SD-2-5　ニューラルネットによる目的達成動作の能動的学習

An Active Learning Machine using Neural Networks to Carry out its Purpose

柴田 克成
Katunari Shibata
㈱日立製作所 中央研究所
Central Research Laboratory Hitachi Ltd.

1．はじめに

計算機と人間における情報の処理方法の取得には、受動的か、能動的かという大きな違いがある。一般の計算機は、ユーザがプログラミングしてやらなければ実行ができないが、人間は外界に自ら働きかけ(動作し)、その変化をみて、より適切な動作を生成するための処理方法を試行錯誤によって学んでいく。この点が人間の柔軟な情報処理を可能にする大きな要因であると考えられる。

能動的な学習を行なうためには、学習、連想能力を持つニューラルネットが有効であり、既に銅谷[1]、依田ら[2]によってそのモデルが提案されている。これは、図1のように状態に対する評価関数を定義し、ニューラルネットを用いることによってこの評価関数の値がより高くなる動作を学習していくものである(以下、評価関数法と呼ぶ)。

本報告では、この評価関数をも経験によって学習することにより、"目的"から目的達成のための"動作"を自動的に生成するモデルを提案する。そして、それを実現するネットワークの構造を示し、簡単なシミュレーションを行ったので報告する。

2．能動的学習と評価関数の自動生成

能動的学習という観点から、各学習方式を比較すると表1のようになる。能動的学習においては、予め与える情報をいかに少なくし、いかに高度な機能が得られるかが大きなポイントとなる。

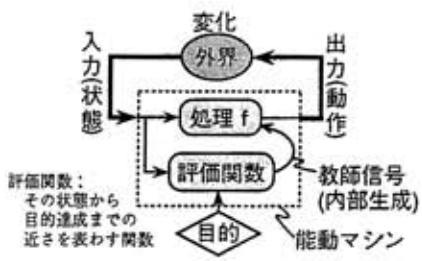

図1 評価関数を用いた能動マシンの構成

表1 学習方式の比較

	計算機	BP法	評価関数法	本研究
処理方法	与える	自動生成	自動生成	自動生成
評価関数	−	−	与える	自動生成
受動／能動	受動	受動	能動	より能動

与える:ユーザが与える　−:概念がない

ニューラルネットの代表的な学習方法であるバックプロパゲーション(BP)法では、処理方法を自動生成できるが、理想の出力である教師信号をいくつかユーザが与える必要がある。また、競合学習(表中では省略)は、入力パターンの出現頻度から学習するため、教師信号なしで学習が可能であるが、目的にあった処理を得ることは難しい。

これに対し、前述の評価関数法では、予めユーザが与えた評価関数によって自らの動作を評価し、教師信号を内部生成させている。これは外界との能動的なループから自ら学習できるという大きな意義を持つが、以下のような問題点も考えられる。

①評価関数は予め決定しなければならない
　→外界の変化への適応及び経験の反映が困難
②評価関数は状態に対して連続な値しかとれない
　→例えば、"食物を食べる"という目的を機械に与えた場合、ある状態において食べられるか否かは2値の離散的な値しかとることができず、上記のモデルへの適用は困難

ここで、この離散的な値をとる"目的"から連続的な値をとる"評価関数"を自ら学習し、生成できれば、上記の問題点を解決できるはずである。次に、そのための具体的方法を示す。

3．ニューラルネットの構成とその学習方法

本報告では、図2のように、外界からの情報を入力して評価関数を出力するニューラルネット(評価関数ネット)と、同じ入力で動作信号を出力するニューラルネット(動作信号ネット)の2つよりなる構成とする。そして、動作信号に乱数成分を加えたものに基づいた試行を繰り返しながら、BP法に基づいて以下の学習を行う。

① 評価関数ネットに対し、目的を達成した時に0.9、初期状態(目的の達成から遠い状態)の時に0.1の教師信号を与える
② ①以外の場合は、評価関数ネットの出力値が時間の経過に対して凹凸なく単調増加するように教師信号を内部生成する
③ 動作信号ネットに対し、評価関数ネットの出力値が、動作によってより高い値に変化するように教師信号を内部生成する

上記②及び③を実現するため、ここでは、それぞれ評価関数ネットの出力を時間で2階微分及び1階微分した値を用い、教師信号を内部生成した

(図2)。また、②の学習を行った後、目的が達成されるまでの過程で、ある状態からある状態により早く到達できる方法があると、その部分は評価関数ネットの出力値の変化がより大きくなる。すると、③によってその方法を学習できるため、ランダムな試行が、徐々に意志を持ち、最短時間経路へと変化していく。

