ニューラルネットを用いた強化学習
－センサからモータまでの合目的的・調和的学習－

Intelligence Based on Reinforcement Learning with Neural Networks
- Purposive and Harmonious Learning of the Whole Process from Sensors to Motors -

柴田 克成[1]　岡部 洋一[2]　伊藤 宏司[3]
Katsunari SHIBATA[1]　Yoichi OKABE[2]　Koji ITO[3]
[1]大分大学　[2]東京大学　[3]東京工業大学
[1]Oita Univ.　[2]The Univ. of Tokyo　[3]Tokyo Inst. of Tech.

Abstract: There still exists a big gap between the present robots and our living creatures with regard to the intelligence. Our living creatures have an advantage of autonomous, harmonious and purposive learning for the non-modularized process from sensors to motors. In this viewpoint, it is expected that reinforcement learning with neural networks breaks through the gap. It has been proposed that raw sensory signals are put into a neural network directly, the output signals of the network are utilized directly as motion commands, and the network is trained based on reinforcement learning. In this learning, the adaptive space reconstruction ability due to the local representation of each sensor cell makes a big role. Some simulation results are introduced to show the effectiveness of this idea.
Keywords: reinforcement learning, neural network, space reconstruction, direct-vision-based reinforcement learning, intelligence

1. はじめに

近年，ヒューマノイドやペットロボットが華々しく注目を浴び，知能ロボットの研究も新らたな段階に入ったと言える．その滑らかで人間や生物に近い動きを見ていると，そのロボットが知的な存在に見えてくる．では，純粋に知能，つまり賢さという点についてはどうであろうか．最近のロボットは，言葉を覚えたり，感情を表現したりと確実に進歩しているように見える．しかし，彼らが言葉の持つ意味を理解しているとは考えられないし，実際に感情を持っているとも考えられない．

筆者らは，この知能における人間や生物とロボットとのギャップを埋めるためには，ニューラルネットを用いた強化学習が大きな鍵となると考えている．本稿では，人間とロボットの違いを改めて考察するとともに，強化学習とニューラルネットの組合わせが，ロボットを大きく人間に近づけると考える理由を説明する．そして，関連したいくつかのシミュレーション結果を紹介する．

2. 人間とロボットの違い

2.1 自律的学習と強化学習

人間は，自ら行動しながら学習し，知識を蓄えていく．一方，ロボットは，設計者としての人間が書いたプログラムにしたがって行動する．したがって，一見，ロボットは賢く見えるが，賢いのはプログラムを書いた人間であり，ロボットはそれを忠実に実行しているに過ぎない．また，人間が書いたプログラミングによる知識付与には限界がある．決まったことしかしない，融通が効かないなど，現在のコンピュータを始めとする機械に対して感じることと同じことをわれわれはロボットに対して感じることになるであろう．このようなことから，現在の知能化の延長線上に生物のような知能は存在せず，いずれどこかで，人間による「知識の付与」からロボット自身による「知識の獲得」へと質的な方向転換を迫られることになると筆者らは考えている．

確かに，ロボットに学習を取り入れるという動きはあるものの，従来の教師あり学習の範疇では，何を教師とすべきかを人間が予め把握し，ロボットに教えてやらなければならない．しかし，これらの情報を与えれば与えるほど，ロボットの柔軟性は失われていく．このようなことから，やはり，ロボット自身が自ら学習していく自律的学習が人間に近づくためには必要であると考えられる．

ここ数年来，強化学習が注目を集めつつある．強化学習は，アクチュエータを用いて行動を起こし，センサ信号からその結果を知るというフィードバックループによって，現在の自分の状態評価とともに，その評価を用いて，行なった行動の良し悪しも判断することができる．また，外部から与えられる情報が少ないために，学習のための時間はかかるが，その分拘束が少なく，自分の状態の評価まで学習によって獲得するため，非常に柔軟に学習ができる．そして，時には人間が予想していなかったような行動までも学習によって獲得することができる．

