ニューラルネットを用いた自律学習システムの研究

桑田 克成

学位請求論文

ニューラルネットを用いた
自律学習システムの研究

指導教官　岡部 洋一 教授

平成８年１１月１日提出

柴田 克成

目次

第1章　序論

1.1 自律学習システムとニューラルネット

近年、コンピュータを始めとする知能システムの進歩はめざましいものがあり、我々の生活の中でなくてはならないものとなっている。にもかかわらず、我々は、これらを使っていく上で、言われたことしかしない、融通が効かない等の印象を受ける。つまり、我々人間を始めとする生物と比較して、数値演算能力ははるかに勝るものの、柔軟性・適応性という面で大きく劣る。本研究では、生物のような柔軟かつ適応的な知能システムの構築を大きな目標として置いている。

我々生物と従来の知能システムは、いずれも何らかの入力を受けてそれに対して何らかの出力をするという構成になっている。しかし、入力から出力への処理方法の獲得が自律的な学習によるかどうかという点と連続値を主体とした処理かどうかという点で相違点があり、それが前述のような差異を生み出していると考えられる。以下、この2つの相違点から、自律学習の必要性とニューラルネット適用の有効性について述べる。

1.1.1 自律学習システム

処理方法の獲得という点に着目すると、従来の知能システムは、主に、図1.1に示したように、人間が入力から出力への処理方法を与えるという形で知能化されている知識付与型の知能システムである。一方、我々生物は、入出力を見ながら自ら学習していくことができる。これをここでは**自律学習**と呼び、このような機能を持つシステムを**自律学習システム**と呼ぶ。中でも特に、図1.2のように運動出力と感覚入力を持ち、外界とのフィードバックループを形成することにより、処理方法を獲得・更新すること（フィードバック学習[Okabe 88]）が重要な役割を果たす。

知識付与型知能システムでは、言われたことを忠実にこなすことは可能であるが、自ら学習する手段を持っていない。しかし、知能システムに対する要求は年々高度になる一方である。この要求を満足するような知識を付与するためには、非常に緻密かつばく大な量のプログラミングが必要になると考えられる。要求される機能のレベルおよびプログラミングの難しさを定量的に表すことは困難であるが、直感的には、例えば機能のレベルと与える知識量の関係がピラミッドの容積と高さのように、要求される機能のレベルが高くなればなるほど、レベルを上げるためのプログラミングはそれ以上に難しくなってくると考えることができる。特に適応ということを考えると、あらゆる場面を予め想定し、それに対しプログラミングしなければならず、大変困難である。従って、初期の頃の知能システムの進歩と比較して、今後の知能システムの進歩の速度は徐々に鈍ってくることが予測される。

一方、自律学習システムである我々生物は、前述のようにフィードバックループを有し、自ら行

動し、外界の状態をセンサによって取り込み、自らの動作が引き起こす外界の変化を知ることによって学習し、知識を獲得していくことができる。従って、機能は学習によって獲得できるため、予め与えなくてはならない情報は少なくても良く、その分、適応的であると言える。また、第三者がいなくても自ら行動できるため、それだけ多くの経験を積むことができるし、行動することにより学習を行うため、必要な学習を能動的に行うことができる可能性も有する。これらのことから、将来の知能システムの発展を考えると、今こそ自律学習システムに着目し、我々生物をお手本としてその能力を向上させていくことが必要と考え、本研究を進めて来た。しかし、どこまでが自律学習かと言っても、その境界は必ずしもはっきりしない。例えば、システムの処理に一つのパラメータを導入し、入出力を見ながら学習によってそのパラメータを変化させていくといった方法も広い意味で自律学習と言える。このような意味で、システムの処理における既定部分と可変部分の比が一つの自律学習の尺度となる。そこで、ここでは、できる限り設計部分を減らし、汎用的な学習によって多くの知識を獲得することを目指す。ただし、可変部分と言っても初期値を与える必要があり、そこに予め知識を埋め込むことは可能である。

知識付与と自律学習の中間的な位置づけのものとして教師あり学習がある。中でも、ニューラルネットの教師あり学習アルゴリズムであるバックプロパゲーション法（以下、ＢＰ法と省略）[Rummelhart 86]は有名である。これは、図1.3のように、人間が入出力サンプルのセットを与えることによって、入出力関係を学習させるため、直接プログラミングしなくても良いという利点があり、人間がプログラミングする際に不得意としたパターン認識などパターン情報の処理をサンプルから

図1.1　従来の知能システム
（知識付与型知能システム）

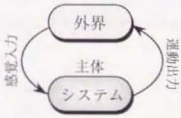


図1.2　生物を始めとする自律学習システム

図1.3　バックプロパゲーション法によるニューラルネットの教師あり学習

の学習によって獲得できるということで威力を発揮した。そして、知識付与型知能システムからの脱却を夢見て多くの研究者が研究を行った。しかし、確かに直接知識を与えなくても良かったが、知識を入出力サンプルという形で付与しなければならなかった。しかも、そのサンプルを作ることが意外と難しく、特に、環境が変化していくような場合、そこにシステムが適応していくことも困難であるということがわかってきた。そして、何より、生物のように、フィードバックループを持っていないため、与えたサンプル以上の機能を学習によって獲得することはできないし、その機能が必ずしもそのシステムにとって最適なものであるかどうかわからないという、知識付与型知能システムと類似した壁にぶつかることになった。ここ数年、ニューラルネットの研究が少し下火になってきているが、この壁が下火になった大きな要因の一つと筆者は考える。

自律学習においては、いわゆるＢＰ法等の教師あり学習と違い、外部から直接理想出力である教師信号を得ることができず、自らの持つ何らかの学習指針に従って学習を進めていかなければならない。この学習の規準をどのようなものにするかが自律学習システムの性能を決定する大きなポイントとなる。自律学習は、外界とのループが重要な役割を果たすため、従来のように、文字認識、画像認識、音声認識、自然言語処理、制御といった個々の機能を独立して研究し、それを後で融合するというアプローチだけではなく、簡単でもいいから外界とのループを形成し、ループを利用した学習、ループを考慮した学習がどうあるべきかを考え、それを発展させていくというアプローチをとることが重要である。また、その学習は、構成の容易さという面からシンプルなほど良いが、学習の汎用性という面から考えても、シンプルな学習則からたくさんの機能を学習できる方がより柔軟性が高い学習則であると考えることができる。与える情報が多ければ多いほどそれに縛られ、学習による自由度が小さくなるからである。この関係を模式的に表したものを図1.4に示す。また、このような意味で、教師あり学習と比較すると、一般的に与えられる情報量が少ないため、より柔軟な学習が期待できる反面、探索空間が広くなるため、学習のために非常に長い時間を要するという欠点がある。

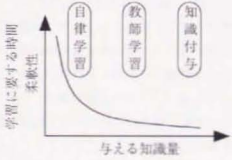


図1.4　与える情報量と柔軟性および学習に必要な時間の関係の模式図

1.1.2 パターン情報の学習・処理とニューラルネット

我々生物と従来の知能システムとの差異を内部での処理形態という点から捉えると、パターンを基盤としているか、シンボルを基盤としているかという違いが浮かび上がる。ただし、ここでパターンとは、距離および微分が意味をなす情報のこととする。

人間のような存在を考えると、一見、論理的思考、シンボル処理こそが高度な知能を支えているように見える。しかし、シンボルはその背後に隠れる様々な知識の象徴として存在し、その土台があるが故に、コミュニケーションや論理的思考といったシンボルの処理が有効になるものと考える。そして、その背後の知識がパターンとして構成されているため、距離が意味を持つことによって汎化能力につながり、微分情報を使ってその先を予測したり、山登り法（最急降下法）が適用できるため学習が効率的に行えるようになる。そして、これが我々生物が柔軟な処理を行うことができる大きな理由であると考える。シンボルとパターンの関係は、ちょうど図1.5のようにシンボルはちょうど氷山の一角のようなものであり、シンボルだけを取り出してその論理的な処理方法を考え、それを高次の処理と呼ぶことは、水面下の氷山の存在を無視して氷山のことを語ることに等しいと考える。

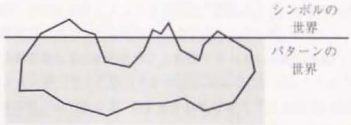


図1.5 氷山を用いたシンボルとパターンの関係の模式図

パターン情報の処理を人間がプログラミングすることは、自由度が大きく、人間にとって大変困難な仕事となる。例えば、文字認識のプログラムを作ろうとしても簡単にはいかない。さらに、我々生物の処理を観察すると、様々な要因が複雑に絡み合って柔軟な処理をしているように見える。例えば、急いでキーボードに入力しようした時のミスの仕方を観察すると、キーボード上で近いキーを打ったり、発音が近いものを打ったり、単語の前後がひっくり返ったり、前後の文字の子音だけ入れ替わったり、近い意味の違う単語と間違えたりと様々な間違い方をすることにしばしば驚く。また、このような機能を人間がプログラミングしようとしてもほとんど不可能であると考えられる。従って、このような処理を実現しようとするならば、時間は掛かっても何らかの評価関数を作って微分情報を用いた学習に頼らざるを得ない。

以上の理由から、パターン情報の処理を得意とし、かつ学習能力を有するニューラルネットは自律学習システムを構築する上で強力な実現手段と考える。パターン情報を扱うニューラルネットの汎化能力は、過去の経験をいかに将来に生かし、柔軟な処理を行うことつながってくる。さらに、人間や他の生物といった自然界の高度な自律学習システムの中枢がニューラルネットで構成されており、言語などのシンボル処理ですらニューラルネットが担っていること、そして、この構成がおそらく長

い年月を経て最適化された結果であることもニューラルネットの使用の妥当性に関する大きな拠り所である。

また、ニューラルネットを使えば、複数の要因を適切にミックスすることが比較的容易にできる。例えば、筆者は以前、ロボットの制御を学習する際に、制御における誤差を小さくするという学習を行うだけで、フィードバック制御とフィードフォワード制御のハイブリッド制御の学習が可能であることを示した[柴田 89]。そしてさらに、ノイズが多い環境で学習した場合には、フィードバック制御が主流になるという適応能力も備えていることもわかった。このように、ニューラルネットは、学習のための評価関数を設定すれば、そのための適切な統合を学習によって獲得していくことができるのである。

一時期、右脳と左脳がパターン的な処理とシンボル的な処理を主に分担しているという知見を基に、それと同様にニューラルネットと従来の計算機を統合させれば人間のような柔軟なシステムができるといった議論がなされた。しかし、筆者は、全く異質な両者を最終的に統合することこそ難しく、統合することで両者の利点が失われる可能性が高いと考える。また、前述の議論のように、シンボル処理とパターン処理は非常に密接な関係にあるため、両者を分離するのではなく、一つの枠組みの中で捉えて行くべきであると考える。つまり、ニューラルネットを用いてシンボル処理を行うことによって、シンボルとパターンのインターフェイスの問題が解決され、ニューラルネットの学習・処理の柔軟さがシンボルを用いた表現の中にも十分に発揮されると考える。また、シンボル処理は、コミュニケーションや論理的思考を行うための必然から獲得できるという期待を持っている。

バックプロパゲーション法（ＢＰ）法は前節で述べたように、ニューラルネットを用いた教師あり学習の代表的なアルゴリズムであるが、ＢＰ法を単独で使うだけでは、教師信号を用意しなければならないため、自律学習を実現することはできないと述べた。しかし、 ＢＰ法は、教師信号と実際の出力の差の自乗を評価関数として最急降下法を適用した単純明快な学習アルゴリズムであり、中間層のニューロンを十分用意し、いくつかの教師信号を与えれば、元となった関数を近似することができるようになる等、その学習能力は強力なものであると筆者は認識する。そこで、本論文のほとんどの部分で、システムが教師信号を自動生成するという形で自律学習を保ちつつＢＰ法を適用するという手法をとった。ＢＰ法の具体的なアルゴリズムについては２.１節で、さらにＢＰ法を用いて強化学習を効率的に行う方法については、２.２節にて述べる。

ＢＰ法の前身である、パーセプトロンについては、小脳パーセプトロン説でも見られるように、それを実現するような機構が脳内に存在する可能性があると考えられている。しかし、ＢＰ法に関しては、逆伝搬誤差がニューロンをまたいで伝搬しなければならない等の理由から、そのような機構は脳内に存在しないと言う見方が大半である。しかし、私は、ニューロンの成長と学習を同一の範疇で捉えることができるのではないかと考えており、その考えに基づくと、神経の成長を促進する神経成長因子（NGF Neuro Growth Factor）[モンタルチニ 79][畠中 92] 等が学習にも関与し、さらには、ニューロンを介して他のニューロンへと伝搬することも考えられるのではないかと期待している。

また、ＢＰ法では、ニューロンの出力関数に非線形関数を用い、さらに、最急降下法を適用しているため、局所最適解（ローカルミニマ）に陥るという問題点がよく指摘されている。しかし、本論文の研究を進めるに当たって、ローカルミニマにトラップされて学習がうまく行かなかったという状況にはほとんど遭遇しなかった。ローカルミニマの例として、Exculsive-OR（排他的論理和）の学習

が挙げられる。この問題は、2次元の入力空間上の4点で教師信号が与えられ、入力空間を教師信号の値によって分離する際に線形分離ができない問題である。しかし、この問題でも中間層のニューロン数を増やせばローカルミニマを回避することができるし、現実に我々が存在する空間は、我々が意識している以上に線形性が強く、また、たくさんのデータを取得することによってこれらの問題をあまり考慮する必要はないと考える。

さらに、最近、ニューロンの出力関数に単調増加型のシグモイド関数ではなく、ガウス関数のようなＲＢＦ（Radial Basis Function）を用いた場合をよく見かける。これは、局所的な情報の和として出力を表現しようとしたもので、わかりやすく、また、ある入力における学習が他の部分に影響を与えない等という点から良く使われている。しかし、前述のように、我々の住む世界は線形性が高いものであり、シグモイド関数のような線形性の強い関数を使う方が効率的に状態を表現できると共に、汎化能力も活かせると考えられる。また、シグモイド関数は、入力が0に近いほど入力に対する出力の感度が大きくなる。これは、100.0と100.1の違いよりも0.1と0.2の違いを大きく感じる我々の感覚と合っており、$-\infty$から∞の入力を0から1の出力に変換する効率的な方法であると考える。以上より、本論文中では、従来通りのシグモイド関数を用いた単純なＢＰ法を学習に用いた。

また、よく、ニューラルネットは中がブラックボックスだから、たとえ学習によって機能が獲得できたとしても意味がないという議論がある。しかし、筆者は、人間や生物自体の処理は前述のように非常に複雑で表現することが困難なものであると考える。従って、生物の機能の本質を探るならば、また、生物のような柔軟な機能を実現しようとするならば、個々の機能を正面から考え、解析していくよりも、その学習方法を解析し、その実現を試みる方がより適切であり、近道であると考える。

1.2　自律学習における学習指針と強化学習

1.1節において、自律学習においては、いかなる学習指針を用いるかが1つのポイントであることを述べた。この学習指針は、汎用的であり、かつ、これによってシステム（エージェント）が有効に働くものでなければならない。以下では、筆者が最も有効と考える強化学習とその他の学習指針について述べる。

1.2.1 強化学習

強化学習とは、図1.6のような仕組みの下で報酬や罰といった強化信号から、より報酬を得られて、罰を避けられる動作を学習することを言う。基本的には、何らかの動作を行って、報酬が得られればその動作を強化し、罰が与えられた時には、次回から同じような状況の下では同じ動作をしないように学習を行う。これは、まさに外界とのフィードバックループを使った自律学習である。この強化学習は、報酬や罰から学習する点が、我々生物においても観察できる非常に自然なアルゴリズムであるということができ、また、報酬や罰という非常に簡単な情報から様々な動作が学習できるという点で非常に強力な学習アルゴリズムであると言うことができる。本論文でも、この強化学習を自律学習の要として捉えていく。

