拘束ルール抽出機構を用いた自律移動ロボットの段階的行動学習[†]

近 藤 敏 之*・伊 藤 紀 彦**・伊 藤 宏 司*

An Incremental Behavior Learning Using Constraint Rules Extraction Mechanism for Autonomous Mobile Robots[†]

Toshiyuki KONDO*, Norihiko ITOH** and Koji ITO*

In the last decade, a number of skillful robots have been developed. However most of them can so far only demonstrate pre-programmed motions according to the external stimuli. In contrast, humans can learn new motions such as catching a ball in spite of his/her high dimensional sensorimotor DOF. In the learning process, it can be hypothesized that the learner actively constrains the DOF by him/her-self using *learning skills*, in this paper referred to as “*schema*”. In this study, a learning method for autonomous mobile robots operating in unknown environments is proposed, where not only a behavior learning mechanism for sensorimotor mappings but also an extraction/re-use mechanism of the schema (i.e. constraint rules for behavior learning) are implemented. Through the results of simulations and real experiments of mobile robot navigation, the validity of the proposed method is clarified.

Key Words: constraint rules extraction mechanism, reinforcement learning, incremental learning, real-world learning, autonomous mobile robot

1. はじめに

近年の工学技術の発展は目覚しく、多様なセンサ、多自由度のアクチュエータを備え、高度な運動能力を有するロボットの開発が盛んに進められている。しかしながら、このようなシステムの挙動を設計の段階であらかじめ把握することは一般に困難な問題であり、このため制御器設計の実際は依然として設計者の経験・知識に頼る部分が大きいのが現状である。一方、ロボットが家庭やオフィスなど多くの不確定要素を含む環境下で頑健に動作し続けるためには、みずからが持つ行動知識を動作環境に合うように適宜調節する、自律的な適応能力を有することが期待される。制御器設計を自律化する適応学習の枠組みとして、近年、強化学習^{1),6)}や進化ロボティクス⁴⁾などの、いわゆる構成論的設計手法の有用性が認められつつあるが、実ロボットに対して同設計法をナীবに適用した場合、評価に要する時間がボトルネックとなり実用時間での学習は望めない。

この問題を解決するひとつの方法として、システムの自由度に拘束をかけて探索空間を低次元化した上で学習を行うことが考えられる。たとえば浅間らは、「ロボットが、未知環境を能動的に「移動」することが「知」を獲得する上で本質的である」という移動知の概念を提唱し、「知的」にふるまう人工システムを設計・実現するための重要な要素のひとつとして、環境と身体拘束条件の自律生成を挙げている³⁾。これはすなわち、学習主体(ロボット)が動的に変化する環境に対して自律的に適応するためには、観測空間から行動空間への写像それ自体よりも、その関係を時空間的に拘束する内的な拘束条件をいかに自律生成するかが重要であることを示唆している。しかしながら、システムが環境との相互作用を通じて自律的に拘束条件を生成していく過程の具体的な実現手法は提案されていない。

一方、川人らは多自由度を有するヒューマノイドの運動学習に「見まね学習」を導入することにより、この拘束条件の生成問題の解決を試みている⁵⁾。ロボットは環境に備えつけられたカメラで手本となる人間の動作を観察してその関節軌道などを計測し、状態予測を行なうための内部モデルを構築する。このモデルに基づいて自身の目標軌道を求めている。このように実在する教師を手本として拘束条件を与える見まね学習のアプローチは、多自由度空間における運動学習のひとつの有効な実現手法であるといえるが、高度な3次元計測システムによる多量の関節軌道データを必要とすることや、試行錯誤の過程を経ないインスタンス学習に特有の、

[†] SICE Annual Conference in Fukui 2003にて発表(2003・8)

* 東京工業大学大学院総合理工学研究科
横浜市緑区長津田町 4259

** 日本アイ・ビー・エム(株)

* Tokyo Institute of Technology, 4259 Nagatsuta, Midori-ku, Yokohama

** IBM Japan

(Received August 19, 2003)

(Revised November 27, 2003)

学習済み運動パターンの汎化性の問題、適用対象がヒューマノイドに限定されるなどの問題点を有する。先の移動知でも提唱されているように、ロボットの行動学習では運動パターンを獲得する際に得た知識（拘束条件）を、新たな動作を学習する際に活用することで、新規運動パターンの学習をより効率化することが可能になると考えられる。