ところが，マルコフ決定過程を前提とした強化学習では，単に，状態から行動へのテーブルを学習によって獲得するだけであり，テーブルのサイズを単に大きくするだけで知能の形成につながるとは考えにくい．

2.2 非モジュール化とニューラルネット

ロボットの開発と言えば，従来，画像認識，音声認識，記憶，プランニング，制御，コミュニケーションなどの各機能モジュールごとに高機能化を図り，それを組み合わせることで，高機能なロボットを作るというアプローチであった．一方，人間や生物についても，脳の構造は視覚野，連合野，運動野のように，構造化されていると一般的に言われている．しかし，少なくとも大脳皮質の構造は，各領野の働きが違う割りに非常に均一であるし，領野間には溝(しわ)があるが，断面図を見ると，領野間と領野内の結合に大きな違いがあるようには見えない．また，各領野の境界に位置するニューロンは，両方の領野に絡んだ働きをしていたり，連合野と呼ばれる部分は，さまざまな刺激に反応し，その働きを明確に示すことは難しい．さらに，各機能は非常に柔軟であり，合目的的である．例えば，視覚情報の記憶を考えてみると，いつも見ている光景でも覚えていないところが多い一方，分かれ道などの必要な情報は良く覚えているということは，われわれ自身良く経験するところである．

機能ごとにモジュールに分割してしまうと，予めそのインターフェイスをある程度設定しなければならないため，柔軟性が大きく損なわれる．また，全体の目的に沿って学習するために，各モジュールがどのように学習すべきかを知ることは容易ではない．さらに，機能モジュールごとに学習しながら，モジュール間で調和を取ることも難しい．このような問題点を考えると，やはり，モジュール分割をしないで，全体を調和的に学習させるべきであると考えられる．

ニューラルネットは，連続値非線形関数近似を学習によって獲得でき，汎化能力によって，近い入力に対しては近い出力を出すことができる．また，EX-OR 問題を解かせた場合の中間層の働きからわかるように，単に教師信号を与えるだけで，中間層の役割分担が自律的に行なわれるという非常に優れた能力を持っている．この能力を利用すれば，予めモジュール分割をすることなく，必要に応じて，自律的，適応的に役割分担が行なわれるようになるのではないかと期待できる．さらに，リカレント型のニューラルネットを用いれば，記憶やダイナミクスを扱うことも可能である．

2.3 ニューラルネットを用いた強化学習

前述のように，自律的学習という面からは強化学習が，モジュール化しない調和的学習のためにはニューラルネットが有効である．そこで，センサからモータまでニューラルネットで構成し，強化学習に基づいて，ニューラルネットの教師信号を生成して学習させる．こうすると，ニューラルネットには常に教師信号が与えられるが，外からロボットを見ると，与えられる信号はあくまでも強化信号のみであるため，拘束が少なく，非常に柔軟な学習ができる．また，強化学習の側から見ると，単にプランニングのためのテーブルの学習から，認識や記憶，制御まで含めた幅広い学習へと飛躍することにつながる．このように，両者を組み合わせることで，お互いの欠点を補い合い，幅広い機能を，自律的，適応的かつ合目的的に学習できるようになることが期待される．こうした考え方を基に，センサの信号をできる限りそのままニューラルネットに入力し，同様に，出力をモータ（アクチュエータ）にできる限り直結するというアプローチを Direct-Vision-Based 強化学習と呼ぶ [1].

一方，視覚信号の前処理などまでニューラルネットで行なわせるのは非常に効率が悪いとの考え方もある．しかし，前処理をあらかじめ与えると，その柔軟性は失われる上，必ずしも目的に沿ったものであるとは限らない．われわれ生体を見ても，マウスのひげを抜くと，対応した神経細胞は退化し，まわりのひげに対応した細胞集団が大きくなったり [2]，縦じましか見えない環境で育てたネコは，横しまに反応する神経細胞の数が減少する [3] などの現象が見られる．このように，いわゆる前処理と考えられる範疇であっても，我々生物は非常に柔軟であることがわかる．また，大脳において，視覚処理をしているであろう視覚野などの領域の構造は，連合野や運動野の構造と似ているし，次章で述べるように，視覚センサ信号は局所的な情報表現をし，それが中間層での柔軟な空間の再構成を可能にするという重要な能力を持っている．