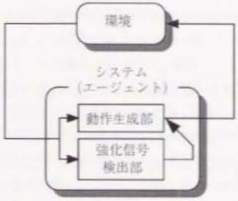


図１.６　強化学習の仕組み

強化学習は、教師あり学習か教師なし学習かという議論がよくなされる。確かに、報酬や罰の検出方法を予め与えなければならないため、教師あり学習であると考えることができるが、出力に対し、直接教師信号を与えるわけではないので教師あり学習でもないといえる。結局、言葉の定義の問題であるが、両者の中間的存在であると考えるのが妥当であろう。

ただ、例えば、ニューラルネットにおいて教師信号と実際の出力の差の自乗を罰として与えれば、教師あり学習に近いこともできる。また、実際の生体でも何を生得的に報酬と感じ、何を罰と感じるかは判断の分かれるところである。我々がシステムに強化学習を適用する際にも、この強化信号をどう設定するかが一つの大きな問題である。このような問題に対し、何を報酬や罰にするかは進化の過程で決定されると言う考え方も出てきている。生き延びるために必要なことを報酬と感じ、死に関係することを罰と感じれば、その生物は生き延びる可能性が高くなるということは、実際の生物を考えた時にもっともで興味深い考え方である。しかし、この場合も、我々がこれを利用しようとすると、進化における淘汰の評価関数をどう設定するかという問題は依然として残る。

また、心理学の分野では、反応遮断化理論[Allison 74] [坂上 94]というものが最近受け入れられているようである。これは、強化学習を刺激-反応という図式で捉えずに、刺激も反応と置き換えることにより、強化を「複数の行動を適切な割合に再配分すること」と解釈することである。これは、ニューロンや個体は、適度な刺激と適度な（モデレートな）動作を好むというモデレーショニズム[Okabe 88]というもう一つの自律学習アルゴリズムの考え方につながるものがある（本節後述）。しかし、この場合も、やはり適度というものをどう設定するかが問題となる。いずれにせよ、実際の生物がどのようにして強化信号を決定しているか、また我々がシステムを作る際にはどうすべきかは今後のさらなる検討が必要である。本論文中では、第６章で、強化学習を認識や認識のための動作に適用することを試みたが、ここでは、認識出力がいかに理想値に近いかを強化信号として利用した。また、第７章では、強化学習の例題として、移動ロボットが目標物に到達するという問題を考えているが、ここでは、移動ロボットが目標物に到達した時に、強化信号が検出されるという設定で学習を行った。

強化学習の源は、心理学の分野でのオペラント条件付けに見ることができる。古くは、１９世紀終わり頃、Thorndike の問題箱の実験[Hebb 72] [東 69]がある。ここでは、ネズミを箱の中に入れ、ドアに仕掛けをし、ドアの外にエサを置くことによって、試行を重ねるに従って、その仕掛けをはず

してドアを開け、エサを獲得するまでの所要時間が徐々に減少することを確認している。さらに、1961年の Skinner のスキナー箱として有名な実験がある[Skinner 61]。この実験は、図1.7のように、レバーを押すと上からエサが落ちてくる仕組みになっている箱の中にネズミを入れると、ネズミはうろうろして偶然レバーを押してエサを得ることを2・3回繰り返しているうちに、自分でレバーを押してエサを得るようになるというものである。これは、あたかもネズミがレバーを押すとエサが得られるという因果関係を把握し、さらにエサを得るためにレバーを押そうという意志を形成したかのように見える。これに対し、ネズミは単なる反射を形成しているに過ぎず、我々人間の因果関係の把握や意志とは明らかに違うと考える意見もある。しかし、所詮、我々人間も経験に基づいて、物事の因果関係を把握し、報酬を求めて意志が形成されているのであり、両者の間に本質的な違いはないと筆者は考える。このネズミの行っている学習こそが、今の知識付与型知能システムにない機能であり、自律学習（強化学習）の重要性を物語っていると考える。

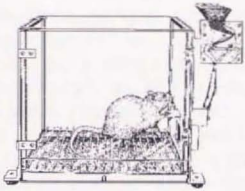


図1.7　Skinner の実験

一方、計算論的な強化学習の研究は、1950 年の Samuel の Checker のゲームの学習に関する研究[Samuel 59] が始まりであると言われている。ここでは、ゲームの木の探索に用いる評価関数に対し、ゲームに勝ったか負けたかを評価関数として強化を行うというものである。

強化学習においては、主体が試行錯誤の成分を含む動作を行った直後に強化信号が得られれば、報酬の場合はその試行錯誤の成分を含めた動作を学習させ、罰の場合は、試行錯誤の成分と逆の成分を加えた動作を学習させることで簡単に動作を学習することができる。しかし、通常は、一連の動作の後に強化信号が得られる場合が多く、これが学習を難しくしている。通常、この問題を単に強化学習と呼ぶ。しかし、本論文では、第6章で強化学習を認識や認識のための動作の学習に適用する際に、単純化のため、毎単位時間毎に評価（強化信号）が得られると言う設定を行っている。そこで、一連の動作の後に遅れて強化信号が得られる場合を遅延強化学習、毎単位時間毎に強化信号が得られる場合を単なる強化学習と呼んで区別する。

この遅延強化学習は、Profit Sharing [Holland 87]のように、得られた報酬を使って直接過去の一連の動作を強化する方法と、報酬から中間的な評価を学習し、その評価値から動作を学習するものとに分けられる。前者は、報酬や罰が得られた時に、そこまでの動作を直接強化する方法であり、前述の

Samuel らの研究はこれに属する。前者は、直接動作を強化するため、学習は速いが、過去の全ての状態を保持しなければならない上、状態に対する評価がなされないため、汎化能力に欠けると筆者は考える。

後者の学習方法としては、1983 年の倒立振子の制御の学習を例題として扱った Barto らの critic-actor アーキテクチャ（ＴＤ学習）[Barto 83] および Watkins らのＱ学習[Watkins 92]が有名である。ＴＤ学習では、現在から将来にわたる強化信号の重み付き総和（未来ほど重みを指数関数的に小さくする）を最大化することを目標にしたものである。そして、この強化信号の重み付き総和を状態（センサ入力）から予測し、それを評価値とし、その評価値が良くなるように動作を学習するという方法である。そして、この評価値はより正しく強化信号の重み付き総和を予測するように１単位時間先の評価値の値を元に更新される。これによって、倒立振子の角度がある値より大きくなったときにペナルティを与えるだけで、倒立振子が倒れないよう制御を学習することができる（第７章参照）。

一方、Ｑ学習は、Bellman の最適化原理 [Bellman 57] に基づき、状態と動作のペアに対して評価値（Ｑ値）を設定し、そのＱ値が最大となる動作を選択するという方法である。そして、やはり１単位時間先の評価値から現在の評価値を学習する。どちらかといえば、Ｑ学習の方が、動的計画法（ＤＰ、Dynamic Programing）から直接導かれたものであり、マルコフ決定過程の上で最適な行動の獲得が保証されており、広く用いられている。しかし、元々動作は状態から状態への遷移を表していると考えれば、状態の評価が決まれば動作に対する評価も一意に決まることになる。従って、動作が評価関数のパラメータになることによって評価関数の入力空間の次元が増えることになり、汎化能力に欠ける上、連続値動作は現在のところ扱うことができないという問題点があると筆者は考える。

その後、Anderson らは、ＴＤ学習に多層ニューラルネットを用い、学習にＢＰ法を用いることによって倒立振子が倒れるまでの回数が飛躍的に伸びることを示している[Anderson 89]。また、最近は、Barto らによって、ＴＤ学習と動的計画法 (Dynamic Programing) との関係からＴＤ学習の収束性を示す議論が行われている[Barto 95-1]。また、Schultz らは、サルを用いた条件反射の学習の実験において、大脳基底核のドーパミンニューロンが、最初は報酬に反応するが、慣れると報酬に反応しなくなり、そのかわり前兆となる感覚刺激に対し反応するようになることを示している [Shultz 93]。これに対し、Barto らは、ドーパミンニューロンの反応がＴＤ学習における誤差に相当するとの考え方に基づき、大脳基底核においてＴＤ学習が行われているのではないかとのモデルが示されている[Barto 95-2][Houk 95]。

これらの動きとは別に、Widraw らは、トレーラーの車庫いれ問題を例にとり、制御対象のダイナミクスを制御信号をランダムに変化させながらニューラルネットで学習させ、このニューラルネットに制御時の誤差を逆伝播させることによって、制御信号の誤差を求め、さらにその誤差から制御信号を生成計算する別のニューラルネットの学習をさせるという学習[Widrow 85]を提案している。また、Werbos らはこの流れを汲み、システムに外界のモデルを持たせるべきであるとし、制御対象のモデルを学習させたニューラルネットと制御信号生成のニューラルネットの様々な組み合わせを提案している[Werbos 90] 。

日本では、遅延強化学習ではないが、鋼谷らの歩行パターンを学習するロボットへの応用[鋼谷 86]が先駆的な研究として挙げられる。ここでは、リンク機構を２つ持ったロボットにロータリーエンコーダを付け、この値を強化信号とすることによって、ロボットに歩行パターンを予め教えることなく

学習によって獲得させている。また、依田らはヒトデの起きあがり問題への適用している[依田 90]。最近では、Onat らは、フィードフォワード型のニューラルネットの中間層の出力値を次の時間に外部からの信号と共に入力層に与える Elman 型のリカレントニューラルネット[Elman 90]をQ学習に用い、リカレントネットに蓄えられた記憶を有効に利用して動作学習を行わせるという研究を行っており[Onat 95]、鋼谷はＴＤ学習を連続時間へ拡張するという研究を行っている[鋼谷 95]。

また、より実用に近い応用としては、ＴＤ学習を Backgammon というゲームに適用した例があり、人間のチャンピオンと互角の勝負をしている[Tesauro 92]。また、浅田らは、視覚センサを持ったロボットにサッカーを行わせる際に、Q学習によってボールをゴールにシュートさせることを学習したり[浅田 95]、敵を避ける動作の学習とシュートの学習を別々に行った後にそれを統合するという面白い試みをしている[Asada 94]。

強化学習の解説記事として [畝見 95] [鋼谷 96] が出ており、詳しくはこちらを参照されたい。

1.2.2 その他の学習指針

強化学習以外のフィードバック学習アルゴリズムとしてモデレーションニズムがある[Okabe 88][甲原 94]。これは、前述のように、各ニューロンの入出力がモデレートなレベルを好むという非常に簡単な学習則である。これによって、生体自身が適度な刺激を求めつつ過度な刺激から逃れることによって、生き延びることができるという考え方である。これは、個々のニューロンレベルで学習することを目指しており、個々のニューロンの学習を統合的に管理する部分を持つ必要がなく、フォルトトレランスの観点からも優れている。現在、人工腕での反射弓が学習によって獲得できることが示されている[浅野 95]が、今後、高次機能実現への道筋が示されることが大いに期待される。

もう一つのフィードバック学習として、筆者は、「できるだけ外界の情報を得るための動作の学習」というもの考える。例えば、我々は何か物音がすれば、そちらに目を向けるとか、動いている物体を追跡するといった動作を行う。これは、外界の情報をよりたくさん得られれば、より正しい認識ができ、より報酬を得られやすくなると考えることによって、強化学習から学習できる機能であると考えることができる。筆者も、これに近い考え方で能動認識を強化学習で学習させようとしており、第6章でこれについて述べる。しかし、外界の情報をよりたくさん得るということは、システムにとってプラスになってもマイナスになることはあまり考えられない上、これに基づく学習則は、特定の場面にしか適用できないものではなく、汎用的な学習則であると言える。また、我々が持っている好奇心というものもこれによるものと考えることも可能である。そこで、これも学習指針の一つとして挙げても良いのではないかと筆者は考える。この問題は、得られる外界の情報量を強化信号として強化学習を行っていると考えることもできる。ただし、外界の情報の量をどのように表すかは問題である。本論文中では、この学習指針はあまり多く用いていないが、第5章の物体追視のモデルにおいて、抽出した空間情報の時間変化が大きい方がたくさんの情報を得ているという観点から、より情報を得られるような動作を学習することを試みている。

以上、フィードバック学習の際の学習指針に関して考えてきたが、我々人間のようにセンサから時々刻々たくさんの情報が得られる場合には、センサの信号から前もって必要な情報を抽出し、抽出した情報から効率的にフィードバック学習を行うことも重要ではないかと考えられる。そこで、複数

種の情報源、特にセンサと運動の間や、異種センサの間に共通して存在する情報が重要な情報であるという考え方から、教師なしでその情報の抽出を学習する方法を考えた。これが、相関情報抽出学習であり、第2章でその概要を述べると共に、第3章で、詳細および空間認識との関連を探る。

さらに、センサから得られる空間情報と時間の関係が重要であると考え、その関係を学習によって獲得することを考えた。そして、センサから得られる信号を入力とし、時間と共に滑らかに変化する出力を学習する方法を提案した。これが時間軸スムージング学習であり、第2章で概要を述べる。そして、これは、センサ信号の統合および遅延強化信号に用いることができるが、これを第4章および第7章で述べる。

また、De Charms, R.C. は「全ての人間は、自己の環境に変化を生み出すことに効果的でありたいと願っており、自分の運命を自分でコントロールし、外界にもて遊ばれたくないと願っている」[Charms 76][丸野 89]という人間の行動の捉え方をしている。筆者は、自分自身を振り返り、この考え方にも賛同する。まず、外界にもてあそばれないためには、まず外界の変化を予測することが必要である。これに関しては、前述の時間軸スムージング学習がこの役割を担うものと考えており、本論文の一つの大きな柱である。そして、外界の予測を行った上で、外界にもて遊ばれず、自分の運命を自分でコントロールするという点に関しては、強化学習がその役割を果たすと考える。また、外界に変化を生み出すことに効率的であるという点に関しては、時間軸スムージング学習と強化学習の組み合わせによって可能であると考える。

さらに人間の学習の中で、模倣というものが大きな役割を示すことも忘れてはならない。中でも、生まれて2、3カ月の子どもが、おとなが口をゆっくり開閉すると真似をしようとする [丸野 89]ことは、大変に驚くべきことである。模倣とはいえ、これを単純な教師あり学習としてすませることはできない。[丸野 89]で指摘されているように、赤ん坊がこのような機能を獲得するためには、1．口を開閉することを見ることができる、2．相手の口と自分の口が対応していることを知っている（手を動かさないで口を動かす）、3．視覚の情報と運動の情報を結び付けることができる、4．真似をすることがいいまたは真似をしなければならないと思っている、等のことが必要である。これらの全ての機能を生後（受精後？）の短期間で学習によって獲得すると考えることはかなり難しく、予め備わっている機能と考えた方が自然である。しかし、おそらく自分が口を動かすとおとなが口を動かしてくれる、または、おとなが口を動かした時に自分が口を動かすと親が非常に喜んでくれるというもっとも身近なフィードバックループが形成され、自分がおとなの動きをコントロールできるということに快感を得ているということが赤ん坊のこのような動きを学習させていると言う面もあるのではないかと考える。

また、より成長してからの模倣に関しても、最初は模倣するとまわりに喜ばれるというレベルから始まり、その後模倣することが問題解決への近道であることを学習し、さらに、模倣すること自体に快感を覚えるようになり、模倣がさらに加速され、高度な機能を身につけることに役立っていると考える。

1.3 基本的立場

本論文の理解を深めるため、この節では、本論文に関する研究を遂行するに当たって筆者がとってきた基本的な立場を明らかかす。

1.3.1 学習の段階性と並列・統合学習

学習においていきなり難しい問題を解くことは困難であるが、簡単なことから徐々に積み上げることで高度な機能の学習が可能になる。例えば、ハトを使ったスキナー箱の実験において、ハトがある場所をつつくとエサが出るという仕組みになっている場合、つつくポイントを徐々に上にもっていってやるとかなり上の方までつつくことができるようになる[東 69]。しかし、もし、最初から高い所をつつかなければエサがでてこなければ学習不可能であろう。これは、我々人間の発達過程が、非常に無力な赤ん坊の頃から少しずつ少しずつ積み重ねていくことからも言えるであろう。例えば、赤ん坊の発達過程で、近知覚（触覚等）から遠知覚（視覚等）へと学習が進む[丸野 89]のもそのためであろう。しかし、発達の過程において、諸処の機能を、例えばある時期は歩行を学習し、ある時期は空間認識の学習をするというようにシーケンシャルに学習していくのではなく、たくさんの機能を並列に、そして徐々にその機能をアップさせているように見える。これによって、相互の機能間の関係を密接に積み上げていくことができ、柔軟かつ高度な知能に結びつくものと考える。この結果、身体の発達と機能の積み上げということから、結果的にある時期に歩行ができるようなり、ある時期にしゃべれるようになるというように学習に流れができると考える。