一方、無限定な環境下でも頑健に行動する生物は、この問題を本質的に解消しているように見える。一般にわれわれは何か新しい運動を行なう際、「既学習の運動パターンとの類似性に基づいて、感覚・運動の自由度に拘束（選択的注意¹⁶⁾や筋の同時活性¹¹⁾など）を働かせること」により、運動学習を効率的に行なっていると考えられる。

またスポーツのトレーニングに見られるように、初心者が複雑な運動を学習する際には、まず容易な条件から学習を開始し、段階的にタスクの複雑さを増すことで最終的に高度な運動技能を獲得することが可能になる。このような方法論は、認知行動学の分野でシェイピング (shaping) と呼ばれる学習の計画問題であり、ロボット工学の分野でも認知発達ロボティクスとして注目を集めている^{2), 12), 13)}。しかしながら、タスクの性質を熟知した設計者が学習のさせ方を明示的に計画する場合、学習の自律性の面で問題がある。

以上の背景に基づき、本研究ではつぎのことを明らかにすることを目的とする。

- (1) 一度獲得された学習のコツ（拘束ルール）は、さらにそれを利用することで類似性があるタスクの学習を効率的にする。
- (2) シミュレータ上で得られた拘束ルールを実ロボットに学習に適用することで、実環境下での学習時間が低減される。

本研究のように学習のコツを知識として抽出し、これを用いて自律移動ロボットの学習を効率化する試みとしてはかに井谷らの研究⁹⁾を挙げることができる。井谷らは、移動ロボットの行動獲得問題において人間が与えた教示情報の再構成により行動ルールを獲得し、未知環境下でルールに基づいたロボットの行動から、不要なルールが削除できることを示している。ここで井谷らが学習初期における人間-ロボット間の教示情報の理解を高めるために、ルールの可読性を重視しているのに対し、本研究では複雑多自由度なロボットが自律学習する上で段階的に行動知識（センサー-モータ間写像および学習の拘束条件）を獲得していく過程に着目し、各ルールはロボットが将来の学習で利用しやすい形式で保存されていく点が大きく異なる。

2. 拘束ルール抽出機構を用いた強化学習法

2.1 自律移動ロボットのナビゲーション課題

後の説明を容易にするため、先に本研究で用いた自律移動ロボットのナビゲーション課題について説明する。ここで、ロボットの学習の目的は、壁に衝突することなく、速やかに未知形状の回廊（幅 1000[mm]）を周回するためのセンサー-

モータ間写像を獲得することである。本研究では実機による実験を想定し、移動ロボット（Pioneer 2-DXe, ActivMedia Robotics 製, Fig. 1）のシミュレータを作成した（Fig. 2）。Pioneer 2 の外形は完全な円形ではないが、ここでは簡単のため、仕様上の最小回転半径 260[mm] の円で近似した。Fig. 1 に示すように、ロボットには壁までの距離を計測するための超音波センサが 16 方向に取り付けられており、本実験では図中に丸で示した 8 基を使用した。また、ロボットの行動は左右の車輪速度 (v_r, v_l) をおのおの独立に指定することによって制御する。

Fig. 1 Pioneer 2-DXe (ActivMedia Robotics)

Fig. 2 A simulation environment (Env.A)

本シミュレータでは簡単のため、(1) 式のようにロボットの運動学のみを考慮した。ここで x, y, ϕ はロボットの中心座標および姿勢、 $2d$ は車輪間の距離、 r_d は車輪の半径である。

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\phi} \end{pmatrix} = \frac{r_d}{2} \begin{pmatrix} \cos \phi & \cos \phi \\ \sin \phi & \sin \phi \\ \frac{1}{d} & -\frac{1}{d} \end{pmatrix} \begin{pmatrix} v_r \\ v_l \end{pmatrix} \quad (1)$$

2.2 学習アーキテクチャ

2.2.1 Actor-Critic 強化学習

Fig. 3 The structure of a sensorimotor mapping

ロボットの知能を設計することは、センサ-モータ間の写像関係を記述することにはかならない。一般に、このセンサ-モータ間写像は、動作環境の性質、超音波センサならびに駆動機構の特性に起因する非線型性を含むと考えられる。したがって、本研究では任意の非線型関数を近似表現することが可能なニューラルネットワーク (artificial neural networks, Fig. 3) を用いて制御器を表現することとする¹⁴⁾。