とは言っても，前処理部を学習させるためには，非常に時間がかかるのは事実である．これに対しては，ニューラルネットの結合の重み値の初期値として，たとえば遺伝といった形で予め情報を与えることによって，過去の学習の結果をある程度活かしていくことが必要であると考えている．この方法であれば，学習の加速はされるが，たとえば環境の変化が起れば，さらなる学習によって重み値は更新されるため，拘束にはならないと考えられる．

3. センサ信号の局所性と大域表現の獲得

3.1 センサ信号の局所性とその利点

一般的に，網膜上の視細胞を始めとし，センサを構成する各センサセルは，受容野というものを持ち，空間上の局所的な情報のみを表現している．一方，一般的に，前処理を行なった後のデータは，例えば，視覚センサ上に見える物体が，センサ全体のどこにあり，どれぐらいの大きさであるかというように，連続値として大域的に表現している場合が多い．このことから，前処理の一つの目的は，個々の局所的なセンサ信号を統合し，大域的な空間情報に変換することであると考えることができる．われわれ人間から見れば，局所的な情報よりも大域的な情

報の方が分かりやすく，また，その統合の仕方はあまり変化の余地がないと考えがちである．

ニューラルネットは非線形関数近似が可能であるものの，一般的に用いられているシグモイド関数は，滑らかで単調増加である．したがって，ステップ関数のように非線形性の強い関数を近似しようとすると，大きな結合重み値を必要とし，学習に時間がかかったり，逆に学習が不安定になったりする．Boyan らは，Hill-car 問題という例題について，シグモイド関数型のニューラルネットでは学習がうまく進まないことを示している [4]．一方，Suttonは，CMAC を用いることで，同じ問題の学習がうまくできることを示している [5]．CMAC は連続値入力を局所化しており，テーブルルックアップに近い形となるため，その後は単に重み付けして足す，つまり，線形の計算ですむようになるからであると考えられる．そして，視覚センサ信号やガウシアンベースの RBF(Radial Basis Function) も，CMAC と同じような働きをしていると考えることができる [1]．

3.2 大域的表現の必要性と適応的獲得

しかし，テーブルルックアップでは，より良い近似を行なうために，状態を細かく分割すればするほど，汎化能力が劣るというジレンマがある．例えば，物体が右側に見えた場合には右に回転し，左に見えた場合は左に回転することが要求されたとし，それをある見え方のときに学習したとする．しかし，少し位置がずれれば，別のセンサセルが反応するようになるため，入力が変化し，一から学習しなければならない．したがって，全体として多くの学習回数が必要となる．しかし，大域的な表現の上では，汎化能力が有効に働くため，何回かの学習だけで，右に見えたら右に，左に見えたり左に回転することを学習できる．ただし，逆に，境界部分では，不連続な出力を求められるので学習は困難である．

この場合，右に見える場合と左に見える場合とでそれぞれを一つの状態として一くくりにすることが最も効率的である．最初からそれが分かっていれば良いが，分からない場合は，学習を通してそういう表現を獲得していくことが望まれる．

ニューラルネットには，入力パターン間の距離が大きくても，与えられる教師信号が近いと，中間層の表現も近くなるという性質がある [6]．これによって，局所信号が入力として与えられても，中間層において必要に応じて統合され，大域的な表現を獲得できるようになる．たとえば，上記の例の場合は，物体が右側に見える場合は，求められる動作は常に同じなので，その結果，見え方によらず中間層の表現も近いものとなる．ところが，左側に見える場合は別の動作が求められるので，別の表現となる．この際，前述のように，局所信号から大域的な表現への変換は，線形に近いものとなるため，右と左の境界部分での学習も容易となる．そして，いったんこの中間層の表現が学習されると，そのような表現が有効な別の学習を行なう際には，その中間層での大域的な表現の上での学習ができるようになり，大幅な学習速度の向上につながると考えられる．われわれが存在する三次元空間の認識は，あらゆるタスクにおいて必要になるため，いくつかのタスクの学習をするうちに脳の中で三次元空間が再構成され，その後のタスクでの学習は大幅に加速されると考えられる．