従来の知能システムの研究では、認識や動作のプロセスを細分化し、ある時はAという方法、ある時はBという方法をという形で、個々の場合に適した方法を考えるという形で知能化の度合いを上げていったと考えられる。一方、前節で述べたように、実際の我々は、常に様々な要因を統合して認識や動作を行っているように見える。個々の機能を個別に考える時には、そのインターフェイスを規定する必要があるが、これを予め規定してしまうことは、適応性という面から好ましくない。従って、従来のように、個々の機能を個別に考えるのではなく、できるだけ入力から出力を統一的に捉えて学習する方法が望ましい。そのためにも、上記のような段階的な学習が必要になる。

学習によらないで、ある程度発達した状態からスタートさせるためには、それだけ完成度の高いモデルを構築しなければならない。そうすると、結局与えたモデルを越えることはできないし、初めから完璧なモデルを作ることが困難であるため、システムが暴走し、人間に危害を加えることになりかねない。その点、赤ん坊のように、無力な状態からスタートすれば、いきなり暴走する心配はない。無力な状態に設定するには、赤ん坊のように身体的に無力な状態とし身体の発達と共に内部の処理も学習していく方法と、動作出力が0になるように内部の処理のパラメータを設定しておく方法が考えられる。前節で述べたフィードバック＋フィードフォワードの制御の学習の際には、後者の方法でスタートさせ、暴走することなく自然に学習を進めることができた。

制御の分野では、川人らが提案しているフィードバック誤差学習が有名である[Kawato 87]。彼が提案しているアーキテクチャは、上記の筆者が提案したものと似ているが、フィードバック制御の部分は予め備えつけられており、フィードフォワード制御をニューラルネットで学習することを試みている。ここでフィードバック制御を備え付けとしたことは、最初からある程度の機能を備え、かつ、暴

走してはいけないというという束縛によると考えられる。学習とは機能を向上させるためのものであり、赤ん坊のように機能を獲得するものという捉え方が一般的にされていないためであると考える。しかし、これからは学習の比重を増やし、機能の獲得という捉え方をするべきであると筆者は考える。

しかし、これに対し、予め備え付けられるものは備えておくべきであり、全て学習によって獲得することはナンセンスであるという反論がなされる。これはもっともなことである。しかし、予め与えたことによってそれに縛られてはならず、そこからさらに適応する機能を持たなければならないと考える。ニューラルネットで学習させることを考えれば、その初期値として情報を与えることは、そこからさらに学習させることができるという意味で有効であると思う。ただ、あらゆる場面で適応できるということは、結局、無からの学習ができる機能を持ち合わせていなければならないと考える。従って、本論文では、敢えてできる限り無からの学習をさせるということを試みた。

1.3.2　実時間学習

ニューラルネットの学習のさせ方は大きく2つに分類することができる。一つは、一般的に行われる方法で、有限のデータを何回も繰り返し学習させる方法（サンプル学習）でもう一つは、環境から逐次データを持ってくる方法である。前者は、過学習という問題を生じ、学習させるほど、また、中間層ニューロンを増やすほど汎化能力がなくなると言われている。しかし、これはある意味で当然であり、与えられたデータに対して最適にフィッティングしようとするニューラルネットの利点なのである。これに対し、実際の我々生物は、実世界との関わりを通して、常に学習を続けるため、そのような問題は起こらないし、もし、限られた範囲のデータしか取得できないような環境にいるのであれば、それに対して適応すれば十分であるということになる。従って、自律学習を目指す上で過学習を問題とすることはあまり意味がなく、実世界との関わりをしながら学習を行えば良い。また、この場合、次から次へとデータが得られるため、一部のデータを使って何回も繰り返し学習することはあまり意味がなく、1つのデータに対し1回だけ学習する。これは、汎化という点から重要であるだけでなく、リアルタイム性からも重要なことである。ただし、連続時間で考えれば、時間をさかのぼってニューロンの内部状態を元の値にセットし直して再び学習させること自体がナンセンスになるので、この問題は自ずと解決する。

しかし、通常の教師あり学習では、いくつかの入力に対して教師信号を人間が用意しなければならないため、どうしてもサンプル学習になってしまう一方、実際の環境とのインタラクションを通して学習するにせよ、環境から常に実際の確率分布に従ってデータをとってきて学習させるにせよ、存在する空間は連続であるため、無限個の入力データがあることになる。そのすべてに対して教師信号を人間が用意するのでは意味がないので、教師あり学習させるためには、その教師信号を自動生成させる手段が必要となる。

また、システムが変化する環境に適応できるようにと考えると、学習は常に続けることが好ましい。また、こうすることで、1回の学習による変化を小さくし、安定した学習ができると考える。一方、生体の発達過程を見ると、臨界期（感受性期）というものが存在し、例えば、仔ネコの片眼を遮蔽するという実験を行うと遮蔽されていなかった眼により強く反応する細胞が増えるということが報告されており、さらに、この効果は生後5週ぐらいで最大になり、それよりも早い時期でも遅い時

期でもその効果が減少することが報告されている[Hubel 72][Blackmore 76]。また、一般的にも、老化するほど学習能力は劣るという認識がある。後者に関しては、若年期にはできるだけ学習し、老年期にはそこまでに学習した結果をできるだけ活かすという学習戦略は、個体の生存を考えた時に、理にかなったものであると感じる。前者の臨界期の問題に関しては、それ以外の時期には学習が行われないというのではなく、前節で述べたような身体の発達と機能の積み上げによる学習の流れの中で、ある機能が形成できなかったため、結果的にそれ以上の積み木を積み上げることができなくなってしまったのであり、その機能にかかわること全てその時期に出来上がるわけではないと解釈することもできるであろう。以上より、筆者は、老化による学習量の減少はあるものの、学習自体は常に進行させるべきものと考えている。

本論文では、上記のような考え方に基づき、基本的にセンサを通して逐次データを得ることができるものとし、また、学習は一回のみ行うこととした。また、学習に完了はなく、常に学習をし続けるようにした。

1.3.3　遺伝（生得）と環境（学習）

心理学の分野で，人間や動物の持つ様々な行動の発達は遺伝によるものか環境によるものかという問題は，一卵性双生児の研究などを通して昔から活発に議論がなされているようである。最近では，Stern，W により提案された輻輳説，つまり，遺伝と環境の両方が影響を及ぼすものであるという説が有力である[田中 90]．筆者は、これらは複雑に絡み合っているものであり、分類すること自体にあまり意味がないと感じる。

馬などの動物は生まれるとすぐに歩くことができるが、人間が生まれた状態はあまりにも無力である。Portmann は、これを生理的早産と呼び、人間は他の高等なほ乳類と比較して約1年早く生まれていると指摘している[Protmann 44]。しかし、ほとんどの人間は大きくなれば歩くことができるようになるし、その歩き方には大きな個人差は見られない。このことから、人間の歩くという機能に関してはあまり学習の必要性がないように見える。にもかかわらず、人間が生まれた時に歩けない理由として、人間は、(1)遺伝子という限られた情報量を、学習という機能により多く費やしているため、(2)前述のように、より無からの学習をすることによって、より高度な機能を身につけることができるため等を考えている。前述の Portmann も、動物の行動は、環境に拘束され、本能によって保証されていると、我々は簡単に特徴づけることができる。これに対して、人間の行動は、世界に開かれ、そして決断の自由を持つ、と言っていいだろうと述べている。

いずれにしても、馬の歩行と人間の歩行に質的に大きな違いはなく、馬の歩行にしても、けがをしてもそれなりの歩き方ができるなど、学習によるところは大きいであろう。このことから、一見生得的な機能とそうでない機能も実はあまり大差はなく、生得的と見られる機能も、そのほとんどは実は学習によっても獲得できるものではないかと考えることができる。そして、馬が生まれつき歩けるのは、前にも述べたように、ニューラルネットの初期値が与えられていると考えることが自然ではないだろうか。

進化の研究において、個体が学習機能を身につけることで進化が大きく加速し、さらにその後、学習後の機能を予め持ったものが出現するというボールドウィン効果というものが知られている

[Baldwin 1896]。これより、学習機能がいかに重要で、進化とも切っても切れない関係であるかということがわかる。そして、人間は、まさに高度な学習機能を身につけたものであり、進化の過程の中でも、他と機能面で大きく差を付けているということができる。ただし、歩行の問題に関しては、馬が学習後の機能を予め持っているものとすれば、馬と人間の関係がボールドウィン効果の中で逆転していることになる。これはまさに前述の(1)(2)の理由によるもので、より高度な学習機能を付けるために学習後の機能を予め持つことを放棄したという逆ボールドウィン効果とも言えるものではないかと筆者は考えている。

1.3.4　生物のモデルか工学的応用か

本研究は、前にも述べたように、自律学習によって、自律的でかつ柔軟で適応性のあるシステムの構築を最終的な目標としている。自律学習のお手本は生物であり、知識付与という概念を捨てて大胆に生物に学んでいかなければならないと考える。このような意味から、本研究は、生物のモデル化を目標としているのではないが、自律学習という点からできるだけ生物から学ばなければならないと考える。ただ、心理学にしても生理学にしてもまだまだ生物のことが完全にわかっているわけではなく、未知な部分が多いし、正しいと思われていることでも研究が進むにつれて違っていたということもある。従って、各機能に関して、あまり心理学生理学に捕らわれ過ぎることがないようにした。

また、知識付与型の知能システムの場合は、その知識を付与することによって工学的にどのような機能が実現できるかということによってその善し悪しを測ることができる。しかし、本研究は、学習能力の獲得を目指し、最終的に機能実現に結びつけるというアプローチである。従って、まずはどのような機能が獲得されたかではなく、どのような学習機能が獲得されたかという観点で評価をされることが望まれる。従って、工学的応用という面からもあまり捕らわれないようにした。

筆者は、本研究が進むことによって、最終的には生物のモデル化という意味でも、工学的な応用という意味でも成果が大いに期待できるものと考えている。このような立場から、自律学習という機能にしぼって両者の中間をねらっていきたい。

1.4　本研究の目的

本研究では、まず、自律学習という観点、つまり、いかにして自ら学習していくか、また、いかに少ない情報から多くの機能を学習によって獲得するかといった観点からシステムがどのような学習をしていくべきかを考える。この学習を、図1.8のようなシステム構成の中で、センサ信号の処理（センサ情報の統合）、および行動の生成（強化学習）にいかに展開していくかという点にある。

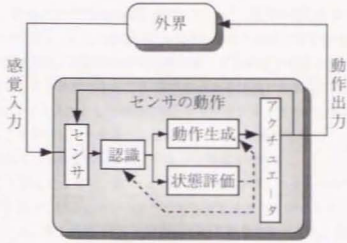


図1.8　本研究で想定した自律学習システムの構成

1.5　本論文の構成

本論文は、学習アルゴリズムと達成する機能の大きな2つの軸を持つ。この序論の後の第2章では、本論文中で用いる主な学習アルゴリズムについて述べる。まず、バックプロパゲーション（BP）法について述べ、その後、BP法を用いることによって簡単に強化学習を行う方法を述べる。その後、筆者が提案している、自律学習を実現する上で必要もしくは有効と考えられる3つの基本的な学習アルゴリズムである、相関情報抽出学習、時間軸スムージング学習、値域拡大学習について述べる。

第3章以降は、第2章で提案した学習アルゴリズムを用いた具体的な応用例へと展開していく。第3章では、相関情報抽出学習と値域拡大学習を組み合わせることにより、複数の情報源からの信号に共通する情報（相関情報）を抽出することを教師なしで学習する方法について述べる。そして、視覚と運動の情報を相関情報を抽出することによって、対象物体との相対位置を抽出することを学習できることを示す。また、抽出する情報が複数次元の場合に、それを直交化する方法について述べる。

第4章では、時間軸スムージング学習と値域拡大学習を組み合わせることにより、局所的な受容野を持つ多数のセンサセルよりなるセンサの信号を統合し、物体の位置などの情報をアナログ値として表現することを教師なしで学習させる方法について述べる。

第5章では、第4章で述べた学習をさらに拡張することにより、視覚システムの学習モデルとして、頭部位置によらない認識、前庭動眼反射、物体追跡の3つの機能を説明できることを示す。これによって、上記の時間軸スムージング学習を使ったセンサ信号の統合の有効性を示す。

第6章では、強化学習を能動認識、つまり、認識および認識のための動作という直接報酬や罰と関係ないものに対しても適用できることを示し、視覚センサの入力を基に認識、および視覚センサの動作を直接教師信号を与えることなく実現する方法を示した。そして、簡単な文字認識の問題を学習させることにより、センサの動作を行わせない場合より、少ない中間層ニューロン数で学習することができることを示した。

第7章では、時間軸スムージング学習の遅延強化学習への適用について述べる。ここでは、遅延

強化学習の評価を、目的達成までの所要時間で行うこととし、所要時間の予測を時間軸スムージング学習で行うというものである。また、所要時間が異なる場合も正確に所要時間を予測するために、時間軸スムージング学習を拡張した評価値の時間変化量一定化学習を提案する。以上を用いて、視覚センサ信号を直接入力としても学習できること、障害物がある場合でもある程度学習が可能であることを示すと共に、視覚センサを入力とした場合には、ニューラルネットの中間層に空間の情報を必要に応じて自己組織することを示し、その自己組織の仕方について検証した。

第8章は、結論であり、本論文の成果および今後の課題を述べる。

本論文は、表1.1に示したように、図1.8で示した自律学習システムのどの部分に適用したかと、提案したどの学習則を用いているかという2つの切り方をすることができる。そして、前者によって章の流れを構成している。第2章は、本論文で用いている学習アルゴリズム全般について概説し、その後、第3章と第4章では、センサ信号の統合・認識の問題、第5章と第6章では認識と認識のためのセンサ動作を絡ませた話し、第7章では、動作生成まで含めた話しという流れとなっている。

第3章と第4章では、（1）複数種の情報源に共通に存在する情報が重要な情報であるという観点と（2）空間情報は時間的に連続にしか変化しないため、それを用いて情報の効率的表現をすべきであるという2つの観点からその学習則を示した。また、第5章と第6章では、（1）より時間的に滑らかに変化する情報を得るように認識や認識のための動作が働くという考え方と（2）認識やセンサ動作を目的達成（強化信号を得る）ための動作として捉える、つまり、認識重点型と動作重点型の2つの考え方に基づく学習則をそれぞれ示した。どちらも同じ機能を実現するための方法が複数存在するが、実際の人間はこのどちらかというよりは、これらの複数の学習方法を組み合わせているのではないかと考えている。

表1.1　本論文における章の構成

	適用部分	用いる主な学習則	内容
第2章	3つの基本学習則の提案		
第3章	認識	相関情報抽出学習 値域拡大学習	マルチセンサまたは センサと運動の情報統合 の学習
第4章	認識	時間軸スムージング学習 値域拡大学習	局所センサ信号の統合の学習
第5章	認識＋センサ動作	時間軸スムージング学習 値域拡大学習	視覚系機能の学習モデル
第6章	認識＋センサ動作	強化学習	能動認識の学習
第7章	動作生成＋状態評価 （＋認識）	時間軸スムージング学習 強化学習	遅延強化学習

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

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

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、

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

$\kappa = k\,\overline{rnd^2}$、　$\overline{rnd^3}$：$rnd\ rnd$ 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。