本研究では、ニューラルネットワークの結合荷重の学習に Actor-Critic 強化学習^{1), 6)} を用いた。ここで Actor-Net はロボットの制御器であり、状態入力 x_t から行動出力 $a(x_t)$ への写像である。一方、Critic-Net は状態評価器であり、状態入力 x_t から状態価値の推定値 $\hat{V}(x_t)$ への写像である。ここで状態価値 V_t は、適宜得られる報酬 r_t に将来の報酬に対してどの程度割り引いて評価するかを決定する割引率 γ ($0 \leq \gamma \leq 1$) をかけて (2) 式のように定義される。

$$V_t \equiv r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \dots = \sum_{i=0}^{\infty} \gamma^i r_{t+i+1} \quad (2)$$

ロボットの報酬 r_t は (3) 式のように、平均移動速度 (第1項)、直進率 (第2項)、距離センサ $d(i)$ の最小値 (第3項) の線形和として算出される。ここで C_1 , C_2 および C_3 は各項を 1.0 に正規化するための係数である。すなわちこの報酬関数の下では、壁との衝突を回避し、より迅速かつ直線的に周回動作することが求められる。

$$r_t = C_1 \frac{|v_r + v_l|}{2} + C_2 \frac{1}{1 + \frac{|v_r - v_l|}{2}} + C_3 \min_{i \in M} d(i) \quad (3)$$

また、評価の推定値 $\hat{V}_t$ と報酬 r_t から、Actor-Net への強化信号となる TD(Temporal Difference) 誤差 ε は (4) 式のように計算される。

$$\varepsilon_t = r_t + \gamma \hat{V}(x_{t+1}) - \hat{V}(x_t) \quad (4)$$

2.2.2 学習の手順

Actor-Net ならびに Critic-Net は以下の手続きにしたがって更新される¹⁷⁾。

- (1) ロボットは環境からセンサ入力 x_t を観測する。
- (2) x_t を Actor-Net に入力し、行動 $a(x_t)$ を計算する。この値を中心とした正規分布から確率的に行動 $\tilde{a}(x_t)$ (すなわち v_r, v_l) を求め、単位時間実行する。

$$\tilde{a}(x_t) = a(x_t) + N(0, \Sigma) \quad (5)$$

- (3) ロボットは環境から報酬 r_t を受け取り、つぎのセンサ入力 x_{t+1} を観測する。Actor-Net の更新量 $\Delta a(x_t)$ を算出するため、(4) 式に基づき TD 誤差 ε_t を計算する。

$$\Delta a(x_t) = \alpha (\tilde{a}(x_t) - a(x_t)) \varepsilon_t \quad (6)$$

ここで、 α は学習率である。 $\varepsilon_t > 0$ ならば、実行した行動 $\tilde{a}(x_t)$ は好ましいと考えられるので、その方向に行動 $a(x_t)$ を修正する。 $\varepsilon_t < 0$ の場合はその逆である。ここでは $a(x_t) + \Delta a(x_t)$ を教師信号として $a(x_t)$ を BP(Back Propagation) 法で学習する。

- (4) TD 誤差を用いて Critic-Net の重みを更新する。 $\hat{V}(x_t) + \varepsilon_t$ を教師信号として $\hat{V}(x_t)$ を BP 法で学習する。
- (5) 1に戻る。

2.3 拘束ルールの抽出と選択

本研究では、学習したニューラルネットから抽出した知識 (拘束ルール) を将来の学習において拘束条件として利用する、状態拘束型適応システム (Fig. 4) を提案する。同手法は 1) 既学習のニューラルネットワークからの拘束ルール群の抽出、2) 未学習環境下で有効に機能する拘束ルールの選択、および 3) 拘束ルールを用いた強化学習の 3 過程からなる。以降で各過程について説明する。

