複数環境学習オートマトンネットワークによる ITS 自律走行制御

学生員 白石 和章* 正員 濱上 知樹*
正員 小坪 成一* 正員 平田 廣則*

Autonomous Driving Control for ITS by using Learning Automaton Network with Multi Environments

Kazuaki Shiraishi*, Student Member, Tomoki Hamagami*, Member, Seichi Koakutsu*, Member,
Hironori Hirata*, Member

In this paper, the application of the learning automaton (LA) network with multi environments is proposed for the adaptive controller for ITS autonomous driving. The LA network, which we have introduced previously, has the ability of the learning which deals with both plural reinforcement signals and information of multiple environments at the same time. The feature is found to be useful for improving the response of the adaptation in the dynamic environment like the highway. In order to evaluate the practical advantage of using the network, we designed the simulational highway system, constructed the autonomous travel controller using the simple LA and the LA network, and executed comparative experiments which evaluate the performance of adaptation response and the collision avoidance. The results show that the performance of LA network with multi environments is superior to the one using simple LA application on its stability and safety.

キーワード：学習オートマトン，高度道路交通システム，走行支援道路システム

Keywords: learning automaton, ITS, AHS

1. まえがき

高度道路交通システム (ITS: Intelligent Transport System) とは、交通体系全体を知能化するためのシステムであり、交通事故、交通渋滞、環境問題など、交通に関わる社会的な課題を解決する技術として、近年注目されている。特に走行支援道路システム (AHS: Automated Highway System) は、運転者の操作を支援、自動化することを目的とした技術であり、近年盛んに研究開発が行われている。

AHSにおいては、走行中に生じるさまざまな環境変化に対応して、適切な支援を行う自律走行制御機構が必要である。このような自律走行制御機構を実現する手法の1つとして、学習オートマトン (LA: Learning Automaton) を応用した自律走行制御機構が提案されている⁽¹⁾。

LAとは、未知環境 (RE: Random Environment) と確率オートマトン (SA: Stochastic Automaton) により定義される学習機械の一種であり、ノイズに強く収束性が保

証されているという利点を持つ。しかし、一般的な LA は、戦略の良否を判断するための信号として未知環境からの強化信号のみが使われる。そのため、動的な未知環境においては、学習が安定しないなどの問題が残されていた。

一方、LA に未知環境からの強化信号と、環境情報の両方を入力することが可能な LA ネットワーク⁽²⁾が提案されている。LA ネットワークは、LA 間のネットワークの重みを用いて学習を行うシステムであり、自律ロボットの制御などへの応用が期待されている。また、複数の未知環境を考慮した LA (複数環境 LA)⁽⁴⁾が提案されている。複数環境 LA は、ネットワーク構造を持つことができ、複数の動的未知環境下での学習が可能であるなどの特徴を持つ。近年、複数環境 LA を用いた群知能ロボットなどへの応用が期待されている。

本論文では、筆者らが提案している複数環境 LA⁽⁴⁾を応用した ITS 自律走行制御機構について述べる。本機構を計算機上に構築し、シミュレーションによりその基本的な性能を評価することを目的とする。

2. では、ITS 自律走行制御機構として、Ünsal らが提案した LA による手法と、本論文で述べる複数環境 LA を応用した手法の2つの手法についてそれぞれ述べる。3. では、

* 千葉大学

〒263-8522 千葉市稲毛区弥生町 1-33

Chiba University

1-33, Yayoi-cho, Inage-ku, Chiba 263-8522

本手法を用いた模擬高速道路シミュレーションによる自律走行実験とその結果について述べ、本手法の有効性を示す。

2. LAを用いた自律走行制御機構

AHSを想定したLAによる自律走行制御機構は、文献(1)によってその可能性が示された。LAによる自律走行制御機構は、単純で基本的な強化信号規則だけを設定すれば、様々な環境において自律的に行動を選択することが可能である。Ünsalらによる自律走行制御機構に用いられたLAのモデルは、強化信号入力のみを持つ従来のLAである。本章では、Ünsalの手法について述べ、その後に複数環境LAを応用した手法について述べる。