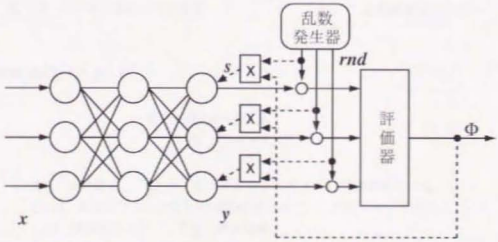


図2.1　ＢＰ法を用いた強化学習の仕組み

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。

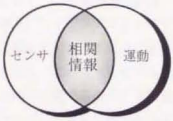


図２.２ センサと運動の相関情報

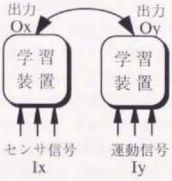


図２.３ 相関情報抽出学習

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。

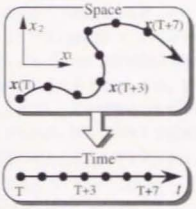


図２.４　空間情報の時間軸へのマッピング

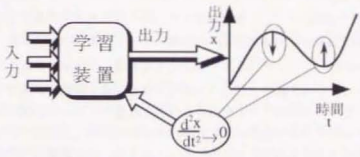


図２.５　時間軸スムージング学習
（矢印は典型的な教師信号を示す。実際は毎単位時間学習を行う）

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。

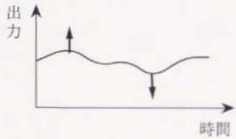


図2.6　値域拡大学習の模式図

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。

第3章　相関情報抽出学習と空間認識モデル

3．1　背景

複数種類のセンサ情報を統合、融合[Luo 89][山崎 90]することにより、単に情報の精度を上げるという量的な効果だけでなく、センサ情報処理の質的な飛躍が望まれている。我々は常に無数のセンサ情報を得ているが、その中から自分にとって必要な情報を抽出し、処理し、適切な行動に結びつけている。この必要な情報の一つとして、前章で複数種類の情報源からの信号の相関情報を挙げ、これを抽出する学習として相関情報抽出学習を提案した。本章では、この学習方法をより具体的に述べると共に、視覚と運動の情報から空間認識の機能を学習によって獲得できることを述べる。

例えば、図3．1に示したように、我々が前に進めば、目に映っている物体は大きくなり、やがて手に届くようになる。そして、ビジュアルフィードバックによって自分の手を物体に近づけることができる。この因果関係、つまりを我々が生活する3次元空間の認識を学習することによって、我々は、食物が目の前に見えたら前に進み、手を伸ばして、食べることができる。この時、我々が生活する空間の情報は、通常、足や手を動かしたという多次元の運動の情報と、網膜上に映る物体が徐々に大きく見えることを反映した多次元の視覚センサからの情報の両者から得ることができる。従って、視覚の情報と運動の情報に共通する情報を抽出することができれば、我々が生活する空間に関する情報を抽出できることになる。また、その情報が両者に共通に存在する可能性が高いことを利用すれば、外部から教師信号を与えることなく学習によってその情報を抽出できると考えられる。

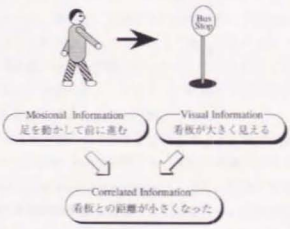


図3．1　運動と視覚の相関情報としての空間情報

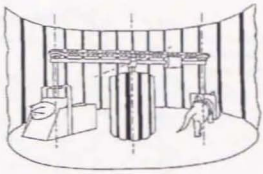


図３.２ Held & Hein の実験（[Held 63]）

図３.２に示したHeldらの実験[Held 63]によれば、２匹のネコを縦縞で囲まれた装置に入れ、片方は自分の意志で動けるようにし、もう片方は相手と全く同じ動きをするようにしてやると、両者は全く同じ視覚体験をしているにもかかわらず、自らの意志で動作できないネコは、装置をはずすとものにぶつかったりしてしまい、正常な空間認識能力を形成することができない。これは、空間認識能力の形成には、視覚の情報だけでなく、自らの意志による運動が必要であるためと考えられる。つまり、複数の情報源からの信号が空間認識機能を学習するために必要であると考えられる。また、逆に、視覚と運動の情報から空間認識能力が形成できる可能性を示している。

また、Aitkenらの実験[Aitken 82]によると、先天的な盲児に対象物との距離や大きさ等を音声情報に変換する装置を取付けると、その盲児は目が見えないにも関わらず、わずか数十回眼前におもちゃを提示するだけでそのおもちゃに手を延ばしてつかむことができるようになる。この結果から、空間認識には必ずしも視覚という特定のセンサが必要なのではなく、また、空間認識能力の獲得が、先天的なものではない、つまり、汎用的な運動－センサ統合学習の存在の可能性を示唆している。

情報の圧縮という観点からは、恒等写像型ニューラルネット（砂時計型ニューラルネットとも言う）が知られている[Rumelhart 88][Zipser 86][入江 90]。この方法では、図３.３のように、多層（通常５層）のニューラルネットを用意し、入力層のニューロン数と出力層のニューロン数を等しくし、真ん中の中間層のニューロン数をそれよりも少なくする。そして、入力データをそのまま教師信号として学習を行うことによって、中間層に圧縮された情報が得られるというものである。

これに対し、Ballardらは、大規模ニューラルネットに適したアーキテクチャとして、複数の恒等写像型ニューラルネットの圧縮された中間層のニューロンの出力をさらに上位の恒等写像型ニューラルネットへ入力すると共に、上位のネットワークでの誤差を下位のネットワークの学習にフィードバックするという方法を提案している[Ballard 87]。

また、片山らは、より生体を意識したモデルを提案している。まず視覚の情報を恒等写像ニューラルネットである視覚ネット(Visual Net)で学習させ、その中間層の出力を体性感覚情報を出力する別のネットワーク(Somato-Sensory Representation Net)へ入力し、出力を求める。そして、その出力に対して実際に得られる体性感覚情報を教師信号として学習させ、その誤差を視覚のネットまで逆伝播させて学習に反映させるというものである[片山 90]。

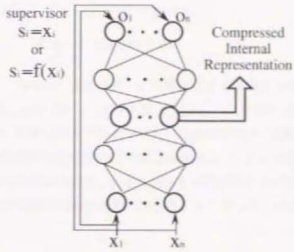


図３.３　恒等写像ニューラルネットワーク（砂時計型ニューラルネット）

Ballard および片山らの方法は、いずれも恒等写像ニューラルネットワークを用いているため、複習種類の情報源から共通の情報だけを抽出することができない、さらに、複数の情報源の内、片方がなくなるとうまく動作しないという問題点がある。

複数種類の情報から、その中に共通に存在する情報を抽出する方法としては、多変量解析の分野で正準相関分析という方法が知られている[田中] [奥野 71]。そして、麻生らは、これを非線形に拡張した非線形正準相関分析を定式化している[麻生 87]。ニューラルネットを用いてこれを近似的に実現する方法としては、筆者と同時期に Becker らによっても提案されている[Becker89]。また、その後も、麻生ら[Asoh 94]、山内ら[山内 95]によって同様な方法が提案されている。これらには微妙に学習方法に違いが見られるが、詳細は 3.2.3 にて述べる。

３.２　相関情報抽出ニューラルネット

3.2.1 相関情報抽出の定義

複数種類の情報の間で共通に存在する情報（以下、相関情報と呼ぶ）を抽出するということを以下のように定義する。まず、与えられた２種類の時間によって変化する情報のベクトルを

$$\boldsymbol{x}(t) = [x_1(t), x_2(t), \ldots, x_m(t)]$$

$$\boldsymbol{y}(t) = [y_1(t), y_2(t), \ldots, y_n(t)] \qquad (3.1)$$

とし、この $\boldsymbol{x}, \boldsymbol{y}$ から次式で表される情報ベクトル

$$r(t) = f(x(t)) = g(y(t)) \qquad (3.2)$$

f, g：ある関数ベクトル

を抽出することとする。これは言い換えれば、$\boldsymbol{x}$ の関数と $\boldsymbol{y}$ の関数が t の変化によらず等しくなるような関数の組 $(f_1(\boldsymbol{x}(t)), g_1(\boldsymbol{y}(t)))$ 、$(f_2(\boldsymbol{x}(t)), g_2(\boldsymbol{y}(t)))$ ... を求めるという問題になる。 例えば、図３.１のような場合、我々が足や手を動かすという情報を $\boldsymbol{x}(t)$、各網膜細胞の出力を $\boldsymbol{y}(t)$とすると、それぞれの多次元情報の相関情報 r（この場合はスカラー）として物体との距離が考えられることになる。そして、この学習が進めば、必要な情報を抽出することができるようになると共に、式（３.２）から、複数種の情報の内の一種類の情報のみを与えれば、その上位の情報を想起できるようになる。

3.2.2 相関情報抽出ニューラルネットの構成と学習法

前節で定義した相関情報を外部か教師信号を与えることなく学習によって抽出する方法を考える。これを行なう相関情報抽出ニューラルネットは、図３.４のように、２つの階層型ニューラルネットによって構成され、それぞれ微小な乱数によって結合の重み値が決定されているものとする。そして、それぞれのニューラルネットに、別々の種類の情報を、例えば、片方に視覚に関連する情報、残りの片方に運動に関連する情報といった具合に入力する。そして、片方のニューラルネットの出力をもう片方のニューラルネットの教師信号として、互いに出力を交換しあい、両者の出力が同じ値に近づくようにバックプロパゲーション（ＢＰ）法[Rumelhart 86]に基づいて学習を行なう。これによって学習が収束し、誤差が０に近づけば、両ネットワークの出力の値は入力の値によらず常に一致しているはずである。また、当然各ネットワークの出力値はそのネットワークへの入力値の関数となっている。よって、その出力値は式（３.２）で表されるような両ネットワークへの入力値の相関情報ということになる。

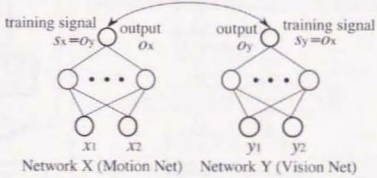


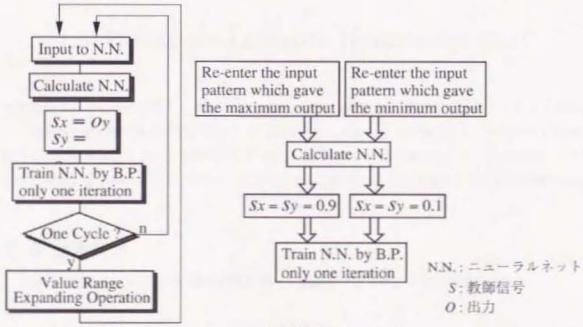
図３.４　相関情報抽出ネットワークの構成と学習方法

しかし、実際に学習を行なうと、

$$f(x(t)) = g(y(t)) = const. \qquad (3.3)$$

とどんな入力を入れても同じ値しか出力しなくなってしまう。これは確かに式（3.2）を満たす上、ニューラルネットにとっては入力の値によらず一定値をとるという学習が非常に容易であるためと考えられる。しかし、これでは情報を抽出したとは言えない。そこで、このような状態を避け、できるだけニューロンの値域を有効に使用するため、2.3節で述べた値域拡大学習を行なう。

まず、入力パターンを変えながら相関情報抽出学習を行うが、この入力パターン何個分かを１サイクルと定義する。そして、１サイクルに一回、そのサイクル中で両ネットワークの出力の和が最大であった時の入力パターンと最小であった時の入力パターンを記憶しておいて再度入力し、出力の値域が０から１の場合には、それぞれ０.９、０.１の教師信号を与えて学習させる。それが終了した後、再び相関情報抽出学習を始める。本論文では、１サイクルを１００パターンとした。１サイクルのパターン数が多いと、値域拡大学習の適用回数が減るため学習が遅くなり、また、パターン数が少ないと、サイクル毎に最大値、最小値が変動したり、相関情報抽出学習の適用回数が減って学習が遅くなってしまう。従って、両者のバランスをとることが必要である。また、前サイクルでの最大値、最小値の際の入力パターンを保持しておき、１サイクル終了後その入力パターンから出力を再び計算し、今サイクルの最大値、最小値と比較して、より大きい（小さい）方に値域拡大学習を適用するという方法も可能である。これにより、よりグローバルに値域拡大学習を行うことが



(a) 主なフロー　　(b) 値域拡大学習のフロー

図3.5　相関情報抽出学習のフロー

できる反面、特殊な入力があるとそれにずっと左右される等柔軟性に欠けるところがある。そこで、前サイクルのデータに減衰率を掛けたものと今サイクルの最大値、最小値と比較するという方法も考えられる。ただし、本論文では、固定環境のみを扱っているため、前サイクルの値の保持は行っていない。フローチャートを図３.５に示す。ここでの学習は全てＢＰ法を利用するが、教師信号は全てシステムの内部で生成されるため、システムの外からは教師信号を与える必要はない。

3.2.3 類似学習則との比較

３.１節で述べたように、本学習の提案と同時期、およびその後に類似した学習則がいくつか提案されている。ここでは、これらの学習則を比較する。

まず、学習が逐次的であるかどうかで大きく２つにわけることができる。筆者の方法と山内らの方法は逐次的、Beckerらおよび麻生らの方法は予めデータセットを用意するという方法である。

山内らの方法は、３.１節で述べた恒等写像ニューラルネットを利用しており、複数の中間層の出力を平均するモジュールを設け、その平均値に各中間層の値が近づくように、恒等写像ニューラルネットと同時に学習を行う。また、値域の確保および複数の相関情報を抽出する際の個々の相関情報間の直交化は恒等写像ニューラルネットの学習を行うことで実現できるようになっている[山内95]。しかし、恒等写像を実現するためにネットワークが肥大化すること、および恒等写像ニューラルネットにおける誤差が抽出する情報にどのように影響するかはっきりしないなどの問題が残る。特に、視覚イメージなどを入力とした場合には、統合した情報から再び視覚イメージを再現しなければならず、統合に直接関係しない恒等写像の復元がうまくいかない可能性が大きい。

Becker らの方法は、与えられたデータセットについて、相互情報量 $I_{a,b}$

$$I_{a,b} = -\left[\int p(a)\log p(a) + \int p(b)\log p(b) - \iint p(a,b)\log p(a,b)\right] \quad (3.4)$$

を最大化するという方法をとっている[Becker 89]。また、麻生らは、片側のニューラルネットの出力を、分散共分散行列が単位行列になるように正規化し、それをもう片側のニューラルネットの教師信号として学習を行うという方法を採っている[Asoh 94]。いずれの場合も、データを予めデータセットを用意しなければならず、逐次的、リアルタイムの学習へそのまま適用することは困難である。

3.3 基本実験

相関情報抽出ネットワークの基本機能を調べる実験を行なった。まず、相関情報を

$$r = x_1 + x_2 = 3\,y_1\,y_2 \quad (3.5)$$

として、２つのネットワークにそれぞれ x_1、x_2 および y_1、y_2 を与えて学習を行なった。ただし、入

力値はx_1、x_2、y_1を０から１の間の値を乱数で定め、その３個の値を式（３.５）に代入して y_2を計算し、これが０から１の間の値になるまでx_1、x_2、y_1を決め直した。この時の相関情報 r に対する学習後の両ネットワークの出力値の平均値（今後単に出力値と呼ぶ）をプロットしたものを図３.６に示す。これから、ネットワークの出力は相関情報 r に対してほぼ一意に決定しており、相関情報が抽出できたと言える。また、この時の学習によって各サイクル毎に求めた両ネットワークの出力値を足したものの最大値、最小値をプロットしたものを図３.７に示す。また、この時の誤差（値域拡大学習時の誤差は除く）つまり両ネットワークの出力値（０から１の間の値）がいかに近づいたかを示す値の変化の様子を図３.８に示す。両図をあわせて見ると、学習の初期には両ネットワークともでたらめな値を出力し、誤差も大きいが、その後時間が少し経過すると、両ネットワークの出力とも０.５に近づく。これは、値域が拡大される前、つまり、入力の値が変化しても両者の出力の値はほとんど変化しない状態で、値域拡大学習の教師信号0.9と0.1に対する誤差が最も小さくなるのが0.5であるからだと考えられる。時間が経過していくと、値域拡大学習によって徐々に最大値と最小値が離れ、それに伴って両出力の値は一旦離れて誤差が大きくなっていくが、さらに学習が進むと再び誤差は０に近づいていくという大まかな動きがわかる。