2.3.1 NN からの知識抽出

ニューラルネットから知識を抽出する方法には大きく分けて 2 通りある¹⁸⁾。ひとつは分解法 (decomposition method) と呼ばれており、各ユニットの出力と入力の重みの関係から論理関数としてルールを抽出する方法^{8), 10), 15)} である。この手法では、離散化された特定の入力 (たとえばすべての入力が 1) における重みの出力に対する影響度からルールを抽出する。しかしながら、本研究では制御器をニューラルネットワークで表現しており、その入出力は連続値であるため、分解法をそのまま用いることはできない。もうひとつは学習後のニューラルネットにテスト入力を与え、その出力を得ることによって学習用のデータを集め、その入出力データベースから帰納的にルールを抽出する方法である⁷⁾。本研究ではこの帰納法に基づいて、学習を終えたロボットが、学習した環境で行動する際の間層ニューロンの発火状態に着目してニューラルネットの枝刈りを行なうルール抽出法を採用した。

2.3.2 拘束ルール抽出過程

拘束ルールは、周回行動を学習し終えたロボットが、学習済みの環境下で再び周回行動を行なう間に、以下の手順にしたがって抽出される。

- (1) ロボットが行動中に、拘束ルール抽出条件 (中間層の出力値が閾値 θ_a を越える) を満たすとき、ニューラル

Fig. 4 The architecture of the proposed schema extraction mechanism

ネットワークの入出力（センサ値および速度指令値）を保持し、以下の拘束ルール抽出操作を行なう。したがって、ここで抽出されるルール数は θ_a に依存する。

(2) センサの値は θ_b を閾値として、 θ_b 以上なら 0 に、 θ_b 以下なら 1 に離散化。

(3) 離散化されたセンサ値、速度指令値（連続値）ならびに活性化した中間層ニューロンの ID 番号（ID はあらかじめ付与）を、拘束ルールとしてメモリに保存。

(4) 1に戻る。

この過程をロボットが既学習環境を一周し終えるまで繰返す。複数種類の環境について同様の作業を行ない、拘束ルールを蓄積する。ここで示した、ニューラルネットから拘束ルールを抽出する方法を Fig. 5 に模式的に示す。

Fig. 5 The architecture of extracting constraint rules

2.3.3 拘束ルール選択過程

センサが同じ状況であってもタスクの目的が異なれば行動価値関数が異なるように、過去に抽出された拘束ルール群すべてが、常に有効に働くとは限らない。本研究では、各ルールの有効性を検証する事前選択過程 (pre-selection process)

を導入することで、タスクおよび環境との相互作用の結果として有効な拘束条件が自律的に選択されるものとした。ここで各拘束ルールの有効性は、ロボットが未学習環境で拘束ルールを行動プリミティブとして適用した際の、ロボットの行動評価から決定される。すなわち未知環境で周回する際に、

- (1) ロボットのセンサ入力を θ_b を閾値として、 θ_b 以上なら 0 に、 θ_b 以下なら 1 に離散化する。
- (2) 保持している拘束ルールのセンサ条件部が、現在のセンサ入力とすべて一致した拘束ルールは活性化され、ロボットはそのルールに記述された行動を出力する。
- (3) 行動の結果得られる即時報酬を活性化された拘束ルールの評価値として、有効な拘束ルールを選択する。

以上の過程を繰返し、拘束ルールの有効性を評価する。このように、選択は未知環境で行なわれるため、各環境で有効に機能すると予測される拘束ルールのみが選択される。

2.3.4 拘束ルールを用いた学習

選択された拘束ルールは、拘束ルールに記された中間層ユニットに関係する結合加重を固定する（すなわち、BP 学習の対象から除外する）ことで拘束を与える。この際、固定されるニューラルネットの結合荷重は、先に学習を終えたニューラルネットのものをを用いた。ここで示した拘束ルールを用いた学習法を Fig. 6 に模式的に示す。

Fig. 6 How to constrain weight parameters

3. 段階的行動学習シミュレーション

3.1 シミュレーション設定

本研究の目的は、未学習状態（ニューラルネットの結合加重がランダムな値に設定された状態）から周回行動を獲得する過程（本稿では一次学習と呼ぶ）において得られた知識から学習に有効に働く拘束条件を抽出し、経路形状の異なる環境で学習する際にもこの拘束ルールを用いること（これを二次学習と呼ぶ）により、学習が高速化されることを示すこ

とである。このため、3つの異なる環境（環境A（Fig. 7）、環境B（Fig. 8）および環境C（Fig. 9））を用いて以下の2種類のシミュレーション実験を行ない、両者の比較を行なう。