脳内で局所的な無数の視覚センサ信号を統合し，三次元空間を再構成できる仕組みについてもう少し具体的に考えてみる．ここで，強化学習の状態評価における「空間から時間へのマッピング」というものが一つのキーワードとなる [7]．強化学習では，報酬が得られるまで，状態の評価値は時間とともになめらかに単調増加するように学習される．Q-learning を含めた TD(Temporal Difference) 型の学習 [8] では指数関数的に，筆者らの一部が提案した TS(Temporal Smoothing) 型の学習 [7] では直線的に増加する．これは，言い換えると，時間的に近い出力に対しては教師信号の値も近くなる．また，時間的に近い入力は，空間的にも近い．したがって，軌道に沿っては，空間的に近い状態に対応する入力は，入力パターンでは距離が離れていても，中間層のパターンでは距離が近くなる．一方，軌道と垂直方向に対しては，行うべき行動が滑らかに変化すると考えられる．このようにして，極座標に近い形で三次元空間が脳内に再構成されると考えられる．これは，SOM(Self Organizing Map)[9] が隣同士のユニットの参照ベクトルが互いに引き合ってトポロジーを保持する仕組みと似ている．

4. 目標物獲得タスク

4.1 タスクの設定と学習結果

ここでは，Fig. 1のような簡単な視覚センサを持つロボットが移動して目標物を捕獲するというタスクを学習させたシミュレーションの結果を紹介する [1]．このロボッ

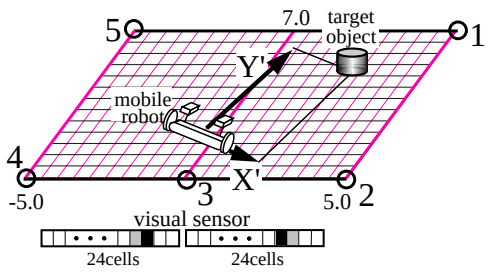

Fig. 1 A task in which a mobile robot with two visual sensors captures a target object

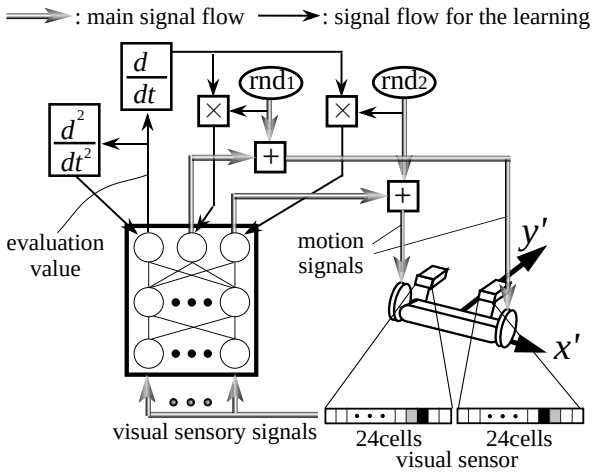

Fig. 2 Architecture and signal flow in the robot.

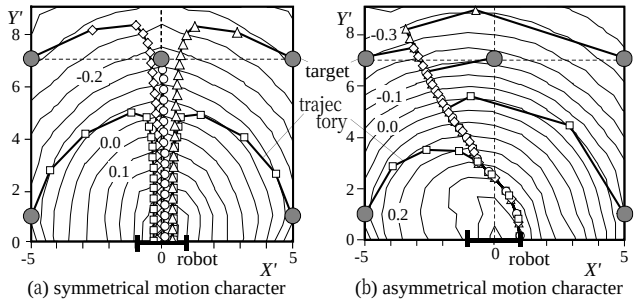

Fig. 3 Difference of the state value function and the robot trajectories on robot-fixed coordinates depending on its motion character.