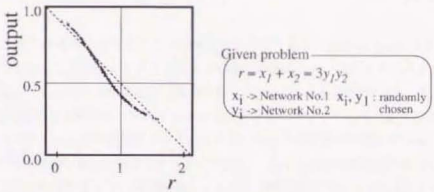


図３.６　基本実験の結果

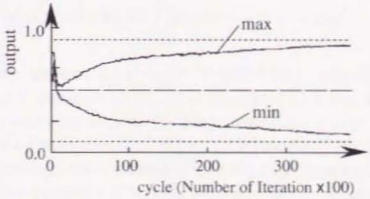


図３.７　学習による出力の最大値と最小値の変化

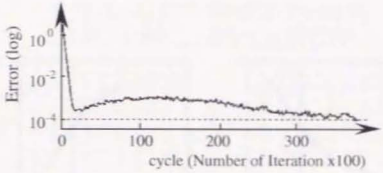


図３.８　学習による誤差（両ネットワークの出力の差）の変化

次に、ネットワークが学習した時の相関情報の分布と実際の出力の値域の分布との関係を詳しく見るため、入力を１個として次の式ような相関関係を持たせて学習を行なった。

$$r_1 = x = 0.1 / y$$

$$r_2 = 0.1 / x = y \qquad (3.6)$$

ここで、式（３.６）の両式における x と y の関係は対称的であり、$r_1=x$ も $r_2=y$ も共に x と y の間の相関情報である。この場合 x を決めれば y が自動的に決まるが、両ネットワークへの入力データが対称になるように x と y の値を交互に乱数を用いて 0.1 から 1.0 の間で決定し、入力データを生成した。これを学習させた時の r_1 と r_2 のそれぞれに対するネットワークの出力値を図３.９に示す。ここでは、ニューラルネットの初期値を変えた１０回の出力分布を重ねて描いている。これを見ると、$x = y = \sqrt{0.1}$ の時に出力はほぼ 0.5 となっており、r_1 と r_2 の対称が保存されていると考えられる。また、この対称性を崩して、奇数回時に $x = 0.1$、偶数回時に $1.0 > x > 0.5$ とするという分布で入力パターンを生成する。ＢＰ法の定義から、入力パターンに対する誤差とそのパターンの出現確率を掛けたものを小さくするように

$$\frac{1}{2}\int_0^1 p(x_1)\,(o_2 - o_1)^2 dx_1 = \frac{1}{2}\int_0^1 p(x_2)\,(o_1 - o_2)^2 dx_2 \rightarrow \min. \qquad (3.7)$$

と学習する。この時、奇数時の入力は $x = 0.1$ という一点だけであり、$0.5 > x > 0.1$ のデータの分布が０であるのに対し、偶数時には $1.0 > x > 0.5$ と入力に幅がある。従って、誤差を小さくしようとすると、$1.0 > x > 0.5$ での出力の値域が小さくなり、結果的に図３.１０のように、$x = y = \sqrt{0.1}$ での出力値が 0.5 よりわずかながら大きく、つまり、入力が 1.0 の時の出力値に近くなっている。これは、入力の分布にいくつかのかたまりがあれば、そのかたまり内での値域は小さくなり、かたまり間の値の差が大きくなる傾向にある。つまり、かたまりを分類する方向に働くと言える。

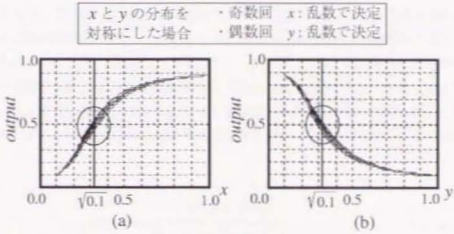


図３．９　入力データに対称性を持たせた場合の出力の分布

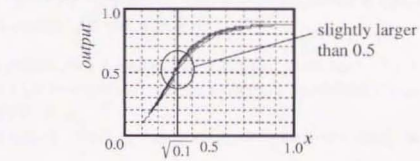


図３．１０　入力データを非対称にした場合の出力の分布

３．４　複数次元の相関情報の抽出

前節までは、抽出する相関情報の次元が１次元の場合について述べた。本節では、これが複数次元になった時、つまり、式（３．２）の相関情報 $\boldsymbol{r}$ が

$$\boldsymbol{r}(t) = f(\boldsymbol{x}(t)) = g(\boldsymbol{y}(t)) \qquad (3.8)$$

のようにベクトルになった場合にどのように抽出するかを述べる。

この場合、抽出したベクトル $\boldsymbol{r}$ の各成分が直交していることが望ましい。予めデータセットが用意されている場合は、非線形正準相関分析の考え方を用いれば比較的簡単であるが[Asoh 94]、ここでは、これを逐次的にかつ近似的に学習する方法を述べる。

直交させるということは、ある成分の分散が大きい時に他の成分の分散が小さくなるように学習すれば近似的に実現できる。抽出する情報が１次元の時は、過去何回かの学習の中で出力値が最大のもの、最小のものについて値域拡大学習を適用したが、ここでは、ある成分の分散が大きく、かつ他の成分の分散が小さい時に値域拡大学習を行うこと（複数出力の直交化学習）でこれを実現する。ただし、ここでは、逐次的に行いかつ計算を簡単にするため、分散の計算の際に、平均値の代わりに前サイクルでの出力の最大値と最小値の平均値を用いた。具体的には、i 番目の出力の独立度 ind_i を

$$ind_i(t) = \frac{\left(out_i(t) - mid_i\right)^2}{\sum_{j=1} \left(out_j(t) - mid_j\right)^2 + \theta} \qquad (3.9)$$

$out_j(t) = \dfrac{\sum_{n=1}^{N} o_{n,j}(t)}{N}$ ：N 個のネットワークの j 番目の出力の平均値、

$o_{n,j}$：n 番目のニューラルネットの i 番目の出力値、N：ニューラルネットの数

mid_i：前サイクルでの N 個のニューラルネットの i 番目の出力の最大値と最小値の平均値、

θ：微小な定数（ここでは、0.001 とした）

と定義し、出力の偏差、つまり、$out_i(t)$-mid_i が正の場合と負の場合について、それぞれこの値の最大値をとる入力パターンを再入力し、偏差が正の場合には0.9、負の場合には0.1の教師信号を与えて値域拡大学習を行う。

例えば、２つのニューラルネットを設け、それぞれの入力を２つの相関情報の線形結合として、

$$in_{1,1} = (r_1 + r_2) / 2.0$$

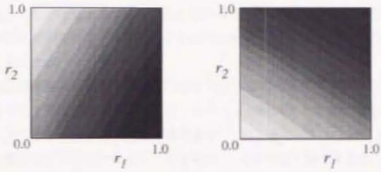
$$in_{1,2} = (r_1 - r_2 + 1.0) / 2.0 \qquad (3.10)$$

$$in_{2,1} = (2\, r_1 + r_2) / 3.0$$

$$in_{2,2} = (r_1 - 2\, r_2 + 2.0) / 3.0 \qquad (3.11)$$

$in_{n,i}$：n 番目のニューラルネットの i 番目の入力値、　r_i：相関情報

のように定め、r_1、r_2 を０から１の間の一様乱数によって決定して学習を行った。図３.１１に、それぞれの相関情報に対する出力の分布を表す。これより、出力１と出力２は、ほぼ相互に直交した成分に反応していることがわかる。



(a) 出力１　　(B) 出力２

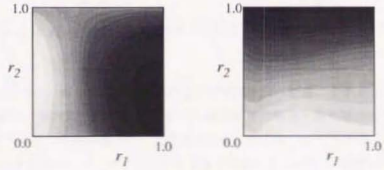
図３．１１　相関情報r_1、r_2に対する２つの出力の分布
（それぞれ、２つのネットワークの出力の平均で、色の濃淡が値の大きさを示す）

また、入力と相関情報の間に非線形の関係を設けると、例えば、１つめのニューラルネットは前のシミュレーションと同じ（（３．１０）式）とし、２つめのニューラルネットの入力を

$$in_{2.1} = r_1 \cdot r_2$$

$$in_{2.2} = r_1 / (r_2 + 0.1) / 10.0 \qquad (3.12)$$

とすると、図３．１２のように、出力分布はほぼ直交していることがわかる。



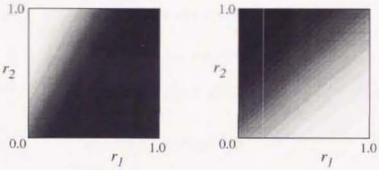
(a) 出力１　　(b) 出力２

図３．１２　相関情報と入力の関係が非線形である場合の相関情報r_1、r_2に対する２つの出力の分布

また、次に、我々が３次元の空間を認識する場合、前後、左右、上下という形で、単に直交しているというだけでなく、それぞれ意味のある情報となっている。この場合、前後、左右、上下に動こうとした場合、それぞれで運動の仕方がかなり違うこと、重力の影響で物体は上から下に落ちる

し、我々は前は見えて後ろは見えないが、左右は基本的に対象である等かなり意味の違う情報となっていることが原因と考えられる。しかし、それ以外に抽出した情報の各成分に意味付けをすることができないかを考えた。

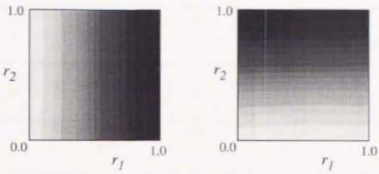
まず始めに、我々の運動が、前後、左右、上下という方向にのみ変化する場合が多いことから、データの分布を変化させることによって各出力ニューロンが抽出する主成分方向が変化するのではないかと考えた。そこで、入力が相関情報の線形結合（(3.10)、(3.11)式）の場合について、データの分布を一様分布とするのではなく、ある時は、r_1を中間の値である0.5に固定し、r_2を0から1の間で変動させ、ある時は、逆にr_2を0.5に固定し、r_1を0から1の間で変動させた。しかし、この場合、出力間の直交化すらできなかった。これは、データの分布が2つの直線上に乗っているため、r_1-r_2平面でのデータの分布は連続ではないため、2つの出力が一旦同じ相関情報を抽出するようになってしまうと、両者が徐々に直交化していくことができないことが原因と考えられる。また、r_1のみ変化させる確率およびr_2のみを変化させる確率をそれぞれ40%とし、r_1、r_2共に変化させる場合を20%とした場合も同様の傾向が見られた。



(a) 出力1　　(b) 出力2

図3.13　r_1、r_2のどちらかを0.5に固定した場合の2つの出力の分布

次に、データの分布は一様とし、複数出力の直交化学習を行った後、1サイクル中の各出力の最大値、最小値を出力した時の入力値に対し、出力の偏差が正の場合と負の場合のそれぞれについて独立度の最大値をとる出力を教師信号として学習を行ったところ、図3.14のように、2つの出力がそれぞれr_1、r_2をきれいにコーディングできるようになった。また、この時、データの分布をr_1、r_2による矩形状ではなく、$r_1^2+r_2^2<1.0$という円形にしたところ、再び2つの出力は直交関係を保っているが、それぞれr_1、r_2をコーディングするということはなくなった。これは、矩形の領域の場合、最大・最小値は領域の頂点に存在し、独立度の最大・最小値はその頂点を含む辺上にある可能性が高い。そして、頂点の値をその辺上の一点の値に近づけるという学習を行うことによって、その辺上の全ての値が等しい値になるためと考えられる。データの領域を長方形としても、同様に各編の方向の成分を出力として抽出することができるようになった。また、この際に、長軸方向がすぐに抽出され、全体的に正方形の場合と比較して収束は早かった。



(a) 出力１　　(b) 出力２

図３．１４　出力間直交化学習の後、各出力の最大値、最小値を出力した時の入力値に対し、それぞれ独立度の最大値、最小値をとった時の出力を教師信号とした学習を行った場合

さらに、これが相関情報の次元が３次元の場合でも有効かどうかを確かめた。そこで、３つの相関情報に対し、各ネットワークにそれぞれ次のような３つの入力を生成し、学習を行った。

$$in_{1,1} = (2\,r_1 + r_2 + r_3) / 4.0$$

$$in_{1,2} = (r_1 + 2\,r_2 + r_3) / 4.0$$

$$in_{1,3} = (r_1 + r_2 + 2\,r_3) / 4.0 \qquad (3.13)$$

$$in_{2,1} = (r_1 + r_2 + r_3) / 3.0$$

$$in_{2,2} = (r_1 + r_2 - r_3 + 1.0) / 3.0$$

$$in_{2,3} = (r_1 - r_2 + r_3 + 1.0) / 3.0 \qquad (3.14)$$

その結果、各相関情報に対する３つの出力は、図３．１５のように、各出力がそれぞれ１つずつの相関情報をコーディングするようになった。このことから、データの分布領域が矩形の場合は、最大および最小出力値の入力パターンに対し、独立度が最大、および最小の場合の出力値を教師信号として与えることによって実現できることがわかった。

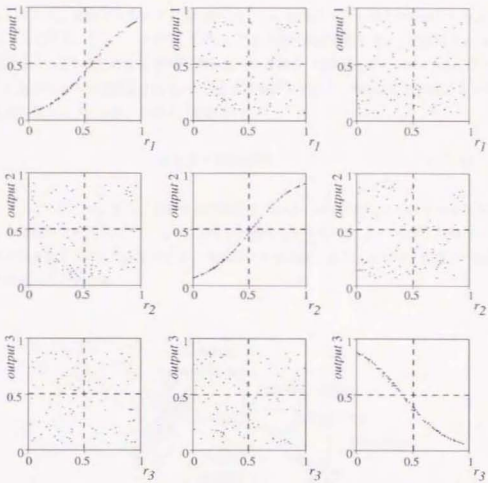


図3.15　相関情報が3次元の場合の学習結果

3.5 空間認識のモデルとシミュレーション実験

3.5.1 空間認識モデル

空間認識は、3.1節でも触れたように、視覚の情報と運動に関する情報の相関情報を抽出することによって得られるものと考える。そこで、相関情報抽出ニューラルネットの2つのニューラルネットの片方（動作ネット）に運動の情報を、もう片方（視覚ネット）を視覚センサから得られた情報をそれぞれに入力する。そして、前章の学習アルゴリズムに従って学習を行なうことにより、空間認識能力が教師なし学習によって形成されると考える。

そこで、まず図3.16のような視覚センサ付き移動ロボットを考える。このロボットは、X,Φ という2つの駆動部分を持つが、視覚センサは固定された棒の上を1次元の運動をするものとする。また、視覚センサは、左右2つのセンサセルよりなり、それぞれ受容野に対し、物体の占める面積の割合を0から1の間の値で出力するものとする。さらに、センサと物体は左右に z だけずれるも

のとする。そして、動作ネットには X,Φ を入力とし、視覚ネットには２つのセンサセルからの出力 e_1, e_2 を入力する。そして、学習は、X,Φ,z の３つの変数を乱数を用いて決定し、e_1, e_2 はその値から決定する。それぞれの変数の値の領域は、$4.0<X<8.0$, $-\pi/4<\Phi<\pi/4$, $-1.0<z<1.0$ とする。ただし、リンク部からセンサが固定されている棒までの距離を２とし、物体までの距離 d は２から10までの間の値になるようにした。つまり、距離 d は

$$d = X + 2\tan(\Phi) \qquad (3.15)$$

で求められるようにした。また、投影された物体は、$d=2$, $z=0$ の時に２つのセンサセルのそれぞれの半分の領域を占めるものとし、$z=-1$ の時に左側のセンサセルに、$z=1$ の時に右側のセンサセルにちょうど収まるようにした。また、d と物体の大きさは反比例し、従って、投影された物体の面積は d の自乗に反比例する。

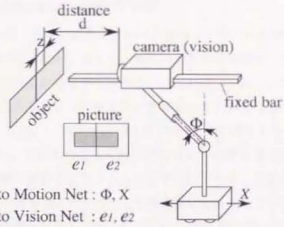
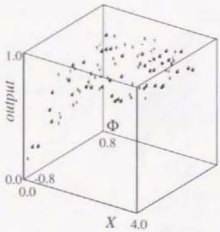