〈2・1〉 ÜnsalらによるLAを用いた自律走行制御機構
本節では、文献(1)で提案されたÜnsalらの手法の概略を述べ、その問題点について考察する。

Ünsalらは、図1に示すように2つのSAを用いた自律走行制御機構を提案した。それぞれのSAは、ロンジチュージナルSA(SA¹)、ラテラルSA(SA²)と呼ばれる。各SAは、未知環境から得られる強化信号*f*に基づき、学習を行い、以下に示す行動出力*y*₁, *y*₂を、それぞれ確率的に出力する。

- *y*¹: ACC(Accelerate: 加速), DEC(Decelerate: 減速), SS(Same Speed: 現在速度維持)のいずれか
- *y*²: SL(Shift to Left: 左車線変更), SR(Shift to Right: 右車線変更), SiL(Stay in Lane: 現在車線維持)のいずれか

各SAは以下で述べる行動選択確率修正式(強化法)に従い、行動選択確率を更新する。

時刻*n*において行動*i*が選択されたとき、*i*の行動選択確率*p*_{*i*}(*n*)及び、その他の行動の行動選択確率*p*_{*j*}(*n*)|*j* ≠ *i*を、式(1),(2)にて更新する。

$$p_i(n+1) = p_i(n) + f \cdot (-k \cdot \theta \cdot H(n)) \cdot [1 - p_i(n)] - [1 - f] \cdot (-\theta) \cdot [1 - p_i(n)] \dots\dots (1)$$

$$p_j(n+1) = p_j(n) - f \cdot (-k \cdot \theta \cdot H(n)) \cdot p_j(n) + [1 - f] \cdot (-\theta) \cdot p_j(n) \text{ for all } j \neq i \dots\dots\dots (2)$$

$$H(n) \equiv \min \left\{ 1; \max \left[\frac{p_i(n)}{k\theta(1 - p_i(n))} - \epsilon; 0 \right] \right\} \dots\dots\dots (3)$$

図1 Ünsalによる手法(LA)
Fig. 1. Ünsal's method (using LA).

ここで、強化信号*f*は、0: REWARD または、1: PENALTY のいずれかの値をとる。 θ は $0 < \theta < 1$ を満たす実数、*k*は $0 < k\theta < 1$ を満たす実数、 ϵ は任意の小さな正の実数である。

車両に搭載された4つのセンサは、表1, 2に示す各規則に従い、0: REWARD または 1: PENALTY のいずれかの出力を出す。それぞれのセンサの詳細は、〈3・1・1〉で後述する実験条件中で示す。各センサの出力は、図2に示される Mapping *F*により強化信号*f*に変換された後、前述のように各SAに入力される。各SAから出力された行動は、それぞれ固定長のバッファに蓄えられる。同一の行動が一定以上連続した場合に、制御が実行される。

Ünsalらによる手法は、強化信号以外の未知環境からの入力を持たない。そのため収束までの時間にばらつきが生じる。この特徴により、実環境下で障害物を回避する場合などにおいて、安定した動作が保証できないなどの問題が生じる。また、未知環境からの入力を持たないため、動的な未知環境においては、行動選択確率の収束に要する時間が長くなる可能性がある。

〈2・2〉 複数環境LAネットワーク LAネットワーク⁽²⁾は複数入力付きLAによりネットワークを構成する集団LAである。LAネットワークは、ベイズ識別法に基づいた強化法により、動的未知環境においても学習可能という特徴を持つ。ここで動的未知環境とはPENALTY 確率分布が、時間*t*により動的に変化する未知環境である。一方、複数環境LA⁽⁴⁾はLAネットワークを容易に構成でき

表1 フロントセンサの出力

Table 1. Output of the front sensor.

Current action	Vehicle in front sensor
SiL	PENALTY
SL	REWARD
SR	REWARD
ACC	PENALTY
DEC	REWARD
SS	PENALTY

表2 左右センサの出力

Table 2. Output of the left and right sensor.

Current action	Vehicle in sensor range	No vehicle in sensor range
SiL	REWARD/REWARD	REWARD/REWARD
SL	PENALTY/REWARD	REWARD/REWARD
SR	REWARD/PENALTY	REWARD/REWARD

図2 Mapping *F* の定義
Fig. 2. Definition of mapping *F*.