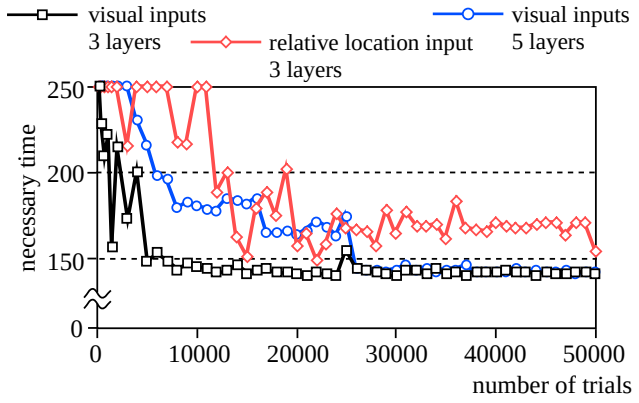

Fig. 4 Comparison of the learning curve depending on the input way and the network architecture.

トは，両脇の車輪を独立に回転させて前後・回転運動ができ，車輪上に２つの視覚センサを持つ．各視覚センサは，24 個のセンサセルよりなり，各セルはオーバーラップのない放射状の局所的な受容野を持ち，全体として 180 度の視野を有する．そして，Fig. 2のように，センサからの信号計 48 個をニューラルネットへ入力し，評価の出力 1 個，動作用の出力 2 個の合計 3 個の出力を用意し，2 つの動作用出力に従って両脇の車輪を動かした．また，ニューラルネットは 3 層，中間層ニューロンは 20 個とし，ニューロンの出力関数は，-0.5 から 0.5 のシグモイドとした．目標物は Fig. 1中の平面内にランダムに置き，目標物の中心がロボットを通り抜けた時にロボットが目標物を獲得したとし，報酬として 0.4 の教師信号で評価値を学習し，それ以外で目標物が自分より後ろにいって視野から消えた場合は，罰として-0.4 の教師信号を与えて学習した．そして，目標物に到達するかまたは見失った場合を１試行とし，再びロボットを元の位置に戻し，目標物の位置を乱数によって決定し，動作と学習を開始した．ただし，学習の初期には，目標物の初期位置をロボットの近くのみとし，学習の進行と共に徐々に範囲を広げた．状態評価の学習は TS 型とし，ここでは，評価値の時間に対する傾きの理想値を常に一定値とした．

Fig. 3(a) は，学習後の目標物の位置に対する状態評価関数と 5ヵ所に目標物を置いた場合のロボットの軌跡をロボットの位置を固定した座標で示したものである．学習を始めた当初は，各センサセルの受容野が放射状に広がることが影響し，評価関数曲面も放射状の尾根や谷ができていた．しかし，学習の進行によって，図のように，滑らかな状態評価値が獲得され，回転して前方に目標物が見えるようになってから前進するという経路で目標物に到達できるようになった．

次に，ロボットの右の車輪を該当出力の３倍，左の車輪は１倍して回転させるという非対称な動作特性を導入し，対称な場合との比較を行った．Fig. 3 (b) にその結果を示す．この図より，評価関数が左右で少し非対称性であることがわかる．また，ロボットの経路を見ると，目標物が右側にある場合でも，左前方に見えるように回転してから前進し，目標物が近づいてくると，ロボットはそれを右側で捕らえている．さらに，絶対座標でロボットの経路を見ると，右の真横に物体が見えると，いったん大きく後ずさりすることが観察された．これらはいずれも，右の車輪を速く回転できる場合に有利な行動であるが，設計者自身前もって予測できなかった行動もあり，強化学習による機能創発を表していると言える．

