第２章　基本となる学習則

本章では、本論文で用いる基本的な学習則について述べる。まず、２.１節では本論文中で用いたニューラルネットのモデルと学習アルゴリズムであるバックプロパゲーション（ＢＰ）法について簡単に説明し、様々なニューラルネットモデルがある中で本論文で用いたものを明確にする。２.２節では、このＢＰ法を用いて効率的に強化学習を行う方法を示す。その後、２.３節以降で、筆者が提案している、自律学習において重要な役割を果たすと考えられる学習アルゴリズム１.相関情報抽出学習、２.時間軸スムージング学習、３.値域拡大学習および複数出力の直交化法についてその概要を述べる。各学習アルゴリズムの具体的な使用方法は、第３章以降で述べるのでこちらも参照されたい。

２.１ ニューラルネットの計算とバックプロパゲーション（ＢＰ）法

始めに、本論文で用いたニューラルネットの前向き計算と学習のための後ろ向き計算（ＢＰ法）を簡単に示す。本論文中では、すべて階層型ニューラルネットの離散時間モデルを用いる。まず、前向き計算に関しては、各ニューロンは、

$$u_j = \sum_{i=1}^{n} w_{ji}\, x_i \qquad (2.1)$$

u_j : j 番目のニューロンの内部状態、x_i : １つ下の階層の i 番目のニューロンの出力値
w_{ji} : 下の階層の i 番目のニューロンから j 番目のニューロンへの結合係数（重み値）
n : 下の階層のニューロン数

$$x_j = f(u_j + \theta_j) \qquad (2.2)$$

f : ニューロンの出力関数、θ_j : j 番目のニューロンのバイアス入力

の２式によって計算する。出力関数は、１.２節で述べたように、シグモイド関数を用いるため、

$$f(u_j+\theta_j) = \frac{1}{1 + \exp\{-(u_j+\theta_j)\}} \qquad (2.3)$$

と表わすことができる。シグモイド関数の値域は、（２.３）式では０から１までであるが、これから一様に 0.5 を引いて-0.5 から 0.5 を値域とした場合もあるが、特に大きな差は見られなかった。これを、階層にしたがって順次計算し、ニューラルネットの出力を計算する。そして、出力ニューロ

ンに対し教師信号 s_j が与えられると、誤差 E が

$$E = \frac{1}{2}\sum_{j=m}^{n}(s_j - x_j)^2 \qquad (2.4)$$

と定義される。そして、最急降下法の式、

$$\Delta w_{ji} = -\eta\frac{\partial E}{\partial w_{ji}} \qquad (2.5)$$

を計算していくと、重み値の変化量Δw_{ji} は

$$\Delta w_{ji} = \eta\delta_j x_i \qquad (2.6)$$

となる。ただし、δ_j の値は、出力層では

$$\delta_j = (s_j - x_j)f'(x_j) = (s_j - x_j)(1 - x_j)x_j \qquad (2.7)$$

となり、中間層では、上位の層のδ_j の値から

$$\delta_i = f'(x_j)\sum_{j=1}^{\infty} w_{ji}\delta_j = (1 - x_i)\, x_i \sum_{j=1}^{\infty} w_{ji}\delta_j \qquad (2.8)$$

と計算できる。本学習は、本論文中のほとんど全ての学習に用いる。また、教師信号が０や１に近いと、ニューロンの内部状態の絶対値が非常に大きくなければならず、よって結合の重み値も大きくなければならない。従って、与える教師信号は、0.1 から 0.9 の範囲で与えるようにした。

２.２ バックプロパゲーション法を用いた強化学習

本論文中で強化学習を行う際には、バックプロパゲーション（ＢＰ）法を用いて簡単に学習をさせている。本節では、このＢＰ法を用いて強化学習を実現する方法について述べる。

図２.１に示したように、まず、階層型ニューラルネットに何らかの入力ベクトル $\boldsymbol{x}$ が入り、出力ベクトル $\boldsymbol{y}$ を計算する。この出力に対し、乱数ベクトル $\boldsymbol{rnd}$ が加えられ、その値を評価器で評価し、強化信号（評価値）Φが得られるものとする。そして、この強化信号が大きくなるように出力ベクトル $\boldsymbol{y}$ を学習するものとする。