図3 複数環境 LA ネットワーク

Fig. 3. An example of LA network with multi environments.

る手法である。本論文では、文献(5)と同様に複数環境 LA ネットワークを用いた応用を考える。

〈2・2・1〉 複数環境 LA ネットワークの定義 複数環境 LA ネットワークの定義を以下に示す。

(定義) K 個に分割された未知環境における、 M 個の複数入力付き SA から構成される単一の階層の複数環境 LA ネットワークを以下の形式で定義する。

- SA^m ... m 番目の複数入力確率オートマトン
- Y^m ... SA^m の出力集合
- $y^m(n)$... 時刻 n における SA^m の出力
 $y^m(n) \in Y^m = \{-1, +1\}$
- $x^m(n)$... 時刻 n における SA^m への未知環境入力ベクトル
 $x^m(n) = (x_1(n), \dots, x_l(n))^T$
- $w^m(n)$... 時刻 n における SA^m への入力重みベクトル
 $w^m(n) = (w_1^m(n), \dots, w_l^m(n))^T$
- $p^m(n)$... 時刻 n における SA^m の行動選択確率分布
 $p^m(n) = \{p_1^m(n), p_2^m(n)\}$
- T^m ... SA^m の強化法
 $p^m(n+1) = T^m(w^m(n), x^m(n), y^m(n), f^k(n))$
- RE^k ... k 番目の未知環境
- $f^k(n)$... 時刻 n における RE^k からの強化信号
 $f^k(n) \in \{-1, +1\}$

図3に、複数環境 LA ネットワークの概念図を示す。各 SA には、入力ベクトル $x^m(n)$ が、それぞれ重みベクトル $w^m(n)$ による重み付けがなされた上で入力される。また、未知環境 RE^k は、 SA^m の出力結果に応じた強化信号 $f^k(n)$ を、各 SA^m に戻す。各 SA^m は、 $f^k(n)$ に基づき、 $w^m(n)$ を、以下に示す強化法により更新する。

(定義) 強化信号 $f^k(n)$ の入力を持つ SA^m の強化法 T^m

を以下に示す。

$$w^m(n+1) = w^m(n) + \Delta w^m(n) \dots \dots \dots (4)$$

$$\Delta w^m(n) = \begin{cases} \text{REWARD 応答 } (f^k(n) = 1) \\ -\zeta \{ \tau(w^{mT}(n)x^m(n)) - f^k(n)y^m(n) \} x^m(n) \\ \text{PENALTY 応答 } (f^k(n) = -1) \\ -\lambda \zeta \{ \tau(w^{mT}(n)x^m(n)) - f^k(n)y^m(n) \} x^m(n) \end{cases} \dots \dots \dots (5)$$

$$p_1^m(n+1) = \begin{cases} \text{REWARD 応答 } (f^k(n) = 1) \\ \Psi(-w^{mT}(n)x^m(n) + \zeta \|x^m(n)\|^2 (w^{mT}(n)x^m(n) - 1)) \\ \text{PENALTY 応答 } (f^k(n) = -1) \\ \Psi(-w^{mT}(n)x^m(n) + \lambda \zeta \|x^m(n)\|^2 (w^{mT}(n)x^m(n) + 1)) \end{cases} \dots \dots \dots (6)$$

$$p_2^m(n+1) = 1 - p_1^m(n+1) \dots \dots \dots (7)$$

$$\tau(\xi) = \begin{cases} -1, & \xi \leq -1 \\ \xi, & -1 < \xi < 1 \\ 1, & \xi \geq 1 \end{cases} \dots \dots \dots (8)$$

ここで、 λ, ζ は学習パラメータ、 $\Psi(\cdot)$ は $[0, 1]$ の単調増関数である。本論文では $\Psi(\cdot)$ としてシグモイド関数を用いる。

