強化学習に基づく能動認識
Active Perception Based on Reinforcement Learning

柴田克成 (PY)，西野哲生，岡部洋一（東京大学先端科学技術研究センター）
〒153 東京都目黒区駒場４－６－１
Katsunari SHIBATA (PY), Tetsuo Nishino, Yoichi Okabe
Research Center for Advanced Science and Technology (RCAST), Univ. of Tokyo
4-6-1 Komaba Meguro-ku Tokyo 153 JAPAN
shibata@okabe.rcast.u-tokyo.ac.jp

abstract
The sensory motions can be thought as one of the motions to achieve a purpose. We propose a structure with a neural network and learning method to make a system learn appropriate motions for the effective perception based on the reinforcement learning by the reinforcement sigal made from the perception result.

１．はじめに

我々生物は，様々なセンサを使って外界の情報を取り込み，それに基づいて適切な行動を行なうことができる．また，外界の情報を効率的に取り込むために，センサ自身を動かし，より良い認識ができる位置にセンサを動かす，いわゆる能動認識[1]の機能も有する．センサを動かして，効率的な認識を行ない，外界の状態をより正しく認識することができれば，より適切な動作を行ない，より効率的に目的（本能）を達成することができる．つまり，生物にとっては，センサを動かすという動作も，食物を食べるために食物に近づくといった動作と同様に，本能を満たすために必要な一連の動作の一つであると考えることができる．

一方，システムの出力に対し，教師信号を与えるのではなく，その出力が良いか悪いかを教えることによって，そのシステムの入出力関係を学習させる強化学習 (Reinforcement Learning) というものが提案されている．この強化学習は主として，機械（ロボット）が目的を達成する時の動作の学習に使われてきた[2][3]．また，最近は，強化学習にニューラルネットを適用することによって，その学習，汎化能力の有効性が確認されている[3][4]．前述のように，能動認識におけるセンサの動作も，一連の目的達成動作と考えることができる．そこで，我々は，ニューラルネットを用いた強化学習で能動認識を学習させることを考えた．強化信号は，通常，動作から時間が経過してから得られる，いわゆる遅延強化信号となるため複雑な学習を要する．しかしここでは，簡単のため，認識結果に対して逐次強化信号が得られるという仮定の下で，その強化信号からセンサの動作と物体の認識を同時にニューラルネットを用いて学習させるためのシステムの構成と学習方法を提案する．

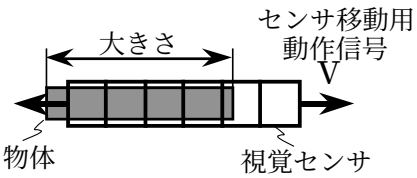

図１ 仮定した視覚センサとその動き

２．システムの構成

ここでは，図１のように，６つの受容野を持つ視覚センサが物体を見ており，その視覚センサが１次元の動きを行ないながらその物体の認識を行なうという例を基に説明を行なう．システムの全体構成を図２に示す．ニューラルネットは，階層構造になっており，視覚センサから入力を得て，出力を計算する．ニューラルネットの出力は，センサの動作用の信号１個と，認識用の出力３個からなる．各出力には，乱数付加部で，微小な乱数を付加する．センサは，センサ動作用のニューラルネットの出力に乱数を付加した値に従って左右に１次元の動作を行なう．一方，認識用の出力は，乱数を付加した後，強化信号生成部に渡され，その出力が理想の出力にいかに近いかを示す強化信号が生成される．そして，強化信号の時間変化（前の時間の強化信号との差）とニューラルネットの出力に加えられた各乱数との積をとることによって，加えた乱数を評価し，より強化信号が大きくなるような出力値を学習する．

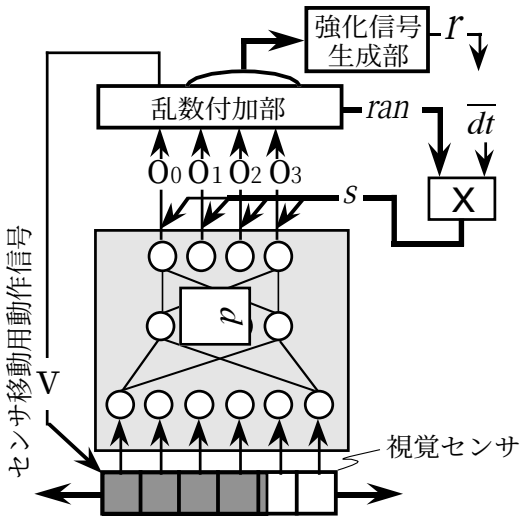

図２ システムの全体構成
（太線はベクトル）

Keywords : Active Perception, Reinforcement Learning, Neural Network
キーワード：能動認識，強化学習，ニューラルネットワーク