この時、加えた乱数が良ければ、Φの変化量$\Delta\Phi$はより大きな値になることから、出力ベクトル $\boldsymbol{y}$ に対する教師信号ベクトル $\boldsymbol{s}$ を

$$s = y + k\, rnd\, \Delta\Phi \qquad (2.9)$$

k：学習の定数

とする。すると、 $\Delta\Phi$が微小であるとすれば、

$$\Delta\Phi = rnd\, \nabla\Phi \qquad (2.10)$$

となるため、（２.９）式は

$$s = y + k\, rnd\, rnd\, \nabla\Phi \qquad (2.11)$$

と変形できる。ここで教師信号ベクトルの期待値 $\bar{s}$ を求めると、

$$\bar{s} = y + \kappa\, \nabla\Phi \qquad (2.12)$$

$\kappa = k\,\overline{rnd^2}$、 $\overline{rnd^2}$: $rnd\ rnd$ の期待値（スカラー）

となり、この学習によって y は徐々にΦの値が大きくなる方向に学習によって移動していくことがわかる。こうすることにより、ニューラルネットの重み値自体に乱数を加えて同様な学習させる場合と比較して効率的に強化学習させることができる。

本学習は、第６章での認識および認識のためのセンサ動作の学習、第７章での状態評価値からの動作の学習、および第５章での前庭動眼反射モデルの眼の動きの学習、物体追跡モデルの眼の動きの学習に用いる。

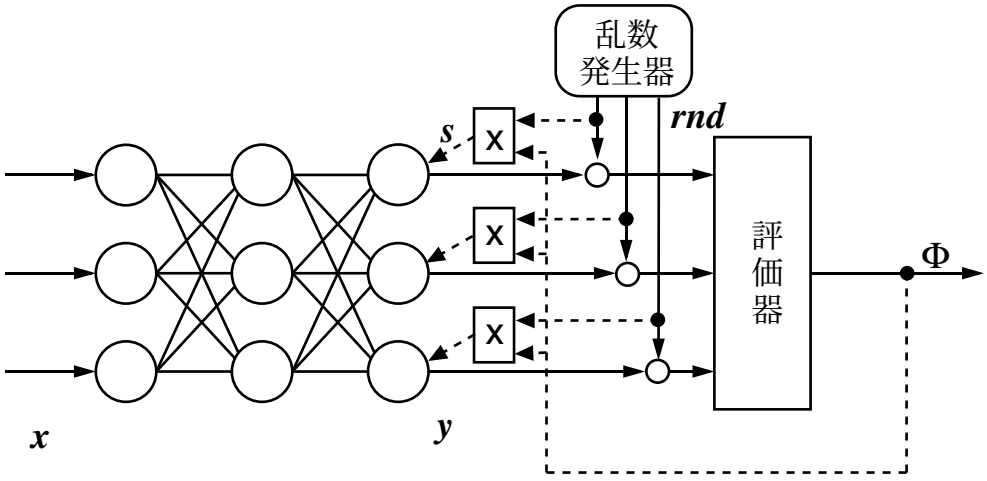

図２.１　ＢＰ法を用いた強化学習の仕組み

２.３ 相関情報抽出学習

我々は生物は、無数のセンサからの信号を受け取り、その中から重要な情報を抽出し、処理し、それに従って適切な動作を行うことができる。この無数のセンサからの信号の中で何が重要な信号かと考えてみると、まず、違った種類（例えば視覚と聴覚）のセンサからの信号の間で共通に存在する（相関の高い）情報とか、センサからの信号と運動の情報に共通な情報ではないかと考えることができる。例えば、我々が概念を形成する過程を振り返ると、複数種類のセンサ信号間に共通に存在する情報を抽出することではないかと考えることもできる（次章の考察参照）。また、センサからの信号と運動に関する信号の間に共通した情報は、運動することによって変化するセンサの情報と言い換えることができる。従って、センサを通して得られた外界の状態のうち、我々が制御できるものということになり、自律学習の観点から重要であると考えられる。例えば、視覚の情報と運動の情報の間の共通な情報として空間的な情報を挙げることができる（次章参照）。そこで、この異なった信号源からの信号を得ながら、図２.２のように、そこに共通に存在する情報、ここでは**相関情報（Correlated Information）**と呼ぶ、を誰から教えられることなく学習によって抽出することが自律学習実現に向けて必要であると考えた。