複数環境 LA ネットワークは、以下のアルゴリズムにより学習を進める。

- (1) ネットワークの構造を与える。
- (2) 時刻 $n = 0$ ですべての SA^m に対する重み $w_j^i(n)$ を一様乱数で初期化する。
- (3) 入力ベクトル $x^m(n)$ を入力層の各 SA^m に与える
- (4) ネットワークの最下層から順に確率分布に従って出力 $y^m(n) \in Y = \{-1, 1\}$ を決定する。
- (5) 未知環境からの応答 $f^k(n)$ を受け、強化法 T^m に基づいて重みを修正する。
- (6) (3) から (5) を終了条件を満たすまで繰り返す。終了条件は重みの変化率が一定の閾値以下になることである。ただし、終了条件は静的な環境においてのみ用いられ、動的な環境では終了条件を用いず (3) から (5) を無限に繰り返すオンライン学習となる。

〈2・3〉 計算機実験における複数環境 LA ネットワークの制約事項および仕様 〈2・2・1〉で示した複数環境 LA ネットワークを用い、本論文で仮定する道路モデルにおいてシミュレーションを行うための制約事項と、具体的な複数環境 LA ネットワークの仕様について述べる。

計算機実験における道路モデルは片側 3 車線の高速道路モデルとする。車線幅は日本道路協会の基準値 (1 種) に準じている。車線は直線とみなし、交差点、信号は考慮しない。この道路モデルにおいて自律走行を実現するための制

図4 複数環境 LA ネットワーク
(自律走行制御機構応用形)

Fig. 4. A LA network with multi environments (an autonomous driving control for ITS).

図5 自律ロボットの置かれる未知環境
および取り得る行動

Fig. 5. Environment where the autonomous robot is placed, and actions which can be taken.

約を複数環境 LA ネットワークに設ける。

図4に、複数環境 LA ネットワークを自律走行制御機構へ応用する場合の構成図を示す。図5に車両が選択可能な行動及びセンサの構成を示す。以下、自律的に動作する車両を自律ロボットと呼ぶ。

自律ロボットは、以下に示す仕様を持つ。

● SA の構成

4つの SA (SA^1 , SA^2 , SA^3 , SA^4) を持つ。各 SA は、 SA^1 : ACC (Accelerate: 加速), SA^2 : DEC (Decelerate: 減速), SA^3 : SL (Shift to Left: 左車線変更), SA^4 : SR (Shift to Right: 右車線変更) のそれぞれの行動に対して、行動選択確率分布に従った発火を起こす。

● 未知環境

3つの未知環境 (RE^1 , RE^2 , RE^3) を想定する。各 RE は、 RE^1 : 自律ロボット前方の未知環境, RE^2 : 同左側未知環境, RE^3 : 同右側未知環境をそれぞれ表す。

● 未知環境情報の入力

未知環境からの入力として、自律ロボットに搭載された4つのセンサ出力を用いる。それぞれのセンサは、 x_1 : 同車線の前方を走行する他自律ロボットとの距離, x_2 : 自身の走行速度, x_3 : 左側の障害物, または左車線を走行する他自律ロボットとの距離, x_4 : 同右側をそれぞれ観測し、[0,1] に正規化した後に、SA に入力

表3 強化信号規則 (前方センサ)

Table 3. Rules of reinforcement signal (front sensor).

Current action	Vehicle in front sensor	No vehicle in front sensor
ACC ON	PENALTY	REWARD
ACC OFF	REWARD	PENALTY
DEC ON	REWARD	PENALTY
DEC OFF	PENALTY	REWARD

表4 強化信号規則 (左右センサ)

Table 4. Rules of reinforcement signal (left and right sensor).