図3.16　シミュレーションで仮定した視覚センサ付きロボット

学習後のニューラルネットワーク（動作ネット）に乱数で決定した２００個の X, Φ を入力した場合の出力を図3.17に、距離 d に対する動作、センサの両ニューラルネットの出力を図3.18に示す。これより、出力は距離 d に対してほぼ１対１の対応関係を形成しており、X, Φ に対しては１対１の対応がとれていないことがわかる。また、距離 d に対する出力の分布が距離 d が小さい時ほど距離 d に対する出力の変化が大きくなっており、図中の距離の自乗に対する反比例の曲線と近い形となっている。これは、前述のように、視覚センサ上での物体の面積（視覚ネットへの入力）の変化は、距離の自乗と反比例するため、出力がそれに引きずられたためと考えられる。



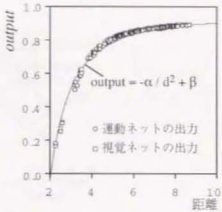


図３.１７ 学習後のX, Φ に対する出力の分布　図３.１８ 学習後の距離d に対する出力の分布

3.5.2 ステレオ画像からの物体との距離の学習

ここでは、図３.１９のような、２つの視覚センサーを持つロボットを仮定する。そして、２つの視覚センサからステレオ画像を得ることによって、物体の大きさ等が変化しても、相関情報としての物体との距離の情報が抽出できるかどうかを調べる。このロボットは、２つの駆動部 Φ、X を持ち、６個ずつのセンサセルを持つ左右２つの視覚センサによって、計１２個の視覚センサ信号が得られる。そして、物体の長さ l、幅 t、水平方向の位置ずれ z が可変であるとする。そして、Φ、X 、l、 t、z を乱数で決定し、その時のそれぞれの視覚センサに映る像を計算し、各網膜細胞はその受容野中に映る物体の面積の割合を出力するものとする。そして、運動の信号であるΦ、X を動作用のニューラルネットに、視覚センサからの１２個の信号をもう片方の視覚用のニューラルネットに入力し、前述のような学習を行なう。Φ、X、l、t、z は乱数を使って毎回変化させて、学習を繰り返す。各変数は、Φ が$0\sim\pi/4$、X が$5.0\sim8.0$、l が$2.0\sim4.0$、t が$0.5\sim1.0$、z が$-1.0\sim1.0$ の範囲とし、距離d は（３.１５）式を用いて計算した。そして、視覚センサの視野を左右 45 度ずつとし、各センサセルの視野は 15 度とし、２つの視覚センサは距離 2.0 だけ離れているとした。また、距離d が 5.0 で t が 1.0 の時に視覚センサの高さと投射された物体の高さが一致するようにした。

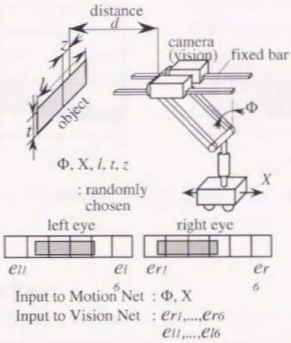


図3.19　2つの視覚センサを持つロボット

結果を、図3.20、図3.21に示す。図3.20は、Φ、X、l、t、zを色々と変化させた場合の物体とセンサまでの距離と両ニューラルネットの出力との関係を示した。Vision Net の出力の方が、出力のばらつきが多少あるものの、両ニューラルネットの出力とセンサ-物体間の距離dがほぼ1対1の対応がとれていると言える。図3.21は、Xおよびlと両ネットワークの出力の平均値との関係を示した。ここでは、Xまたはlと出力の間に1対1の対応関係はない。これらのことから、両ニューラルネットの出力は、物体の大きさや、個々のセンサ入力によらない、物体とセンサとの距離を学習を通して抽出することができるようになったと言える。

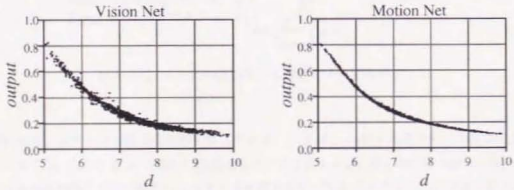


図3.20　ロボットと物体との距離dと両ニューラルネットの出力

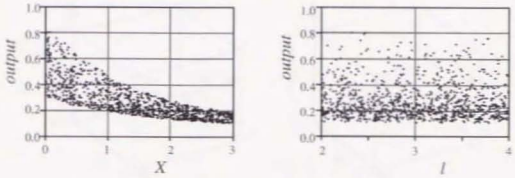


図３．２１　ロボットの駆動部X および物体の長さlと両ニューラルネットの出力

3.5.3　２次元の相対位置抽出

次に、ロボットが２次元の運動をする場合について、シミュレーションを行った。ロボットは、図３．２２のように、x,y,Φ の３つの駆動部分を持ち、前後と上下の２次元の運動を行う。また、視覚センサは２つの視細胞を上下に並べた構成とした。そして、３つの駆動部分の情報を運動ネットに、２つの視覚センサの情報を視覚ネットに入力し、それぞれのニューラルネットの出力を２つ設け、3.4節で説明した独立度を用いた学習を行った。

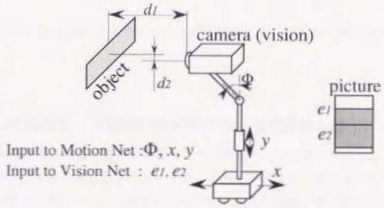


図３．２２　２次元の運動をする視覚センサ付きロボット

学習の結果、ロボットと物体の相対位置（前後 d_1、上下 d_2）に対する各出力の分布は図３．２３のようになった。これを d_1 および d_2 へ投射したデータを見ても、１対１の対応は得られないが、それぞれを６０度回転させた方向からこのデータを読めると、図３．２４のように、ほぼ１対１の関係が得られていることがわかる。

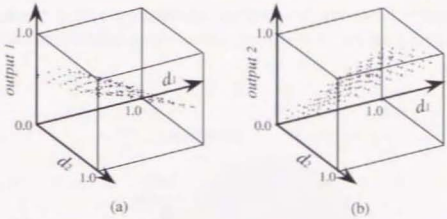


図３．２３　ロボットと物体の前後、上下の位置に対する各出力の値の分布

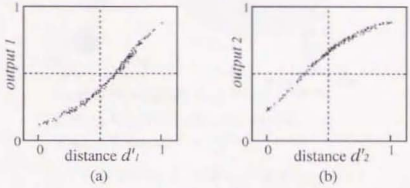


図３．２４　d_1, d_2軸のそれぞれから６０度回転させた方向より見た出力の分布

3.5.4 物体との接触検知（３種類の情報源からの相関情報の抽出）

次に、運動の情報と視覚の情報以外に触覚センサの情報が入った場合を考える。ここでは、図３．２５のようなロボットを考える。このロボットは、２次元平面上に存在するものとし、固定された物体が図の網目で表されたロボットの胴体である円状の部分に接触、または内部に侵入した場合に触覚センサが１を出力し、接触していない場合は０を出力するものと仮定する。そして、３つめのニューラルネットを用意し、そこに触覚センサの信号を入力する。そして、１番目のニューラルネットの出力を２番目のニューラルネットの教師信号とし、２番目の出力を３番目の教師信号、そして３番目の出力を１番目の教師信号として与えることによって学習を進める。この場合、触覚センサの出力が１つしかなく、かつ２値であるため、実際には、この値を教師信号として運動と視覚のニューラルネットを学習させているような状況になる。この学習によって、運動と視覚のニューラルネットの双方で、物体との接触の検知ができるようになった。また、ここで、運動と視覚の入力は連続値であるため、接触するかしないかのあたりでは、出力が、０から１の領域に対して、0.5付近の値を出力してくる。このことは、例えば我々が視覚や運動の情報だけで接触の検知をする場

合、明らかに離れているような場合は接触しないと判断できるが、微妙なところでは判断が難しく、例えば、車に乗っていて実際にはぶつからない場合でもぶつかりそうだと感じることと似ていると考えられる。

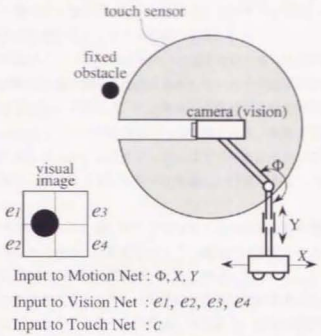


図3.25 接触センサ、視覚センサを有するロボット

3.6 考察および今後の課題

本学習では、複数のニューラルネットの出力が同じになるように学習を進めることによって、ある情報源からの入力が欠如した場合にでも他の情報源からの信号を用いて出力を求めることができるという利点がある。しかし、この得られた出力をさらに上位の機構に渡すと考えた時に、複数の出力をどのように処理するか、また、複数の情報源のうちいくつかの情報が欠如した場合に、そこからの情報をマスクし、正しく得られた情報源の出力をどのようにして選択していくかという点に関しては触れておらず、今後の課題である。山内らのように、複数の出力の平均をとるという方法[山内 95]では、一部の入力情報が欠如した場合にうまく情報を表現できないことになる上、その中から一つの情報を選択するにも、はっきりとした選択基準が見あたらない。これに関しては、現在核情報源からの情報が正常に得られているかいないかを何らかの方法で検出し、それによって出力にマスクをかけるという方法が考えられる。ただ、この時係数の合計が1になる必要がある。このように、入力のマスクをするような処理および学習には、ニューロンの積和演算の前に、入力信号同士の掛け算を行うシグマ－パイユニット[Rumelhart 88] の導入が有効であると考えられる。

さらに、ここで用いられている値域拡大学習や出力間直交化学習は、各サイクル毎に、そのサイクルで最大値および最小値を出力した入力パターンを再入力してそれぞれ0.9、0.1という教師信号で学習させている。しかし、実際は、ニューラルネットは連続時間の中で動作しているはずであり、このようにニューラルネットの動作をサイクルで分断し、過去の入力をニューラルネットに再入力することは非常に無駄が多い。値域拡大学習を通常の学習の中に組み込むこと、または、上位の機構からのフィードバック信号から値域を確保することによってサイクルといった人為的なものを考慮しなくてもいい学習方法を考えていくことが必要である。

また、第4章で、時間軸スムージング学習を用いた局所センサ信号統合の学習について述べる。筆者は、本章で述べた機能と次章で述べる機能は共存しているのではないかと考えているが、そのための方式を考える必要がある。本章では、センサからの信号が各単位時間毎にランダムに変化するという仮定で行った。しかし、この仮定は非常に不自然であり、次章で述べるように、通常、センサからの信号は時間と共に滑らかにしか変化しないはずである。センサ信号が時間と共に滑らかにしか変化しないとなると、サイクルをどう扱うか等難しい問題がある。この辺をいかに取り扱うかは今後の課題である。

次に、実際の脳との比較を考える。脳における連合野を含む大脳皮質の構造を見てみると、全体的に層構造になっており、その中は、錐体細胞のように縦に伸びるものやカハール水平細胞のように横方向に伸びるものおよび星型の顆粒細胞等によって構成されている。そして、外部とは、異なる皮質領域間を結ぶ連合線維や皮質からさらに内部に伸びる投射線維等によって結ばれている。この構造と本論文で提案しているネットワークの構造の類似から、錐体細胞が階層型ニューラルネットに相当し、カハール水平細胞、連合線維および投射線維が上記に示したような複数のネットワークの出力を相互にやりとりする機構に相当し、空間認識やセンサ情報の抽象化の役割を担っている可能性があるのではないかと考える。

これに対し、サルの頭頂連合野では、触刺激やいろいろな関節運動の組み合わせに反応するニューロンがある領域や能動的な運動や視覚性の刺激の組み合わせに反応する領域があることが知られている[酒田 76]。そして、後者は、空間の認知に関与していると考えられている[酒田 82][津本 86]。また、生まれて直後のサルの目を縫い合わせると、正常のサルと比較して、頭頂連合野において視覚刺激や視覚刺激と体積感覚刺激の両方に反応していたニューロンの数が極端に経ることも報告されている[Hyvarinen 81]。

さらに、我々の概念の形成過程を考えると、複数のセンサ情報の間で相関する情報を抽出するという形でセンサ統合が行なわれているのではないかと考えられる。例えば "ピアノ" という抽象化された概念は、図3.26のように、ピアノの形、ピアノの音、「ピアノ」という字の形、「ピアノ」という発音等の複数の要素から形成されており、ピアノを前にしている時にピアノの音を聞いたり、「このピアノは...」等の言葉を聞くことによって、それらの情報の相関情報としてピアノの概念が学習によって形成されていくものだと考えられる。そして、我々は、ピアノの音を聞いても、また、ピアノという文字を見ても "ピアノ" という概念を想起できる。このように、学習後、単一のセンサ入力からその上位の情報を想起できるという点も、本論文中の相関情報抽出ネットで説明することができる。

最後に、山内らは聴覚と視覚の情報から母音の認識への応用を行っているが[山内 96]、上記の概

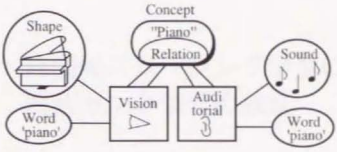


図3.26　概念形成のモデル

念形成モデルの実証など、さらにアプリケーションを見つけることも重要である。また、本方法で抽出した複数種の情報に対し、さらに相関情報を抽出する等してより抽象的な概念形成に結びつけていくことも重要であると考える。

3.7 まとめ

異種情報源の情報から相関情報を抽出することをニューラルネットによって学習する方法を提案した。これは、複数のニューラルネットを用意し、異種情報源の信号をそれぞれのニューラルネットに入力し、その出力を他のニューラルネットの教師信号として与えるという相関情報抽出学習と値域拡大学習を組み合わせた簡単な学習で実現できる。

また、この学習では、値域拡大学習によって出力の値域が0.1から0.9の間になるように学習を行うが、それ以外の出力値に関する規定は明示的に与えていない。そこで、学習後の相関情報に対するニューラルネットの出力の分布を調べたところ、入力データにおける、相関情報の分布に偏りがある時は、入力データの分布密度が大きいところほど出力の値域が小さくなる傾向があることがわかった。

また、抽出する相関情報が複数の場合に対し、各出力に対する独立度を定義し、独立度の大きいものに対し、値域拡大学習を適用する方法を提案した。そして、入力情報と相関情報の間に線形性が強い場合、きれいに直交化された複数の相関情報が抽出できることがわかった。また、相関情報の分布領域がn次元立方体の場合に、その領域内の各出力の最大値および最小値を独立度の最大値、最小値の場合の出力値に近づけるという学習を追加することにより、立方体の各軸に対応する相関情報を各出力として抽出することができた。

また、動作の情報と視覚の情報をそれぞれのニューラルネットに入力して学習させることによって、物体との距離を抽出させることができた。また、視覚センサを複数個用意し、その情報を視覚ネットに入力することにより、物体の大きさなどによらない距離の情報を教師なしの学習によって抽出できるようになった。また、ロボットの動きが2次元になった場合、物体とロボットの2次元の相対位置の情報を抽出することができた。そして、抽出した2つの成分が直交していることがわかった。さらに、触覚センサ入力を追加し、3つのニューラルネットを用いて学習することにより、視覚情報や運動情報からも物体との接触の検知を学習によって獲得できることがわかった。

第4章　時間軸スムージング学習に基づく局所センサ信号の統合

4.1 背景

画像理解の分野において、画像情報の空間的な滑らかさという拘束を設けることによって、不良設定問題の正則化が行われている。また、我々生物の視覚システムにも、フィリングインという機能があると言われている。これは、情報が得られない空間的な部分が存在した場合、そのまわりの情報からその部分を補間する働きであり、例えば、網膜上の盲点は通常画像の情報を得ることができないが、我々がこれを自覚しないのはフィリングインの機能によるものであると言われている。この機能は、まさに我々の存在する外界が持つ普遍的な特徴、空間的な拡がりを持つ情報は、「空間的に滑らかに変化している部分が多い」ということを利用したものであると言うことができる。

一方、第2章で述べたように、我々の住んでいる外界から得られる様々な情報は、空間的に滑らかに変化している場合が多いだけでなく、力学系であり、時間と共に連続的に、そして滑らかに変化している場合が多い（空間情報の時間的滑らか仮説）。このことは、一見、極当然のこととして軽視されがちであるが、これは我々の空間認識において非常に大きな情報であることは確かであり、学習において大いに利用できるものと考える。