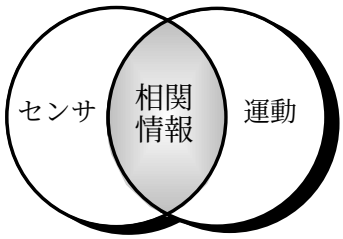

図２.２ センサと運動の相関情報

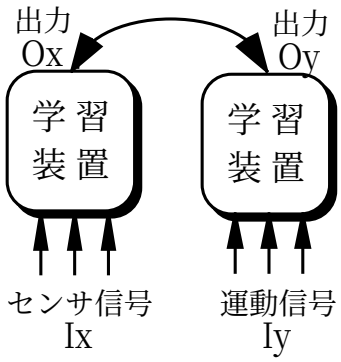

図２.３ 相関情報抽出学習

この相関情報を式で表すと、

$$\begin{aligned} r(t) &= O_x(t) = O_y(t) \\ &= f(I_x(t)) = g(I_y(t)) \end{aligned} \qquad (2.13)$$

$r(t)$：相関情報、　$O_x(t)$：センサ信号を入力とした学習装置の出力、
$O_y(t)$：運動信号を入力とした学習装置の出力、　$I_x(t)$：センサ信号入力、
$I_y(t)$：運動信号入力、　f, g：ある関数。

と表すことができる。この相関情報の抽出は、図２.３のように、２つの学習装置（階層型ニューラルネット）にそれぞれの信号を入力し、両者の出力が同じになるようにという簡単な学習を行え

ばよい。この学習は、それぞれの学習装置の出力を相手の学習装置の教師信号として与え、教師あり学習を行うことにより実行できる。これを**相関情報抽出学習 (Correlated Information Ectracting Learning)** と呼ぶ。ただし、（２.１３）式は、*r(t)*が時間によらず常に一定値をとる場合も解となってしまうため、何らかの方法で値域を拡大する必要がある。これに対しては、２.３節の値域拡大学習を利用することができる。また、相関情報がベクトルの場合は、値域拡大学習を拡張した複数出力の直交化学習によって出力間の直交化が近似的に実現できる。

第３章では、この学習の詳細を述べ、基本的な性質をシミュレーションで調べると共に、運動と視覚の情報から空間情報を抽出できることを述べる。

２.４ 時間軸スムージング学習

2.4.1 空間と時間の対応付けと時間軸スムージング学習

我々生物は、センサを通して空間的な情報を得て、将来という時間に向けて適切な動作をしていくことができる。これを実現するためには、自分が時間という流れの中で現在どういう状態にいるか、また、自分が動作を行うことによってどのように自分や外界の状態が変化するかを把握する必要がある。ここで、自分を含む我々が住んでいる世界が、決定論的に変化しているとすると、状態遷移に分岐がなく、自分や外界の状態（ここでは、空間情報と呼ぶ）は図２.４のように時間軸という１次元の軸上にマッピングできるはずである。これは、言い方を変えれば、無数次元の空間情報を１次元で表現している、または、無数次元の空間の中にポテンシャルを形成しているということになる。もちろん、このマッピングは多対１の関係になるため、時間の情報から逆に空間の情報を再現することは不可能であるが、予想通りの状態変化をした時には、将来の空間上の位置からの時間軸上へのマッピング先を予測することができる。また、もし、状態の変化に予想外のものがあり、かつ、その予想外の状態変化によってその後の状態が変化する場合は、時間軸上での本来の値とのずれによってこれを認識できるという点で広い意味での状態の予測というものにつながってくる。これを実現するアルゴリズムが、**時間軸スムージング学習 (Temporal Smoothing Learning)** である。ただし、時間は無限まで続くため、それを有限の出力値で表現することはできないため、現在何をしようとしているかとか何に注意を向けているか等の違いによって時間を区切るといった工夫をすることが必要である。