また，この時，視覚センサ信号を直接入力した場合と，物体の前後，左右の相対位置を 2 つの連続値で大域的に表現して入力した場合，および，後述の視覚センサ信号を入力した 5 層のネットで，真ん中の中間層を 2 個にしぼったものとで学習の様子を比較したものを Fig. 4に示す．学習係数は，いくつかのものを適用し，最適のものを選んだ．これより，視覚センサ信号を入力したものは，

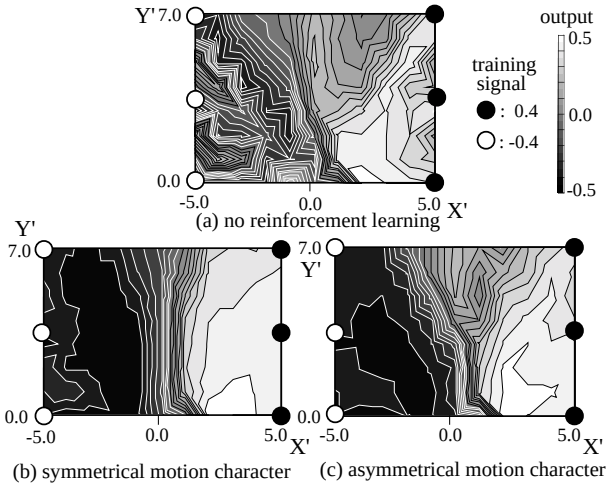

Fig. 5 Comparison of the object location coding in the hidden layer.

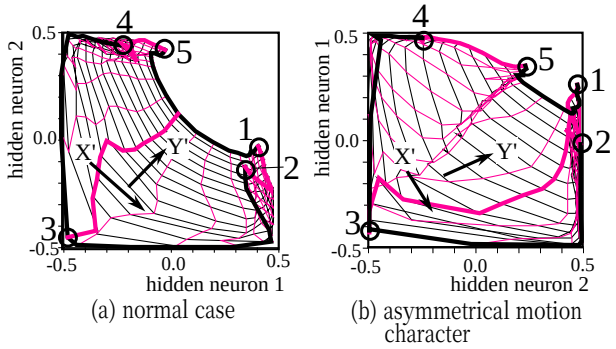

Fig. 6 Comparison of the object location coding in the hidden layer of 5-layered network.

学習が速く，安定しており，相対位置を入力したものは，誤差が他と比較してあまり減少しないことがわかる．また，5 層のものは，学習が遅いものの，最終的には誤差が小さくなっていることがわかる．5 層のネットの上の 3 層と相対位置入力のネットワークは構造が同じであることから，単にニューロン数が 2 つということだけでなく，その表現法が学習に影響することがわかる．

4.2 中間層での空間情報のコーディング

ここでは，中間層がどのようなコーディングをしているかを調べた．まず，前節で学習した 3 層のニューラルネットを用い，出力層に中間層からの重み値をすべて 0 にしたニューロンを付加し，Fig. 5に小さい丸で示した物体の位置 6ヵ所のみについて右側に見えたら 0.4，左側に見えたら-0.4 で教師あり学習をさせ，物体の位置に対する出力の分布がどうなるかを調べた．強化学習を行う前のネットワークを用いたときには，出力は放射状に広がり，大きな尾根と谷ができていることがわかる．それに対して，強化学習を行った後では，右と左をきれいに分離していることがわかる．さらに，動作特性を左右非対称にした場合は，正面より少し左に向かって評価関数の壁ができていることがわかる．これは，Fig. 3(b) の右回転するか，左回転するかの動作の境界部分に相当すると考えられる．