実験1

未学習状態から、各環境（A～C）を独立に学習させる。1試行は5000ステップの行動選択（1ステップは0.2[sec]とした）からなり、各環境について異なる乱数シードを用いて50試行の学習を行なう。50試行の内、最終的に周回行動が達成された比率を学習の成功率 S_1 とする。

実験2

実験1と同様に、未学習状態から環境Aで周回行動を学習し、拘束ルールを抽出する。このとき獲得した拘束ルールを用いて未学習の環境Bで行動し、有効に働くと予測される拘束ルールを選択する。選択された拘束ルールは、ロボットが環境Bで周回行動を学習する際に、一部の結合加重が更新されないように拘束する。環境Bにおける学習の終了後、再度拘束ルールを抽出する。2つの環境（A, B）で蓄積された拘束ルールを用いて未学習の環境Cで行動し、有効に働くと予測される拘束ルールを選択し、これを用いて周回行動を学習する。実験2においても異なる乱数シードを用いて50試行の学習を行ない、周回行動の成功率 S_2 を求める。

ここで、 $S_1 < S_2$ となれば、拘束ルールを用いた学習の有効性を定量的に示すことができると考えられる。

3.2 シミュレーション結果

上記の手順にしたがって行なった実験1と実験2の結果を示す。実験に用いた学習係数は、Table 1に示す値とした。

実験1では環境A～Cで一次学習を行なった。Fig. 7～

Fig. 7 A trajectory of robot in Env.A (Condition1)

Fig. 8 A trajectory of robot in Env.B (Condition1)

Fig. 9 A trajectory of robot in Env.C (Condition1)

Table 1 Parameters

γ :	discount factor	0.95
α :	leaning rate for TD learning	0.1
η :	learning rate for BP	0.01
T :	the number of trials	5000
t :	the number of actions in a trial	2000
θ_a :	threshold for schema extraction for hidden units	0.9
θ_b :	threshold for schema extraction for sensor units	0.4

Fig. 9に示すロボットの行動軌跡から、おのおのの環境においてロボットが周回行動を学習したことが確認できる。

しかしながら、Fig. 14に示すように、環境B, Cにおいて一次学習を行なった際の成功率は共に4%と低く、これらの環境において5000ステップの行動選択で周回行動を実現することの困難さが伺える。

Fig. 10 A trajectory of robot in Env.B (Condition2)

Fig. 11 A trajectory of robot in Env.C (Condition2)

つぎに、実験2の結果について示す。拘束ルールの抽出で行なう枝刈りでは、出力の値が大きい中間層ユニットに關係する部分のみが残される。たとえば、Fig. 12は環境Aを移動中のロボットの中間層活性値の推移を示しているが、

Fig. 13 A Trajectory of robot in Env.B according only to the schema extracted in Env.A

同図に示された大きく変化する4つの中間層 (unit0, unit1, unit5, unit9) に接続された結合のみを有効にしたニューラルネットワークでも周回行動を実現することが可能なことから、ほかの中間層ニューロンに接続する結合を枝刈りすることは有効であると考えられる。

つぎに拘束ルール選択の結果を示す。抽出された拘束ルールの有効性は環境依存であり、常に有効に働くとは限らない。このため本研究では、ルールを適用する前にルールの有効性を評価する事前選択を導入している。本実験では、新しい環境でロボットが行動し、ロボットのセンサの入力パターン (閾値 θ_0 以上なら 0, それ以下なら 1 に離散化する) と拘束ルールが記憶しているパターン (同様に閾値 θ_0 で離散化) がすべて一致する場合に、その拘束ルールが選択されるものとした。本実験では検出距離が 600[mm] のところでセンサ値を離散化することを想定し、 θ_0 を 0.4 とした。たとえば、環境 A で学習を終えたニューラルネットワークから抽出された拘束ルール 10 個をプリミティブとして用い、環境 B で周回行動させた結果、3 個のルールを有効なルールとして選択することができた。環境 A で獲得した拘束ルールを用いて、環境 B で周回行動を行なった際の移動軌跡を Fig. 13 に示す。同図に示すように、拘束ルールのみで行動した場合でも環境 B の途中まで到達することができることから、ルールは周回行動で有効に機能する行動ルールを抽出できていると考えられる。