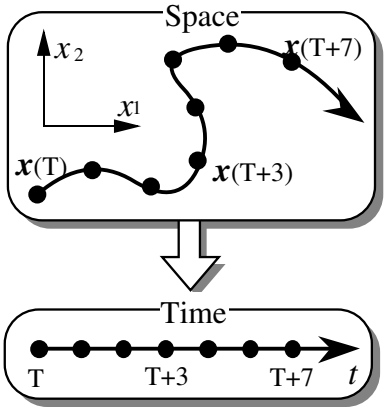

図２.４ 空間情報の時間軸へのマッピング

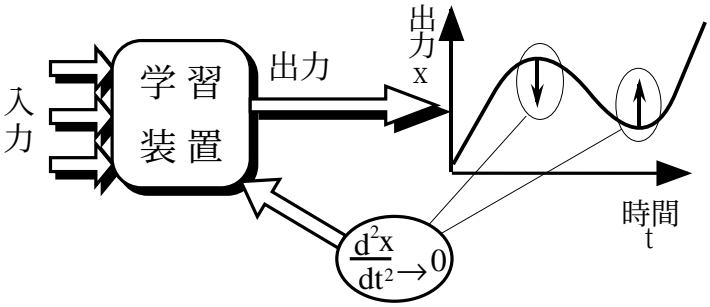

図２.５ 時間軸スムージング学習
（矢印は典型的な教師信号を示す。実際は毎単位時間学習を行う）

図２.５のように、センサの信号を入力とする学習装置（階層型ニューラルネット）を考える。ここで、空間と時間の対応付け（空間から時間軸への投射）を行うということは、時間と学習装置の出力 $x(t)$ が１対１の関係になればよい。つまり、出力が時間の変化とともに単調に増加または減少していればよいということになる。また、時間の変化に対する出力の変化量を平等に割り当てようとすると、時間に対する出力曲線は、直線になることが望ましい。そこで、出力の時間変化を見ながら、出力 $x(t)$ の時間による２階微分値を０に近づけるという非常に簡単な学習を行う。つまり、誤差 E およびそこから求められる教師信号 $s(t)$ を

$$E(t) = \frac{\kappa}{2}\left(\frac{d^2x(t)}{dt^2}\right)^2 \qquad (2.14)$$

$$s(t) = x(t) + \kappa\frac{d^2x(t)}{dt^2} \qquad (2.15)$$

κ : 学習の定数

と出力の時間による２階微分値を誤差として教師あり学習（バックプロパゲーション法）を行うことによって実現できる。ただし、ここでは、通常のバックプロパゲーション法による学習のように、同じ教師信号による学習は繰り返し行わず、毎単位時間毎に小さい学習係数で１回だけ行う。これによって、時間とともに出力が滑らかに変化するようになり、センサの情報を入力すれば、時間の情報が得られるということになる。ただし、この学習も、相関情報抽出学習と同様に、出力が時間にかかわらず一定値の場合も解となるため、何らかの方法で値域を拡大しなければならない。本論文では、強化学習と局所センサ信号の統合にこの学習を用いるが、これについて以下に述べる。

2.4.2 遅延強化学習と時間軸スムージング学習

前章で述べたように、強化学習では、一連の動作の後に得られる報酬や罰からいかにそれまでの動作を学習するか（遅延強化学習）が大きなテーマになっている。Barto らは、現在から将来にわたって得られる報酬を現在に近いものほど重要視するという指数関数による重み付けをしたものの総和を評価値とし、それを最大化するという観点から定式化を行っている[Barto 83]。ここで、単一の報酬源を考えると、重み付けの効果によって、報酬から時間的に遠ざかれば遠ざかるほど評価値が指数関数的に下がることになるため、評価値を最大化するということは、報酬が得られるまでの時間を最短化することであると解釈することができる。ここで、時間というファクターが登場する。

前述のように、時間軸スムージング学習を用いると、センサの情報から時間の情報を得ることができる。そして、それに加えて、報酬が得られた時点でその出力が最大の値になるように学習すると、この出力は、時間的に報酬に近いほど高く、遠いほど小さい値を出すようになる。つまり、ニューラルネットの出力は、現在のセンサの状態から、報酬が得られるまでの所要時間を予測しているということになる。

そこで、今度は、自分の動作を、その予測値がより大きくなるように学習させる。そして、これによって変化した所要時間を学習し、さらにその予測値から動作を学習しということを繰り返していくことによって、報酬を得るための最適に近い動作を学習することができる。詳細は、第７章にて述べる。