また、これも第2章でも述べたが、我々生物の持つセンサの多くは空間的に局所的な受容野を持ったセンサ細胞をたくさん並べることによって空間情報を獲得している。例えば、目という視覚センサは、たくさんの網膜細胞から構成されているが、個々の細胞は、空間的に局所的な受容野しか持っていない。にもかかわらず、我々の意識の中では、個々の細胞の発火を意識することなく、物体の位置などの空間情報を連続的なものとして知覚することができる。また、1.4節で述べたように、空間的な情報を用いて学習を行うことを考えた場合、微分の情報が重要な役割を果たすことから、空間の情報を連続値として表現することは非常に重要であると筆者は考える。これらのことから、我々の脳の中では、多数のセンサからの信号を統合し、連続的な空間情報を出力するニューロンが存在する可能性が考えられる。

4.2 空間情報の時間的滑らか仮説と空間情報の抽出

前節で述べたように、空間情報は滑らかにしか変化しないように見える。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。

ここで、図４.１のように、動く物体と視覚センサおよび階層型のニューラルネットから構成されるシステムを考える。視覚センサは、複数の網膜細胞（センサセル）よりなり、各細胞は、受容野の内、投射された物体が占める面積の割合を出力するものと仮定する。そして、その出力を階層型ニューラルネットへの入力とし、ニューラルネットの中の層間のニューロン同士は区別なく全結合させる。従って、ニューラルネットの側から各センサの信号の空間的な位置関係の情報を得ることはできない。

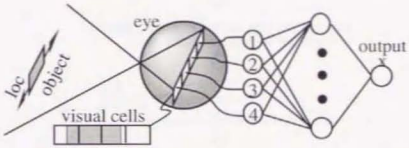


図４.１　時間軸スムージング学習を用いて空間情報を抽出するシステムの構成

この時、物体が視野内を滑らかに動いている、ここでは左右に単振動をしていると仮定すると、投射された物体の位置も徐々に動き、各網膜細胞の出力の時間変化は図４.２(a)のようになる。各網膜細胞は局所的な受容野しか持っていないため、それぞれ矩形波状の発火パターンとなる。この時、ある信号の後にある信号が発火すれば、その２つの信号の源である網膜細胞は隣同士である可能性が高い。従って、我々は図４.２(a)の発火パターンだけ見れば、網膜細胞１、２、３、４がその順番で並んでいることがわかる。

ニューラルネットが内部状態を保持しないと仮定すると、その出力は網膜細胞の出力だけの関数となっており、かつここで変化している空間情報は物体の位置だけである。従って、ニューラルネットの出力が時間と共に滑らかに変化していれば、このニューラルネットの出力は物体の位置を抽出したことになる。そこで、ニューラルネットの出力が時間に対して滑らかに変化するように、つまり時間の２階微分値を減少させるように学習を行う。すると、出力は時間の経過と共に図４.２(b)のように変化するようになる。これを物体の位置と出力の関係に置き換えると、図４.２(c)のように両者の間に１対１の対応がとれて、物体の位置の情報が抽出できることが期待される。

また、物体が前後に動いている場合は、物体が遠くにある時は網膜上では小さく、近い時には大きく映るため、図４.３のように各網膜細胞の出力が変化することが考えられるが、この場合も左右に単身どうさせた場合と同様な学習を行うことにより、図４.２(b)および図４.２(c)のような出力を得ることができると考えられる。

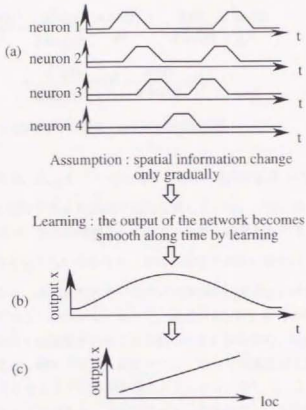


図４．２　時間軸スムージング学習による局所センサ信号の統合の原理

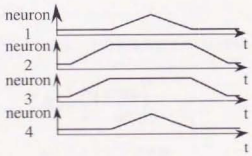


図４．３　物体が前後に動作している場合の網膜ニューロンの出力の変化

ここで、「出力を時間的に滑らかにする」ことが空間情報を抽出することにつながる様子をもう少し厳密に考えてみる。ニューラルネットの出力を x とすると、x の時間軸方向の滑らかさは $\frac{d^2x}{dt^2}$ であらわすことができる。この時、空間情報の内、物体の位置 loc だけが変化しているとすると、

$$\frac{dx(t)}{dt} = \frac{dx(t)}{d\,loc(t)} \frac{d\,loc(t)}{dt} \qquad (4.1)$$

$$\frac{d^2x(t)}{dt^2} = \frac{d^2x(t)}{d\,loc(t)^2}\left(\frac{d\,loc(t)}{dt}\right)^2 + \frac{dx(t)}{d\,loc(t)}\frac{d^2loc(t)}{dt^2}$$

$$= \frac{d^2x(t)}{d\,loc(t)^2}v(t)^2 + \frac{dx(t)}{d\,loc(t)}a(t) \qquad (4.2)$$

$v(t)$:物体の速度、$a(t)$:物体の加速度

と変形することができる。式（４.２）の右辺第１項は、速度の自乗 $v(t)^2$ を係数として$\frac{d^2x}{dloc^2}$、つまり、物体の位置 loc に対する x の曲線の凹凸を表している。一方、第２項は加速度の大きさ a を係数として$\frac{dx}{dloc}$、つまり、物体の位置 loc に対する出力 x の傾きの大きさを表している。従って、$\frac{d^2x}{dt^2}$を０に近づける学習を行うということは、物体の速度が大きい場合は、物体の位置 loc に対する出力 x の凹凸を減少させ、加速度が大きい場合には物体の位置 loc に対する出力 x の傾きの絶対値を０に近づけようとすることがわかる。従って、前述のように、物体が視野内を単振動しているとすると、振動中心あたりでは速度が大きく加速度が小さくなるため、loc に対する x の凹凸を減少させるという学習が進み、端の方では、速度が小さくなって加速度が大きくなるため、loc に対する x の曲線はフラットになるように学習が進むことになる。そして、この曲線に凹凸がなくなれば、１対１の対応がとれるようになり、ニューラルネットの出力として物体の位置が抽出できることになる。

４.３ 学習アルゴリズム

前節の原理より、ニューラルネットは、出力の時間変化の凹凸を減少させるために、

$$E(t) = \frac{\kappa}{2}\left(\frac{d^2x(t)}{dt^2}\right)^2 \qquad (4.3)$$

κ：スムージング定数

を誤差信号として学習を行う。つまり、教師信号 $s(t)$を

$$s(t) = x(t) + \kappa\frac{d^2x(t)}{dt^2} \qquad (4.4)$$

としてバックプロパゲーション（ＢＰ）法[Rumelhart 86]によって学習を行う。ただし、教師信号は内部生成するため、全システムから見れば外部からの教師信号は必要ない。この学習を時間軸スムージング学習と呼ぶ。また、本論文中では、スムージング定数 κ は0.5とした。そして、式（４.４）を差分近似すると、

$$s(t\text{-}1) = \frac{x(t) + x(t\text{-}2)}{2} \qquad (4.5)$$

という式に変換され、これを実際の学習に用いた。つまり、過去2回分のニューラルネットへの入力値を保持しておき、2単位時間前の入力をニューラルネットへ再入力してでてきた出力と現在の出力の平均値を求め、それを教師信号として1単位時間前の入力に対して学習を行った。ただし、ここでの学習も、同じデータで収束するまで何回も繰り返し学習を行わず、1単位時間あたり与えた教師信号で1回だけ学習した。

しかし、この学習を行うだけでは、出力$x(t)$は

$$x(t) = const. \qquad (4.6)$$

となり、情報を抽出できないことになってしまう。そこで、出力の値域を確保するために、過去の出力値の平均 $\overline{x(t)}$ を

$$\tau \frac{\overline{x(t)}}{dt} = -\overline{x(t)} + x(t) \qquad (4.7)$$

τ：時定数

の時定数τを大きくすることにより求め、現在の出力値$x(t)$との偏差が非常に大きい場合にのみ、さらにその偏差を大きくするという値域拡大学習を時間軸スムージング学習の代わりに行う。

以下、初期状態からスタートして初めて値域拡大学習が適用された時の実際のデータを図4.4に示しながら、その学習の様子を示す。まず、図4.4(a)に出力値$x(t)$およびその平均Eの変化の様子を示す。時定数τが大きいため（本論文中では100に設定）、$x(t)$の値は変動するものの、平均値 $\overline{x(t)}$ の値はあまり変化しない。

偏差が大きいかどうかの判定は、ここでは、試行錯誤の結果、4次の偏差を用いた。4次の偏差$dx(t)$は、

$$dx(t) = (x(t) - \overline{x(t)})^4 \qquad (4.8)$$

として求め、さらに、その時間平均 $\overline{dx(t)}$ を次式のように求め、

$$\tau \frac{d\overline{dx(t)}}{dt} = -\overline{dx(t)} + dx(t) \qquad (4.9)$$

これによって正規化した $d\tilde{x}(t)$

$$d\tilde{x}(t) = \frac{dx(t)}{\overline{dx(t)}} \qquad (4.10)$$

を計算する。$dx(t)$ および $\overline{dx(t)}$ を図４.４(b)に示す。このグラフは縦軸をlogスケールで描いてある。４次の偏差をとっているため、$dx(t)$ は大きく変化しているが、時定数τが大きいため（本論文中では式（４.７）と同様100に設定）、$\overline{dx(t)}$ はあまり変化しない。$\tilde{dx}(t)$ を図４.４(c)に示す。４次の偏差を用いたことによって、偏差が大きいところだけがさらに強調されることになり、２０単位時間のところだけが大きな値となっていることがわかる。

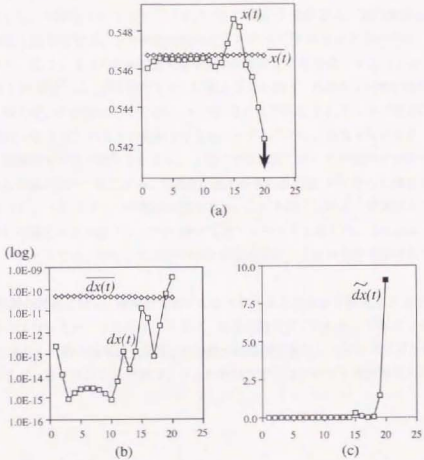


図４.４　各信号の変化の様子

そして、一様乱数 *rnd* を用いた

$$d\overline{x}(t) > \xi * rnd * (1-\overline{x(t)})^2 \qquad if\ x(t) \le \overline{x(t)} \qquad (4.11)$$

$$d\overline{x}(t) > \xi * rnd * \overline{x(t)}^2 \qquad if\ x(t) > \overline{x(t)} \qquad (4.12)$$

ξ:選択確率の調節用定数

という不等式によって値域拡大学習を適用するか否かを確率的に決定し、不等式が成り立った場合に

$$s(t) = 0.1 \qquad if\ x(t) \le \overline{x(t)} \qquad (4.13)$$

$$= 0.9 \qquad if\ x(t) > \overline{x(t)} \qquad (4.14)$$

という教師信号で学習を行った。ただし、ここでは、出力関数をシグモイド関数とし、値域は０から１の連続値とし、不等式（４.１１）、（４.１２）で用いた定数ξは、試行錯誤の結果1000とした。図４.４で示した場合では、２０単位時間のところで $d\overline{x(t)}$ の値が大きいため、不等式（４.１１）が成り立ち、式（４.１３）の値域拡大学習が適用された。不等式（４.１１）の $(1\text{-}\overline{x(t)})^2$ や不等式（４.１２）の $\overline{x(t)}^2$ は、出力値が０や１に偏ることを防ぎ、０から１の間に均等に値をとるように加えた。例えば、出力値の時間平均が１に近い場合は不等式（４.１１）で選ばれる確率が高くなり、この時式（４.１３）のように教師信号を0.1とすることで、出力全体を小さくすることができる。また、値域拡大学習が適用されると、（特に学習初期においては値域が十分に確保されていないため）出力の値が大きく変化する。すると、過去の平均出力値 $\overline{x(t)}$ からの偏差が非常に大きくなり、（４.１１）、（４.１２）の不等式が成り立つ。これを防ぐために、値域拡大学習が適用された際に、$d\overline{x(t)}$ の値を大きな値（ここでは10^4とした）にセットし直した。これによって、以下に述べるシミュレーションでは、平均して、ほぼ1000単位時間強に１回の割合で値域拡大学習が適用された。

図４.５に時間の経過に対し、教師信号がどのように与えられるかを模式的に示す。また、図４.６にこの学習のフローチャートを示す。学習は、毎単位時間毎に行われ、不等式（４.１１）または（４.１２）を満たした時は値域拡大学習、その他の時は時間軸スムージング学習を適用する。ただし、前述のように、ＢＰ法による学習は、与えた教師信号に基づいて１単位時間あたり１回だけ学習する。

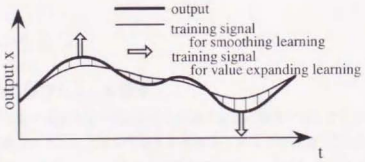


図４.５　教師信号の模式図

図４．５　教師信号の模式図

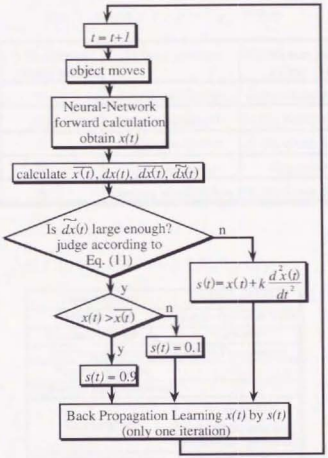


図４．６　学習のフロー

４．４シミュレーション

4.4.1 物体が左右に動作している場合

ここでは、前述の多数の視覚センサの出力から物体の位置の情報が抽出できるかどうかをシミュレーションによって検証した。シミュレーションは図４．１のような環境で行った。そして、(a) 網膜細胞の数、(b) 物体の動き方、(c) 物体の動作範囲の３つを変化させた。行ったシミュレーションの組み合わせを表3.1に示す。

始めに、表４．１の(1)(2)(3)のように、物体の動き方を単振動、動作範囲をちょうど視野と一致させた場合について、網膜細胞の数を１０、２０、３０個と変化させて学習を行った。その他のパ

ラメータを、表４.２に示す。ニューラルネットの階層は３層とし、中間層のニューロンは３０個とした。そして、重み値の初期値は-0.1から0.1の間から乱数によって決定した。また、物体は、93.1単位時間で１周期の運動をするものとした。

表４.１　行ったシミュレーションの条件

	(1) Number of retinal neuron	(2) Way to move	(3) Motion range of the object
(a)	10	simple oscillation	in the visual field
(b)	20	simple oscillation	in the visual field
(c)	30	simple oscillation	in the visual field
(d)	20	constant speed	disappear
(e)	20	random acceleration	in the visual field

表４.２　シミュレーションで用いたパラメータ

Object motion	simple oscillation
Motion range	just in the visual field
Structure of network	Layered
Learning algorithm	Back Propagation
Number of layers	3
Number of hidden neurons	30
Learning rate	0.1 (usually) 1.0 (only when applying expanding operation)
The initial weight value	from -0.1 to 0.1
Oscillating period	93.1 time steps
Time constant τ	100 time steps
Smoothing constant k	0.5
Width of 1 visual cell	1
Number of retinal neurons	30
Size of the object	2.5

学習の結果を図４.７以降に示す。図４.７はシミュレーション(３)、つまり、網膜細胞の数が３０個の場合の学習の様子を示したものである。それぞれのグラフは、物体の位置 loc に対するニューラルネットの出力をプロットしたものである。学習前は、出力の曲線はほぼフラットであるが、微小な乱数によって決められた重み値によって多少の凹凸がある。図４.７の(a)は、初期の微小な凹凸を示すために、他のグラフと比較して縦軸を１００倍拡大して描いてある。学習が進むと、値域拡大学習によって偏差が特に大きいところだけさらに偏差を拡大する学習が行われるため、図４.７(b)のように、曲線上に山と谷が一つづつできる。この時図４.７の(a)と比較すると、初期状態において