次に，前述のように，真ん中の中間層ニューロン数を 2 つにしぼった 5 層ニューラルネットで学習を行った．そして，強化学習を行い，2 つの中間層ニューロンが物体の位置をどのようにコーディングしているか調べた．各層のニューロン数は，入力側より，48-20-2-10-3 とした．Fig. 6に，中間層の 2 つのニューロンの値を縦軸と横軸にとり，Fig. 1中の格子上に物体を置いた時の中間層の値をプロットし，格子を作成したものを示す．学習前は入力層-中間層間の結合の重み値が微少な乱数で決定されているため，2 つの中間層ニューロンとも真ん中の 0.0 近辺の値をとった．しかし，学習が進むにつれ，使用している中間層の値の領域が大きくなり，100000 試行後には，Fig. 6(a) のように，中間層ニューロンのかなりの領域を使って物体の位置を表現していることがわかる．また，図中の数字付の○印は，Fig. 1中の○印に物体を置いた場合に相当する．これより，物体が車輪近くに存在する状態を拡大して表現していることがわかる．これは，この状態が，物体を取り込めるかどうかの境界であるため，ロボットにとって重要，つまり，より細かく見る必要がある領域であることを反映していると考えられる．

次に，非対称動作特性を導入した場合について調べた．すると，Fig. 6(b) のように，ロボットの正面を表す線（図中の○3から伸ている薄く太い線）が大きく曲がっていることがわかる．そして，ロボットの近くではロボットの右側を，遠くでは左側を，つまり，ロボットの経路周辺を拡大して表現していることがわかった．これは，ロボットが右に回転するか，左に回転するか，前進するかの境界部分である．このように，周りの環境は全く同じでも，ロボットの動作特性を変化させるだけで，重要なところを重点的に，そして適応的に表現する能力があることがわかる．

5. その他の機能の学習

上記で示した例題の他に，**センサを動かす能動認識機能**を学習させたり [10]，手先のリーチング動作の学習から **hand-eye coordination** を獲得できることを示した [11]．また，リカレントニューラルネットを使って，エージェント間で衝突回避のため**交渉のコミュニケーション**を学習させたり [12]，文脈に基づく選択的注意の課題を学習させて，必要な**文脈を抽出**し，**選択的注意**が行えること，さらにその記憶に**連想記憶**の能力があることを確

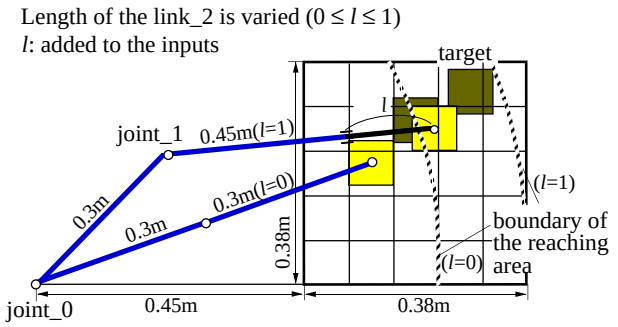

Fig. 7 Learning of arm reaching movement to an object on the visual sensor with variable link length.

認した [13]．また，ニューラルネットは用いていないが，複数のエージェントに強化学習で電車の乗り降り問題を学習させることで，エージェント間に**個性や社会性が発現**し，それが環境などの要因によって適応的にどちらかに分化すること [14]，さらに，もらった報酬の一部を，状況に応じて他のエージェントに分配することを学習することもできた [15]．このように，強化学習によってさまざまな機能が獲得できることがわかった．

ここでは，リーチングの学習 [11] について簡単に紹介する．Fig. 7のように，5×5 の視覚センサ上に，2 関節リンクの手先と目標物が見え，両者は区別できず，第 2 リンクの長さは可変とした．そして，画像情報と，各関節角，角速度，および第 2 リンク長をニューラルネットに入力し，2 つの関節トルクを動作信号として出力する．手先が目標物に接触すると，報酬がもらえ，これをもとに強化学習を行なう．学習によって，手先の滑らかなリーチング動作が獲得された．入來 [16][17] は，サルの頭頂葉で，えさが届く範囲にあるかどうかを表現しているニューロンや，手先にえさが近づくと発火するニューロンの存在を示している．また，サルに道具を持たせると，それらのニューロンは，道具が届く範囲まで発火したり，道具の近くでも発火するようになると報告している．さらに，サルの心の中で道具が手の一部として扱われているとの解釈から，シンボリックシステムおよび「知」との関連性に言及している．そこで，道具を可変リンク長と置き換えて考え，前述の学習したニューラルネットの中間層を観察すると，このサルの前頭葉のニューロンと同様な働きをするニューロンを観察することができた．