提案手法の有効性を定量的に示すため、実験 1 と実験 2 を複数回 (50 回) 実施し、周回行動学習の成功率を比較した結果を Fig. 14 に示す。同図より一次学習 (Condition 1) のみでは環境 B では 4%, 環境 C では 4% と、環境 A の 64% の成功率から比べると非常に低い成功率となることが確認できる。これに対し、二次学習 (Condition 2) を行なった場合には環境 B では 21.9%, 環境 C では 71.4% と成功率が向上することが確認できる。この結果より、抽出された拘束ルールは別の環境下で学習する際に無駄な試行錯誤を抑制することによって有効に機能すること、ならびに複雑な課題の学習には段階的に拘束ルールを抽出することが有効であることが示された。

Fig. 14 Comparison of success rate

4. 実機による検証実験

一方、実環境における本手法の有効性を確認するため、先に示した状態拘束型適応システムを用いて実機による検証実験を行なった。Fig. 1 で示したように、本実験では ActivMedia 社の Pioneer 2 ロボットを用いた。Pioneer 2 は周囲 16 方向に超音波センサを搭載しているが、本実験ではこのうち 8 方向のセンサを用いた。センサの値は、たとえば計測された距離が最大検出距離 1000[mm] のときに 0.0, 最近接 0[mm] のとき 1.0 となるように、[0, 1] の連続値として線形変換され、これをニューラルネットワークへの入力とした。

シミュレーションにおいて得られた拘束ルール群の、実環境での有効性を確認するため、拘束ルールのみで周回行動を行なった結果を Fig. 15 に示す。同図に見られるように、行動プリミティブとして用いる拘束ルールの条件部が 0/1 に離散化されているために、シミュレータと実機のセンサ特性の違いが吸収され、実機でもシミュレーションと同様な行動軌跡を容易に実現できることがわかる。この事前選択において選ばれたルール群を拘束条件として用いて実機上で強化学習を行なった結果を Fig. 16 に示す。

一般に、シミュレータ上で学習により獲得した制御器が、実環境でもそのまま有効に機能することは難しく、一方、実環境でゼロから学習を行なうアプローチも無駄な試行錯誤が多く現実的ではない。本研究で示すように、シミュレータ上で得られた制御器から局所的な拘束ルール群を抽出し、実環境での有効性を評価しながら選択的に適用する状態拘束型強化学習のアプローチは、シミュレーションで得られた知識を実世界で有効に活用する一手法として有用であると考えられる。

5. ま と め

本研究では、高次元入出力を有するロボットの制御器を強化学習を用いて構築する際に問題となる、無駄な試行錯誤を低減するための一手法として、学習後の制御器から拘束条件をルールとして抽出し、それをさらに複雑なタスクを学習する際の拘束条件として利用することで学習に要するコスト

Fig. 12 Activation of hidden units. (EnvironmentA)

Fig. 15 A trajectory controlled according only to the schema extracted in simulations

Fig. 16 A trajectory of robot learned with the extracted schema

を低減する、状態拘束型強化学習法を提案した。ここで抽出されたルール群はすべてが常に有効に働くとは限らないため、本研究では各ルールの有効性を検証する事前選択操作を導入することで、タスクおよび環境との相互作用の結果として自律的に拘束条件を選択する方策を採用した。

タスク達成に求められる行動の複雑度が段階的に増加する移動ロボットのナビゲーション学習課題において、拘束条件を自律的に抽出・選択し、学習に適用することで、より複雑な課題を効率的に学習できることを確認した。

また、シミュレーション上で抽出された拘束ルール群を実

ロボットの学習に適用した結果から、同手法は実環境下で強化学習を行なう際にも拘束条件として有効に機能することが示された。通常、シミュレーションで得られた制御器はそのまま実環境で利用することはできないが、学習の拘束条件として利用することで、シミュレーションから実環境への知識の流用が可能であることを示した。