4. 目標物取り込み問題とシミュレーション結果

上記の機能の検証のため、簡単なシミュレーションを行った。ここでは、このニューラルネットを組み込んだロボットを仮定し、目標物を取り込むという目的だけを与えて、目標物に近づくという動作を生成できるかという問題を考えた。

具体的には、図3のように、両脇に車輪を持つ移動ロボットでシミュレーションを行なった。そして、ロボットから目標物を見た時の前方向と横方向の距離を入力し、評価関数及び両輪の回転角を出力として求め、これに乱数成分を足した値に従って動作をする。学習前は、両ニューラルネットの出力は常に一定値で、乱数成分だけから動作が決まるように設定した。また、学習の初期や誤った学習をした時には簡単に目標物に到達できないため、学習の加速手段として、動作中に目標物を近づけてやる。ただし、ロボットには目標物に近づいたことが良いか悪いかの情報は一切与えない。また、ロボットが目標物を取り込むことなく通り過ぎた場合には、0.1の教師信号で評価関数の学習を行い、再び初期状態から学習を開始する。

こうして、前章で示した学習を行いつつ、目的達成までの動作を8000回行わせた。図4に目標物の位置及び学習回数によって、ロボットがどのように動作するかを示した。この図から、学習が進むと目標物に到達できるようになり、さらに、その経路が変化してより早く目標物に到達できるようになっていることがわかる。また、図5にロボットから見た目標物の距離に対する評価関数を示す。これを見ると、評価関数の値は目標物との距離が近いほど高くなっており、ロボットが目標物との距離が近いほど取り込みやすい状態であることを学んでいると言える。

5. 結論及び今後の課題

報告者の提案したニューラルネットの構成で、評価関数をも能動的学習によって取得できることを示した。これにより、"考える機械"の実現に一歩近づいたものと考える。今後は、より複雑な問題の解決のため、中間目標の自動設定、過去の履歴の評価関数への反映、冗長情報(視覚等)の効果的な利用法等が課題と考える。

6. 謝辞

本研究を進めるに当たって、ご指導ご討論頂いた当研究所の方々に感謝致します。

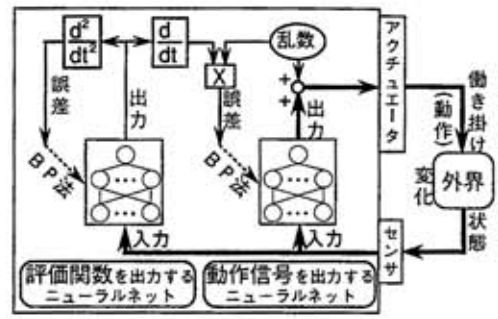

図2 ネットワークの構成と学習方法

図3 目標物を取り込むロボット

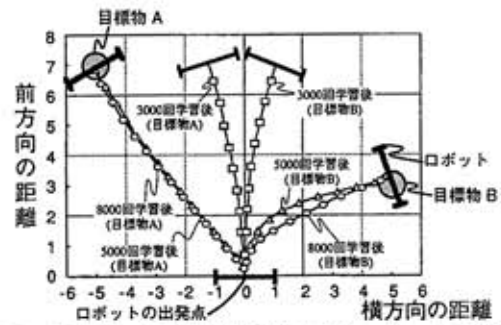

図4 目標物の位置と学習回数によるロボットの経路
(20単位時間毎にプロット)

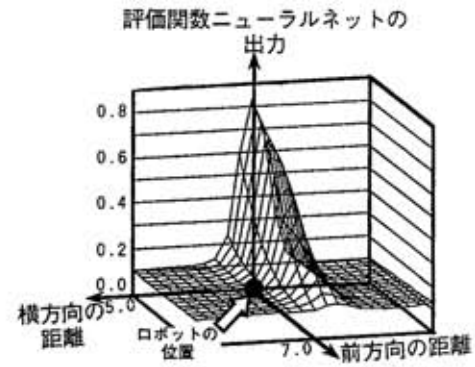

図5 評価関数ニューラルネットの出力
(8000回学習後)

参考文献

[1]銅谷："運動パターンの自己組織化"、第16回SICE学術講演会予稿集、pp.961-964 (1986)

[2]依田、他："本能に基いて運動系列を学習する運動モデル"、信学会論文誌 vol.J73-D-I No.7 (1990)