Current action	Vehicle in left and right sensor	No vehicle in left and right sensor
SL ON	PENALTY	REWARD
SL OFF	REWARD	PENALTY
SR ON	PENALTY	PENALTY
SR OFF	REWARD	REWARD

Current action	Vehicle in left sensor	Vehicle in right sensor
SL ON	PENALTY	REWARD
SL OFF	REWARD	PENALTY
SR ON	REWARD	PENALTY
SR OFF	PENALTY	REWARD

される。

● 強化信号

強化信号は、現在の行動と各センサの出力に基づき、表3、4に示す規則により出力される。また、自律ロボットは表3、4の強化信号規則により、おおまかな行動選択パターンが決定される。表で示した条件以外で SA が発火した場合は、すべて PENALTY とする。

● ロンジチュージナル制御

複数環境 LA ネットワークの出力が直接実行される。ただし ACC と DEC が同時に発火した場合は、現在速度を維持するものとし、また、予め決められた上限設定速度以上には、速度は上がらないものとする。

● ラテラル制御

複数環境 LA ネットワークの出力は、一旦固定長バッファに蓄えられる。同一の行動がバッファ中の一定以上の割合を越えた場合に、制御が実行される。これは車線変更を安全に行うための処理である。なお、左右の車線に車線変更可能な場合、表4の強化信号規則により左車線変更を優先する。また、SL と SR が同時に発火した場合は、無効な行動として処理される。なお、計算機実験において車線変更が実行された場合、結果は「車線変更成功」と「車線変更失敗 (衝突)」の2通りのみとする。また、車線変更開始から車線変更終了までの時間は考慮しない。

3. 計算機実験

〈2・1〉で述べた Ünsal らの手法⁽¹⁾と〈2・2〉〈2・3〉で述べた本手法を比較し、本手法の有効性について評価するために2つのタスクについて実験を行う。

〈3・1〉 追い抜きタスクによる評価実験 前方を走行する、他の自律ロボットを追い抜くタスクにおいて評価実験を行う。

〈3・1・1〉 タスク設定 図6に本実験で行うタスクを示す。このタスクは前方を一定速度で走行する自律ロボットを、後方の自律ロボットが衝突を回避しながら、追い抜くタスクである。なお後方を走行する自律ロボットは前方センサ障害物回避範囲内に、他自律ロボットが存在しない距離からスタートする。

本タスクにおいて、以下の2点について評価する。

(1) 車線変更に要した時間

前方の自律ロボットがセンサ感知されてから、後方の自律ロボットの車線変更が行われるまでの時間間隔を調べる。前方の自律ロボットの速度は70km/hに固定し、後方の自律ロボットの上限設定速度は120km/hに設定する。1回の追い抜き動作を1試行として、それぞれの手法について、1セット40試行の実験を10セットずつ行う。

(2) 車線変更に要した時間のばらつき

前述車線変更に要した時間が、試行毎に、また、相対速度差によりどの程度ばらつくのかを調べる。前方の自律ロボットの固定速度は、70km/h から110km/h の間で10km/h の単位で設定し、後方の自律ロボットの上限設定速度は120km/h に設定する。各手法、各相対速度の組合せにつき、500回の試行をそれぞれ行い、その統計的な特徴について調べる。

いずれの実験においても、計算速度は50iteration/secとする。計算機実験において設定した条件を以下に示す。

- 自律ロボットのサイズを車両幅2[m]、車両長4[m]とする。
- センサの障害物感知範囲を図7に示す。

図6 実験を行った状況

Fig. 6. An experimental environment.

図7 センサ構成図

Fig. 7. Sensors range of a robot.

- Ünsalらの手法におけるパラメータを以下に示す。

- パラメータ: $\theta = 0.06$ $k = 4$ $\epsilon = 0.01$
- 計算速度: 50[iteration/sec]
- バッファの大きさ: 50[iteration]

- 本手法におけるパラメータを以下に示す。

- パラメータ: $\zeta = 0.1$ $\lambda = 0.01$
- 計算速度: 50[iteration/sec]
- バッファの大きさ: 50[iteration]

〈3・1・2〉 結果 結果の比較に先立ち、追い抜きタスクの遂行において、本手法がどのような行動確率分布の変化を生じさせるかを調べる。図8に、自律ロボットのロンジチュージナル制御における加速、減速に割り当てられたSAの行動選択確率の変化を示す。また、図9に、本手法におけるラテラル制御における左車線変更、右車線変更についての行動選択確率の変化を示す。いずれも、10試行後の応答であり、前方自律ロボットを感知してから追い抜きが完了するまでのステップ毎の行動確率を表している。

これらの結果から、障害物を避けるため減速SAおよび車線変更SAの行動選択確率が、前方自律ロボット感知後、即座に1.0に変化していることが確認できる。車線変更後は設定速度に達するため、加速の行動選択確率が高くなっている。これは「車線変更に伴う環境の変化」に適した行動の選択確率が増加したことを意味しており、動的未知環

図8 ロンジチュージナル制御SAの行動選択確率の変化
(複数環境LAネットワーク)

Fig. 8. The action probability of longitudinal SA (a LA network with multi environments).