6. おわりに

ニューラルネットを用いて強化学習を行なうことで，センサからモータまでのさまざまな機能を，自律的，適応的，合目的的，調和的に獲得させることを提案し，ロボットの知能につながる可能性を示した．

謝辞

本研究の一部は，文部省科学研究費重点領域研究「創発システム」(No. 264)，奨励研究 (A)(No.13780295) および学術振興会未来開拓学術研究推進プロジェクト「生物的適応システム」(JSPS-RFTF96100105) の補助の下で行わた．ここに謝意を表します．

参考文献

[1] 柴田克成, 岡部洋一, 伊藤宏司: ニューラルネットワークを用いた Direct-Vision-Based 強化学習, 計測自動制御学会論文集, **37**(2), pp. 168–177 (2001)
[2] 塚田裕三編: 図説 脳, 日経サイエンス社, pp. 50 (1983)
[3] Blackmore, C., and Cooper, G.F.: Development of the brain depends on visual evnironment, *Nature*, **228**, pp. 477–478 (1970)
[4] J.A. Boyan & A.W. Moore: Generalization in Reinforcement Learning: Safely Approximating the Value Function, *Advances in Neural Information Processing Systems*, MIT Press, **7**, pp. 369–376 (1995)
[5] R. S. Sutton: Generalization in Reinforcement Learning: Successful Examples Using Sparse Coarse Coding, *Advances in Neural Information Processing Systems*, MIT Press, **8**, pp. 1038–1044 (1996)
[6] K. Shibata & K. Ito: Reconstruction of Visual Sensory Space on the Hidden Layer in Layered Neural Network, *Proc. of ICONIP(Int'l Conf. on Neural Information Processing)'98*, **1**, pp. 405–408 (1998)
[7] 柴田克成, 岡部洋一: 時間軸スムージング学習, 電気学会論文誌 C 分冊, **117-C**(9), pp. 1291-1299 (1997)
[8] A. G. Barto, R. S. Sutton & C. W. Anderson: Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems, *IEEE Trans. SMC*, **13**, pp. 835-846 (1983)
[9] T. Kohonen: Self-Organized Formation of Topologically Correct Feature Maps, *Biol. Cybern.*, **43**, pp. 59–69 (1982)
[10] 柴田克成, 西野哲生, 岡部洋一: Actor-Q アーキテクチャに基づく能動認識学習システム, 電子情報通信学会論文誌, **J84-D-II**(9), pp.2121-2130 (2001)
[11] 柴田克成, 杉坂政典, 伊藤宏司: 強化学習によるリーチング運動の獲得, 電子情報通信学会技術研究報告, NC2000-170, pp. 107-114 (2001)
[12] 柴田克成, 伊藤宏司: 利害の衝突回避のための交渉コミュニケーションの学習と個性の発現, 計測自動制御学会論文誌, **35**(11), pp.1346-1354 (1999)
[13] K. Shibata: Formation of Attention and Associative Memory based on Reinforcement Learning, *Proc. of ICCAS01*, pp. 9–12 (2001)
[14] K. Shibata, M. Ueda, and K. Ito: Emergence of Individuality and Sociality by Reinforcement Learning, *Proc. of Fifth of Int'l Sympo. on Artificial Life and Robotics (AROB)2000*, **2**, pp. 589 – 592 (2000)
[15] K. Shibata and K. Ito: Autonomous Learning of Reward Distribution for Each Agent in Multi-Agent Reinforcement Learning, *Intelligent Autonomous Systems*, Vol. 6, pp. 495–502 (2000)
[16] 入來篤史: サルの道具使用と身体像, 神経進歩, **42** (1), pp. 98–105 (1998)
[17] 入來篤史: 道具を使う手と脳の働き, 日本ロボット学会誌, **18** (6), pp.786–791 (2000)