2.4.3 センサ信号の統合と時間軸スムージング学習

我々の住んでいる世界では、動いている物体は、慣性の法則に従い、突然消えたり、突然現れたり、原因もなく動いている方向が突然変化することはない。だからこそ、我々は物体の動きを予測し、それに基づいて適切な動作を行うことができると考えることができる。ここでは、これを**空間情報の時間的滑らか仮説**と呼ぶ。

ところが、一般的に、センサは非常にたくさんのセンサ細胞を使って空間の情報を受け取る。従って、個々のセンサ細胞が検知する信号は必ずしも時間的に滑らかではない。にもかかわらず、我々は、例えば物体の位置などの空間の情報を個々のセンサ細胞の出力を意識することなく連続的に認識をしている。つまり、空間の情報の時間的滑らかさという拘束を利用して頭の中で空間を再構成していると考えられる。別の言い方をすれば、無数の空間情報のうち、時間と共に変化する情報を抽出しているということもできる。そして、時間軸スムージング学習によって、時間と共に変化す

る空間情報を抽出することができる。ただし、ここでは値域拡大のために２.３節の値域拡大学習を用いる。詳細は、第４章において述べ、第５章にて、視覚システムの学習モデルへの応用について述べる。

２.５ 値域拡大学習と複数出力の直交化学習

前述のように、相関情報抽出学習および時間軸スムージング学習では、共に出力に拘束を設けるだけであり、出力の値域を確保することはできない。時間軸スムージング学習を強化学習に応用した場合には、最終的に報酬を得た時点での出力が最大に、スタート地点の出力を小さくという学習になるため、値域は確保されるが、空間情報の抽出の学習においては、何を最大値や最小値にすべきかという明確な定義ができない。ところが、相関情報抽出学習の場合は、複数の出力が常に一定の値になり、時間軸スムージング学習の場合は時間の経過による出力の変化が０という解を持ち、いずれの場合も学習の容易さからそのような解に陥る。そこで、出力に対する拘束を生かしつつ、出力の値域を確保するために、**値域拡大学習 (Value Range Expanding Learning)** を行う。基本的には、図２.６のように、過去の出力の最大値の時、および最小値の時の入力パターンを記憶し、そのパターンを再入力して、出力の値域の最大値、最小値をそれぞれ教師信号として学習させることによって実現できる。しかし、過去のパターンを記憶して再学習させることは、非常に無駄が多く、適応性にも欠けるため、過去の出力の平均と偏差の平均を１次遅れを用いて計算しつつ、偏差の大きい出力に対して値域拡大の学習を行わせることができる。ただし、この方法でも偏差の大きい時だけ特殊な学習を行う必要があるため、よりスマートな学習則への改善の余地があると考える。

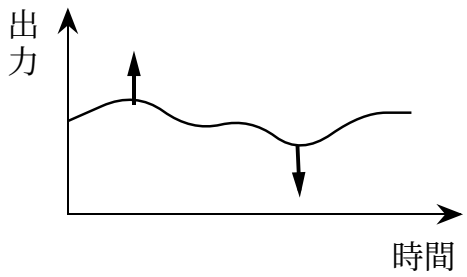

図２.６ 値域拡大学習の模式図

また、時間軸スムージング学習や相関情報抽出学習によって情報の抽出を行い、かつ、複数次元の情報を抽出する場合を考える。この場合は、複数個の出力を設け、さらにその複数個の出力がうまく情報を分担するように学習を行う必要がある。予め入力データが決まっている場合は、相互情報量を最大化する等の手段があるが、逐次的に学習を行う場合には適用が困難である。そこで、他の出力の偏差が小さく、該当する出力の偏差が大きい時に該当する出力の値域を拡大する学習を行うという方法によって、逐次的な学習の場合でも近似的に複数の出力を直交化することができる。これを**複数出力の直交化学習**と呼ぶ。

本学習は、前述の相関情報抽出学習および時間軸スムージング学習と併用する。それぞれ、第３章および第４章において述べる。ただし、第４章では、複数次元情報の抽出は現時点では行っていないため、直交化については、第３章において述べる。