図9 ラテラル制御SAの行動選択確率の変化
(複数環境LAネットワーク)

Fig. 9. The action probability of lateral SA (a LA network with multi environments).

図 10 試行毎の車線変更時間の推移

Fig. 10. The required time for changing a lane.

境に適應できたと言える。

図 10 に Ünsal らによる手法と本手法のそれぞれについて、車線変更選択までの時間を試行毎の平均値で示す。本手法を用いた場合、数回の試行で学習による効果が現れ車線変更選択までの時間が減少しているのに対し、Ünsal による手法では減少傾向は見られない。これは LA のみによる手法は、過去に経験した未知環境における学習結果を記憶する機構がないためと考えられる。これに対し、本手法は、ネットワークの重みにより未知環境を学習することができ、同一の未知環境に再び遭遇した際に、即座に対応を取ることが可能である。

図 11 に、Ünsal らによる手法を用いた場合の、初期速度差毎の車線変更に必要な時間及びその標準偏差を示す。また、図 12 に、本手法を用いた場合の同様の結果を示す。

これらの結果から、複数入力 LA ネットワークを用いた場合、LA のみを用いた手法に比べ行動実行までの時間が短いことがわかる。さらに LA のみを用いた手法は行動出力決定までの時間が安定していないのに対し、本手法ではほとんどばらつきがないことがわかる。

このように、本手法は重みが収束した状態であれば、行動選択に必要な時間が短くなり、また行動実行までの時間的なばらつきも抑えられる。その結果、行動を速やかにかつ安定して出力することが可能である。このような特徴は、Ünsal らの手法のような LA のみを用いた手法に比べ、より確実な動作を求められる実機における応用に適しているといえる。

〈3・2〉複数台同時走行実験における評価実験 〈3・1〉で行った計算機実験の自律ロボットを道路モデル上で複数台走行させ、その衝突回数を測定する。

〈3・2・1〉タスク設定 このタスクは道路モデル上で複数台で衝突を回避しながら走行することを目的とする。自律ロボットは〈3・1〉で設定した条件と同一である。衝突回数とは、単位時間、一定区間での衝突回数を 1 車線あたりの自律ロボット密度ごとに測定したものである。初期配置は等間隔に配置し、設定速度は 80km から 100km の間でランダムとする。この設定条件で 1 時間あたりの 1[台/(km・lane)] から 8[台/(km・lane)] までの衝突回数を測定する。

〈3・2・2〉結果 500 回の試行を行った結果の平均を図 13 に示す。Ünsal の手法に比べ本手法は、衝突回数が少ない。Ünsal の手法において、多く観察された衝突の原因は、行動選択確率の収束時間に伴う行動実行の遅延で

図 11 Ünsal らの手法における車線変更に必要な時間

Fig. 11. The required time for changing a lane (using Ünsal's method).

図 12 複数環境 LA ネットワークにおける車線変更に必要な時間

Fig. 12. The required time for changing a lane (using a LA network with multi environments).

図 13 事故発生回数対自律ロボット密度

Fig. 13. The number of collision occurrence against the density of robot.