３．学習アルゴリズム

センサの動作の学習と，認識の学習は，同時に行ない，かつ，認識結果を強化信号に用いる以外は，全く同じ学習アルゴリズムを適用する．つまり，認識に対する学習も，直接教師信号を与えるのではなく，強化信号を基に学習を行なう．強化信号 r は，基本的には以下のように作成する．

$$r = -\sum_{i=1}^{3} (o_i - t_i)^2 \qquad (1)$$

ここで，o_i はニューラルネットの各出力（i=0 はセンサ移動用の動作信号，i=1,2,3 は認識結果），t_i は各認識用出力に対する理想出力を表わす．ニューラルネットの各出力には，学習のために乱数付加部で乱数を加えるので，実際は，各出力に加える乱数をrnd_i とすると，

$$r = -\sum_{i=1}^{3} \{(o_i + rnd_i) - t_i\}^2 \qquad (2)$$

また，センサに対する動作信号（速度） v は，

$$v = o_0 + rnd_0 \qquad (3)$$

とし，微小な乱数成分を含んだ動作を行なう．また，学習は，ニューラルネットの出力に対し，その後加えられた乱数と強化信号の変化量との積を加えたものを教師信号としてニューラルネットに与え，教師あり学習（バックプロパゲーション法）によってニューラルネットを学習させる．

$$s_i = o_i + rnd_i \cdot r \quad (i = 0,1,2,3) \qquad (4)$$

これによって，物体認識を学習しながら，認識しやすい位置へのセンサの動作をも同時に学習できる．

４．シミュレーションとその考察

図１の環境で，大中小（それぞれ視覚センサの受容野６個分，４個分，２個分に相当）３種類の物体の大きさの認識を学習させた．物体が視覚センサの端に位置する時は，その物体の大きさは認識できないため，正しい認識のためには，物体をある程度視野の真ん中へ持ってくるという動作が必要となる．認識の理想出力としては，大中小それぞれに対し，（１,０,０）（０,１,０）（０,０,１）とした．ニューラルネットは，６－３－４の３層とした．また，視覚センサは，各受容野の中に占める物体の割合を出力するものとした．学習は，まず物体と視覚センサの相対位置は，視覚センサから物体がある程度見える範囲内で，乱数を用いて決定し，物体が物体を見せてから６０単位時間または６個のセンサ出力の合計が１より小さくなるまで続け，その後次のパターンを見せるという方法で繰り返し行なった．

その結果，いずれの大きさの物体を見せた時も，学習によって，物体を中心に捉えるようにセンサが動作し，その中心において正しく認識ができるようになった．我々が物体を認識する時も，視野の中心に物体を捉えるが，こうした学習によって説明が付くのではないかと考えられる．また，われわれの目の網膜上の神経細胞の密度が中心ほど大きいということは，物体を中心で捉えることが非常に効率的であり，その中心部分に神経細胞を高い密度で配置することが有効であるということと関連しているものと思われる．また，中間層ニューロンは３個で学習がうまくいった．これは，動作しなければ区別不可能なパターンがあったり，また，そのようなパターンを除いても，中間層ニューロン３個では，全ての見え方に対して正しく認識できるようなニューラルネットを形成することは不可能であると考えられることから，非常に効率的な認識を実現することができたと言える．

５．結論

能動認識におけるセンサの動きを強化学習によって学習させるためのシステムの構造と学習方法を提案した．これによって，簡単な認識の問題をセンサを動かして効率的に学習させることができた．

参考文献

[1] 山崎弘郎, 石川正俊 (編著), "センサフュージョン：実世界の能動的理解と知的再構成", 科学技術庁監修 (1992)

[2] A. G. Barto, R. S. Sutton and C. W. Anderson, "Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problrms", IEEE Trans. SMC-13, Pp835-846 (1983)

[3] K. Shibata and Y. Okabe, "A Robot that Learns an Evaluation Function for Acquiring of Appropriate Motions", Proc. on WCNN '94 San Diego, vol.2, pp. II-29 - II-34 (1994)

[4] R. J. Williams, "Simple Statistical Gradient-Following Algorithm for Connectionist Reinforcement Learning", Machine Learning, 8, pp.229-256 (1992)