謝辞 本研究の一部は文科省科研費(課題番号:14350227および14750362)によるものである。ここに謝意を表わす。

参考文献

- 1) C.W.Anderson: Learning to Control an Inverted Pendulum Using Neural Networks, IEEE Control System Magazine, **9**, 31/37 (1989)
- 2) M.Asada, S.Noda, S.Tawaratumida and K.Hosoda: Purposive Behavior Acquisition for a Real Robot by Vision-Based Reinforcement Learning, Machine Learning, **23**, 279/303 (1996)
- 3) 浅間, 太田, 石黒, 矢野, 土屋, 伊藤, 湯浅: 移動知発現のシステム原理とその工学的実現に関する研究, 計測自動制御学会システム・情報部門学術講演会 (SSI-2002) 講演論文集, 295/298 (2002)
- 4) S. Nolfi and D. Floreano: Evolutionary Robotics — The Biology, Intelligence, and Technology of Self-Organizing Machines —, A Bradford Book, MIT Press, Cambridge, MA (2000)
- 5) C.G.Atkeson, J.Hare, F.Pollick, M.Riley, S.Kotosaka, S.Schaal, T.Shibata, G.Tevatia, S.Vijayakumar, A.Ude and M.Kawato: Using humanoid robots to study human behavior, IEEE Intelligent Systems: Special Issue on Humanoid Robotics, **15**, 46/56 (2000)
- 6) R.S.Sutton and A.G.Barto: Reinforcement Learning, A Bradford Book, The MIT Press, Cambridge, MA (1998)
- 7) M.W.Craven and J.W.Shavlik: Using sampling queries to extract rules from trained neural networks, Proc. of the Eleventh International Machine Learning Conference, 37/45 (1994)
- 8) L.M.Fu: Rule generation from neural networks, IEEE Trans. on System, Man, and Cybernetics, **24**-8, 1114/1124 (1994)
- 9) 井谷, 古橋: 自律移動ロボットによる教示情報の理解, 計測自動制御学会論文集, **38**-11, 966/973 (2002)
- 10) M.Ishikawa: Structural learning with forgetting, Neural Networks, **9**-3, 509/521 (1996)
- 11) 井澤, 近藤, 伊藤: 状態予測機構を用いた強化学習による運動学習モデル, 計測自動制御学会論文集, **39**-7, 679/687 (2003)
- 12) T.Omata: Learning with Assistance based on Evolutionary

Computation, Proc. of the IEEE International Conference on Robotics and Automation 2180/2185 (1998)

- 13) T.S.Peterson, N.E.Owens and J.L.Carroll: Towards Automatic Shaping in Robot Navigation, Proc. of the IEEE International Conference on Robotics and Automation (2001)
- 14) R.Pfeifer and C.Scheier: Understanding Intelligence, A Bradford Book, The MIT Press, Cambridge, MA (1999)
- 15) G.G.Towell and J.W.Shalvik: Extracting refined rules from knowledge-based neural networks, Machine Learning, **13**, 71/101 (1993)
- 16) 日本認知科学学会編: 認知科学辞典, 共立出版 (2002)
- 17) 西岡, 柴田, 伊藤: モデル型 TD 強化学習による動的環境での行動獲得, 第 11 回自律分散システム・シンポジウム資料, 285/288 (1999)
- 18) 真下, 宮野: 発見科学とデータマイニング, 共立出版, 24/33 (2001)

=====

[著 者 紹 介]

近 藤 敏 之 (正会員)

1999 年名古屋大学大学院工学研究科計算理工学専攻博士課程後期課程修了。博士 (工学)。同年, カリフォルニア大学サンディエゴ校客員研究員 (日本学術振興会特別研究員)。2000 年東京工業大学大学院総合理工学研究科知能システム科学専攻助手, 現在に至る。複雑適応システムのロボット工学への適用に関する研究に従事。計測自動制御学会システム・情報部門学術講演会 2002 優秀論文賞受賞。日本ロボット学会, 電子情報通信学会, IEEE などの会員。

伊 藤 紀 彦

2003 年東京工業大学大学院総合理工学研究科知能システム科学専攻修士課程修了。自律移動ロボットの行動学習に関する研究に従事。現在, 日本アイ・ビー・エム (株) 勤務。

伊 藤 宏 司 (正会員)

1969 年名古屋大学大学院工学研究科修士課程応用物理学専攻修了。70 年同大学工学部自動制御研究施設助手。79 年広島大学工学部第 2 類 (電気系) 助教授, 92 年豊橋技術科学大学情報工学系教授を経て, 96 年東京工業大学大学院総合理工学研究科知能システム科学専攻教授, 現在に至る。運動・行動の学習と制御, 自律ロボット, 人間共存ロボット, 義肢制御に関する研究に従事。日本ロボット学会, 電子情報通信学会, IEEE などの会員。工学博士

=====