あった。これは、〈3・1〉に示されている、行動選択確率の収束時間のばらつきの大きさが関与していると考えられる。一方、本手法において自律ロボット密度が高い場合、極端に速度差のある自律ロボット同士の衝突が発生した。これは減速性能の限界等により、車線変更条件(車線変更バッファ内における車線変更行動の割合が一定以上)を満たす前に衝突したものである。しかし、本手法における衝突発生率は Ünsal の手法と比較して明らかに低く、本手法の有効性が確認できる。

4. むすび

本論文では、ITS における自律走行制御の実現手法として、複数入力を持ち未知環境からの入力も複数持つ、複数環境 LA ネットワークを応用し、Ünsal らによる手法と比較検討した。本手法の特徴は動的未知環境において「速やかな行動選択」「行動選択に要する時間のばらつきの少なさ」「状況毎における学習」を実現できることである。計算機実験の結果、ITS 自律走行制御における制御機構に適していることが、明らかになった。

今後の課題として車々間通信、路車間通信の考慮、合流制御などへの拡張が考えられる。

(平成 14 年 7 月 1 日受付、平成 15 年 2 月 18 日再受付)

文 献

- (1) Cem Ünsal, Pushkin Kachroo, and John S. Bay: "Multiple Stochastic Learning Automata for Vehicle Path Control in an Automated Highway System", *IEEE Trans. System, Man, and Cybernetics*, Vol.29, No.1, pp.120-128 (1999-1)
- (2) F. Qian and F. Unno: "A Learning Automaton Network and its Dynamics", *T. IEE Japan*, Vol.111-C, No.10, pp.490-499 (1991-10)(in Japanese)
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」, 電学論誌 C, **111**, 10, pp.490-499 (1991-10)
- (3) T. Toyama, S. Koakutsu, F. Qian, and H. Hirata: "Learning of Controlling Behavior Using Learning Automaton Networks with Communications", *T. IEE Japan*, Vol.118-C, No.5, pp.781-786 (1998-5)(in Japanese)
外山達彦・小坪成一・銭 飛・平田廣則:「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」, 電学論誌 C, **118**, 5, pp.781-786 (1998-5)
- (4) F. Qian and H. Hirata: "A Learning Automaton with Context Inputs on Multi-Teacher Environments", *T. IEE Japan*, Vol.119-C, No.5, pp.575-582 (1999-5)(in Japanese)
銭 飛・平田廣則:「N 環境付き多入力学習オートマトン」, 電学論誌 C, **119**, 5, pp.575-582 (1999-5)
- (5) A. Tanaka, F. Qian, and H. Hirata: "A Multi-Teacher Learning Automata Resolution of Cooperative Pursuit Problems", *T. IEE Japan*, Vol.121-C, No.4, pp.703-709 (2001-4)(in Japanese)
田中昭信・銭 飛・平田廣則:「複数教師付き学習オートマトンの追跡問題への適用」, 電学論誌 C, **121**, 4, pp.703-709 (2001-4)
- (6) K.S.Narendra and M.A.L.Thathachar: "Learning Automata an Introduction", Prentice-Hall International (1989)
- (7) M.L.Tsetlin: "On behavior of finite automata in random media", *Automation and Remote Control*, Vol.22, No.11, pp.1345-1354 (1961)
- (8) A. Widodo and T. Hasegawa: "Autonomous Cruising Traffic Flow Simulators for the ITS Including the Inter-Vehicle Communication Network", *IEICE*, Vol.J82-B, No.11, pp.2002-2009 (1999-11)(in Japanese)
Ari.Widodo, 長谷川孝明:「車々間通信ネットワークを含めた高度交通システムの評価のための自律走行交通シミュレータ」, 信学論誌 B, **J82-B**, 11, pp.2002-2009 (1999-11)

白石 和 章 (学生員) 1971 年 8 月 10 日生。2001 年広島国際学院大学電子工学科修士課程修了。同年、千葉大学大学院自然科学研究科人工システム科学専攻博士後期課程に入学、現在に至る。学習オートマトン、強化学習の研究に従事。

濱上 知 樹 (正員) 1966 年 3 月 7 日生。88 年千葉大学工学部電気工学科卒業。セコム(株)勤務を経て、99 年在職のまま千葉大学大学院自然科学研究科生産科学専攻後期課程修了。2001 年同助手となり現在に至る。大規模システムの数理、マルチエージェント、機械学習、創発システムの応用研究に従事。博士(工学)。IEEE、電子情報通信学会会員。

小 坪 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。同年千葉大学工学部助手、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。94~95 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。

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持つ。工学博士。2001 年電気学会業績賞受賞。IEEE(fellow), 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。