(t)\} \quad (A.6)$$

$$(l_2 = 1, \dots, r; l_2 \neq j_{a^{m-1}}^*)$$

さらに, 学習オートマトン $Z_{j_1^* \dots j_{a^{m-2}}^*}$ が結合されている動作 $\alpha_{j_1^* \dots j_{a^{m-2}}^*}$ をもつレベル $(N_a^m - 2)$ の学習オートマトン $Z_{j_1^* \dots j_{a^{m-3}}^*}$ について考える. (8), (A.3), (A.6)式より, 動作 $\alpha_{j_1^* \dots j_{a^{m-2}}^*}$ に対して, つぎの関係が成立する.

$$\begin{aligned} E\{v_{j_1^* \dots j_{a^{m-2}}^*}(t)\} &= \max \{E\{v_{j_1^* \dots j_{a^{m-2}}^* l_2}(t)\}, \dots, \\ &\quad E\{v_{j_1^* \dots j_{a^{m-2}}^* r}(t)\}\} \\ &= E\{v_{j_1^* \dots j_{a^{m-2}}^*}(t)\} \\ &= E\{\bar{\beta}_{j_1^* \dots j_{a^{m-2}}^*}(t)\} \end{aligned} \quad (A.7)$$

また, $Z_{j_1^* \dots j_{a^{m-3}}^*}$ の $\alpha_{j_1^* \dots j_{a^{m-2}}^*}$ 以外の動作 $\alpha_{j_1^* \dots j_{a^{m-3}}^* l_3} (l_3 = 1, \dots, r; l_3 \neq j_{a^{m-2}}^*)$ に対しても, (7), (8)式より, 次式が成立する.

$\alpha_{j_1^* \dots j_{a^{m-3}}^* l_3}$ が最下レベルの動作のとき:

$$E\{v_{j_1^* \dots j_{a^{m-3}}^* l_3}(t)\} = E\{\bar{\beta}_{j_1^* \dots j_{a^{m-3}}^* l_3}(t)\} \quad (A.8)$$

$\alpha_{j_1^* \dots j_{a^{m-3}}^* l_3}$ が最下レベルの動作でないとき:

$$\begin{aligned} E\{v_{j_1^* \dots j_{a^{m-3}}^* l_3}(t)\} &= \max \{E\{v_{j_1^* \dots j_{a^{m-3}}^* l_3 1}(t)\}, \\ &\quad \dots, E\{v_{j_1^* \dots j_{a^{m-3}}^* l_3 r}(t)\}\} \end{aligned} \quad (A.9)$$

ここで, (A.9)式の右辺において, $E\{v_{j_1^* \dots j_{a^{m-3}}^* l_3 i}(t)\}$ が最下レベルの動作に対するものであるならば,

$$E\{v_{j_1^* \dots j_{a^{m-3}}^* l_3 i}(t)\} = E\{\bar{\beta}_{j_1^* \dots j_{a^{m-3}}^* l_3 i}(t)\} \quad (A.10)$$

となり, また, 最下レベルでない動作に対するものであるならば, (A.5)式より,

$$\begin{aligned} E\{v_{j_1^* \dots j_{a^{m-3}}^* l_3 i}(t)\} &= \max \{E\{\bar{\beta}_{j_1^* \dots j_{a^{m-3}}^* l_3 i 1}(t)\}, \\ &\quad \dots, E\{\bar{\beta}_{j_1^* \dots j_{a^{m-3}}^* l_3 i r}(t)\}\} \end{aligned} \quad (A.11)$$

となる. よって, (6)式に注意すると, (A.7)~(A.11)式より, レベル $(N_a^m - 2)$ の学習オートマトン $Z_{j_1^* \dots j_{a^{m-3}}^*}$ に

ついても, つぎの関係が成立する.

$$E\{v_{j_1^* \dots j_{a^{m-3}}^* j_{a^{m-2}}^* l_3}(t)\} > E\{v_{j_1^* \dots j_{a^{m-3}}^* l_3}(t)\} \quad (A.12)$$

$$(l_3 = 1, \dots, r; l_3 \neq j_{a^{m-2}}^*)$$

以下同様に, 一般にレベル s の学習オートマトン $Z_{j_1^* \dots j_{a^{s-1}}^*}$ において, 次式が成立する.

$$E\{v_{j_1^* \dots j_{a^{s-1}}^* j_{a^s}^* l_s}(t)\} > E\{v_{j_1^* \dots j_{a^{s-1}}^* l_s}(t)\} \quad \forall t \quad (12)$$

$$(l_s = 1, \dots, r; l_s \neq j_{a^s}^*)$$

Q. E. D.

【著者紹介】

最上 義夫 (正会員)

1975年, 京都工芸繊維大学工芸学研究科修士課程修了. 同年徳島大学工学部情報工学科(現知能情報工学科)助手, 現在に至る. 学習オートマトン, 確率分布定数システムの観測器・制御器の最適配置問題の研究に従事. システム制御情報学会, 電子情報通信学会などの会員.

馬場 則夫 (正会員)

1974年, 京都大学大学院博士課程修了. 同年徳島大学工学部情報工学科助手, 75年同講師, 81年同助教授, 88年知能情報工学科教授, 90年大阪教育大学教育学部教授, 現在に至る. この間約2年間, フランスINRIA研究所およびオーストリアIIASA研究所客員研究員. 学習オートマトン, ゲーミング, 数理計画法, 環境工学, ロボット工学, ニューラルネットなどの研究に従事(工学博士). 著書: New Topics in Learning Automata Theory and Applications (Springer-Verlag), PC-9801パソコンゲーミング手法(日刊工業新聞), 数理計画法入門(共立出版). IEEE, システム制御情報学会, 日本ロボット学会, 日本オペレーションズ・リサーチ学会などの会員.

杉浦 公彦

1995年, 徳島大学工学研究科博士前期課程知能情報工学専攻修了. 同年東レエン지니어リング(株)勤務, 現在に至る. 学習オートマトンの研究に従事. システム制御情報学会の会員.