山と谷であったところが拡大されていることがわかる。ただし、初期状態で最大値をとっているところは、その近くに最小値をとるところがあるため、それに引きずられ、結果的に２番目に大きい値をとっていたところが山となっていることがわかる。さらに学習が進むと、図４．７の(c)のように、さらに値域が拡大し、曲線全体が滑らかになってくる。そして、さらに進むと、山と谷の部分が視野の端の方へ移動して、物体の位置と出力が１対１の対応がとれるようになっている。これは、前述のように、視野の端の方では物体の速度が０に近づくため、曲線をより平らにしようとする学習が働くためと考えられる。網膜細胞の数を変えたシミュレーション（１）および（２）もほとんど同じような経緯をたどり、最終的に、物体の位置と出力は１対１の対応がとれるようになった。ただし、最終的に得られる曲線が右上がりになるか左上がりになるかは、初期値によって変化する。

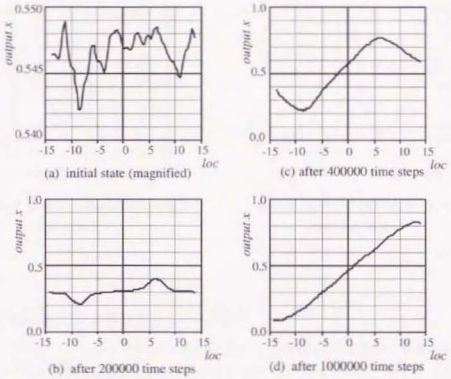


図４．７　学習の進行による出力の変化（網膜細胞３０個）

次に、物体の動作が単振動ではなく、一定速度で動作をし、視野からはみ出す場合についてシミュレーションを行った（表４．１の(4)）。この時の結果を　　図４．８に示す。この場合、物体が視野から消えた場合は、網膜細胞のすべての出力、つまり、ニューラルネットへのすべての入力が０となる。ところが、ここでは、ニューラルネットは記憶を持たず、出力は入力に対して一意に決定されるため、物体が視野の右端から消えようが、左端から消えようが、ニューラルネットへの入力が同じになるので出力も同じ値になる。また、出力を時間に対して滑らかにするという学習を行うことから、物体が視野の右端および左端に見える時は、物体が見えない時と近い出力値となる。従って、物体の位置に対する出力の曲線は、　図４．８のように、ちょうど１周期のサインカーブに近

い形となり、１対１の対応関係を得ることができない。

次に、物体が単振動ではなく、乱数によって決定された加速度によって動作をする場合（表４．１の(5)）のシミュレーションを行った。ここでは、物体が視野の端に到達した時には、逆向きの小さな速度で戻るという設定とした。この場合、学習後の物体の位置に対する出力は、図４．９のように、前のシミュレーションのようにきれいな曲線にはならず、１対１の対応もとれなかった。これは、加速度が乱数で決定されているため、式（４．２）の第２項の影響により、加速度の絶対値が大きい場合にスムージング学習によって傾きを０に近づけようという力が学習によって働いてしまうためと考えられる。このことから、本学習は加速度が小さい運動、つまり滑らかな運動について有効であると言える。

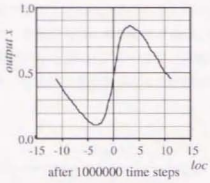


図４．８　物体が視野から消える場合の出力の様子

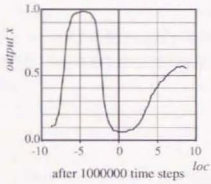


図４．９　物体が乱数で決定された加速度で動作する場合

4.4.2 出力の分布に関するシミュレーション

ここでは、入力データによって出力の分布がどのように変化するかを物体が前後に単振動する場合と、物体の大きさ自体が単振動で変化する場合を比較した。まず始めに、図４．３のように、物体が前後に動く場合にも同様の学習を行うことによってその情報が抽出できるかどうかをシミュレーションによって調べた。物体は前後に単振動するものとし、視野からははみ出ないものとする。図４．１０に、学習後の物体の位置 *loc* および網膜上に映った物体の大きさ *size* に対するニューラルネットの出力値をプロットしたものを示す。この場合も、物体の位置と出力は１対１の対応関係がとれていることがわかる。

次に、物体の位置に対する出力と網膜上の物体の大きさに対する出力を比較してみる。ここでは、物体の位置と網膜上の物体の大きさは反比例をしており、物体との距離が１の時に、網膜上の物体の大きさは４、２の時に２、４の時に１となる。従って、物体の位置が遠くになるほど物体の位置の変化に対する網膜上の物体の大きさの変化、つまり、$\frac{d\ size}{d\ loc}$ の値が小さくなる。にもかかわらず、物体の位置が振動中心の2.5付近で出力値が0.5に近い値であるのに対し、網膜上の物体の大きさが

2.5 付近では出力値は 0.5 よりかなり小さくなっていることがわかる。

比較のために、網膜上の物体の大きさ自身を単振動させて学習させた場合の物体の大きさに対する出力値を　　図４.１１に示す。この場合は、物体の大きさの振動中心である 2.5 付近で出力がほぼ 0.5 になっていることがわかる。これらから、時間軸に対して出力を滑らかに変化するように学習することによって、出力値の確率密度を均等化する働きがあることがわかる。ただし、図４.１０をよく見ると、物体の位置が４付近で出力値がフラットな領域が広くなっている。これは、２章で述べたように、振動の折り返し付近では出力値がフラットになるように学習が進む。すると、この時、物体の境界にある網膜細胞の出力は、ニューラルネットの出力に対し影響を及ぼさないように学習される。ところが、前述のように、物体が遠くにある時は、物体の位置が変化しても、網膜上に映った物体の大きさはほとんど変化をしない。よって、このシステムでは、物体が遠くにある場合、物体の位置が変化してもそれを検出できない状態となる。これは、網膜細胞の数を増大させれば避けることは可能であると考えられる。

さらに、物体が左右に単振動する場合の結果と比較すると、いずれの場合も学習が速く進んでいることがわかる。これは、この場合、各網膜細胞の出力の和という簡単な形で物体の位置が抽出できるため、学習が容易であったことが原因と考えられる。

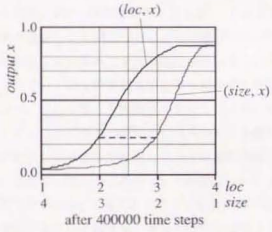


図４.１０　物体が前後に動く場合

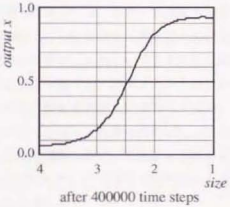


図４.１１　物体の大きさが変化する場合

４.５　考察および今後の課題

物体が視野の中を１次元の単振動運動をしている場合、提案した教師なし学習によってその位置を抽出することができた。しかし、実際に我々の住んでいる世界では、多くの物体は動かず、また動いても一方向に進んでいくものが多い。そのような意味で、物体が一定速度で視野を通り過ぎるという設定のシミュレーション（４）が最も現実に近いと考えられる。しかし、この場合、物体の位置と出力の間に１対１の対応付けをさせることはできなかった。我々人間の場合を振り返って考

えてみると、視野の右端から物体が消えれば、過去の履歴を元にして、視野よりも右の方に物体があるということを脳内に内部表現として獲得し、これに基づいて学習できると考えられる。従って、リカレントニューラルネットを用いて学習を行うことにより、そのような内部表現を形成することできるのではないかと考えている。ただ、右から消えたものはいつまでも視野の右側にあるとは限らない。従って、時間の経過と共に情報の確信度が減少し、新たな情報を獲得するようにしなければならない。

また、本シミュレーションにおいては、視野内の単振動という設定以外にも、動いている物体が１つであるとか、物体が１次元の運動しかしない等、まだまだ人工的な設定が多い。複数次元の運動を分離することに関しては、前章の相関情報抽出学習と一緒に用いた複数出力の直交化学習と同様な方法によって可能ではないかと考える。また、複数の物体が存在する場合については、注意の機構を導入し、注目すべき物体を切り出す等の対処が必要であると考えられる。これらも含めて、より現実的な環境に適応できるような学習アルゴリズムを考案することが今後の大きな課題である。

さらに、本学習では、うまく学習させるためのパラメータの設定が難しい。特に、値域拡大学習を適用するかどうかを決定する（４.１１）、（４.１２）式の ξ は、小さ過ぎると物体の位置と出力の間に１対１の関係が得られず山や谷がいくつかできてしまい、大き過ぎると値域拡大学習が適用される回数が減り、学習が遅くなると考えられる。また、 ξ の大きさは物体の運動とも大きくかかわる問題であり、物体の運動の振動数が大きければ ξ は小さい方が良く、大きいと学習の速度が遅くなる。また、振動数が小さければ、 ξ は大きくなければ情報を抽出できない。逆に考えると、複数の運動を ξ を調節することで分離するといった可能性も考えられる。また、値域拡大学習の適用を決定する不等式で４次の偏差を用いることも、出力曲線上の余計な凹凸に値域拡大学習を適用しないで、１対１の対応関係を得られやすくするための苦心の策で、４次ということに特に必然性はない。これらをより洗練させていくことは、本学習の妥当性を言う上で必須であると考える。

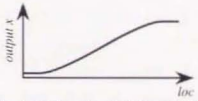
しかし、空間情報は時間的に滑らかにしか変化しないという仮説を利用した学習は不変的、汎用的であり、我々生体が利用している可能性は非常に高いと考えられる。本章の初めにも述べたように、空間的に滑らかであるという拘束については、既に、画像理解の際に、不良設定問題の正則化の際に用いられ、有効であることが示されている[Poggio 85][横矢 91]。空間情報が時間的にも滑らかに変化するという拘束も、さらに研究を進める必要があると考える。

しかし、そもそも空間情報の時間的滑らか仮説が理にかなっているのかという議論も考えられる。例えば、動いている物体が壁の後ろに隠れてしまった時、見た目には不連続に変化しているように見える。しかし、この時、物体が連続的に動いていると考えることによって、壁の反対側の端から物体が出てくることが予測できるのである。また、物体が壁に当たって跳ね返るように、空間情報の変化が滑らかでないように見える場合には、これを連続的な状態の変化として捉えることにより、因果関係の把握につながると考えることができる。そして、逆にもし物体が壁をすり抜けたとすると、通常の連続的な状態変化が起こらないことで違和感を感じるのであると考えることができる。

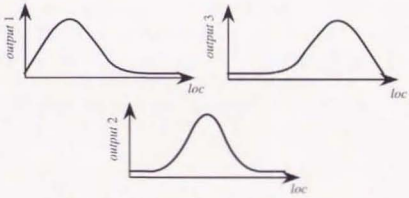
また、本論文では、多数のセンサ信号を統合し、空間情報をアナログ情報として出力することを学習させている。1.4.1 で述べたように、アナログ化された情報は微分情報をとることが容易であり、学習という側面で非常に有効であると考えられる。また、強化学習を行う際に、遅れて得られる強化信号から各状態の評価を学習する際に、本論文で提案したスムージング学習を応用できることも

確認されている。従って、本論文で行ったセンサ信号の統合と強化学習を融合させ、より柔軟で知的なシステムの構成が可能であると考える。ただ、時間軸スムージング学習を用いて強化学習を行うだけでもある程度センサ信号の統合ができることがわかってきている。そこから考えると、センサ側で時間軸スムージング学習を適用するということは、強化学習を加速させるという効果が考えられる。また、時間軸スムージング学習は適用するものの、値域の確保については強化学習における学習信号をここまで伝搬させることによって行う方法も考えられる。

ただし、現在の所、生体内で空間情報をアナログコーディングしたようなニューロンが見つかったという報告はない。ただ、図４．１２の(a)のように、１ニューロンによるアナログコーディングの代わりに、(b)のように複数のニューロンでファジーのメンバーシップ関数のような形でコーディングしている可能性は高い。このような場合でも、本論文で提案した空間情報の時間的滑らかさを利用した学習を行うことが可能であると考える。いずれにせよ、データの表現方法等、生理学等の分野とのすりあわせをしていく必要があると考える。



(a) Representation of spatial information by one neuron



(b) Representation of spatial information by multiple neurons

図４．１２　物体の位置の表現方法

４．６　まとめ

空間情報が時間的滑らかであるという仮説が多数の局所的な受容野を持つセンサ信号を統合する際に有効であることを示した。そして、階層型ニューラルネットの出力を時間的に滑らかにするという時間軸スムージング学習を行うことによってセンサ信号を統合し、空間情報をアナログ値とし

て抽出できることを示した。さらに、シミュレーションによって、物体が視野内を左右および前後に一次元の単振動運動をしている場合について、ニューラルネットの出力が物体の位置に対して滑らかに、かつ1対1の関係が成り立つように学習することができた。しかし、物体が視野から消えてしまう場合や物体の加速度を乱数で決定させた場合は、物体の位置と出力の間に1対1の関係を実現することはできなかった。

それから、本学習も前章の学習と同様に値域拡大学習によって出力値の値域を拡大させるという学習はさせるものの、中間的な値に対する明示的な規定はない。そこで、学習後のデータの分布を調べたところ、入力データの時間変化に依存し、入力データが時間によってあまり変化しない部分では出力の解像度が上がり、入力データが大きく変化するところでは出力の解像度が下がることがわかった。

第５章　局所センサ信号統合化学習による視覚系機能の学習モデル

本章では、前章で提案した時間軸スムージング学習による局所センサ信号統合化学習を基本として、頭部位置によらない物体位置認識の学習モデル、動眼前庭反射の学習モデル、物体追跡の学習モデルが説明できることを示す。これらは主に、時間軸スムージング学習による局所センサ信号統合化学習の有効性を肉付けする立場で進めた研究であり、生体のモデルとしては必ずしも適切でない部分が多々あるが、最終的には両者が収束することを期待している。

5.1　頭部位置によらない物体位置認識の学習モデル

我々生物は、頭や目が動いても正しく物体の位置を認識することができる。例えば頭を大きく動かしてもまわりの物体が動いているとは感じない。ところが、前章で提案したシステムでは、頭や目の位置に関する信号の入力がないため、このような機能を実現することが不可能であった。ところが、頭や目の動きが滑らかでないと仮定すれば、前章のシステムに頭や目の位置を入力として加えてやれば、出力を時間に対して滑らかに変化するようにという前章で提案した学習アルゴリズムによって頭や目の位置によらない物体の位置の認識ができるようになると考えた。

具体的な例を挙げて説明する。図5.1のように、頭部位置が可変の視覚センサを考え、頭部の位置および視覚センサからの出力を階層型ニューラルネットへの入力とした。

物体と視覚センサ（網膜）との位置関係は、図5.2のように、物体の位置および頭部の位置の差によって表され、視覚センサの出力もその値に従って変化する。頭部位置は、ここでは乱数で決定

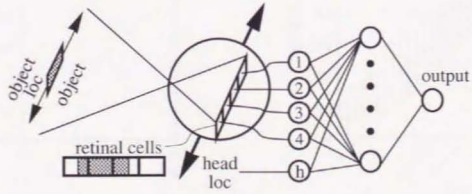


図5.1　頭部位置によらない物体の位置認識を学習するシステムの構成

するものとし、物体は前章と同じく、視覚センサの視野からはみ出ない範囲で左右に単振動をしているものとする。シミュレーションの環境は、前章のものとほとんど同じで、視覚センサのセンサセル数は３０個とし、各センサセルの幅は１とした。投射された物体の大きさは 2.5 とし、頭部の位置は、中心から左右に 4.0 の間で毎単位時間乱数で決定した。また、物体は視野から消えない範囲（振幅 9.75）で左右に単振動しているものとする。

学習後、物体の位置と頭部の位置を変化させた時の、物体の位置に対するニューラルネットの出力値をプロットしたものを図５.３に示す。この図から、頭部の位置が変化しているにもかかわらず、物体の位置と出力が１対１の関係になっていることがわかる。また、頭は動いているにもかかわらず、頭部位置の入力を０に固定した場合の出力を図５.４に示す。この場合は、頭部位置を補正することができず、物体の位置と出力の間に１対１の関係を得ることができなかった。以上より、頭部位置をニューラルネットに入力することにより、頭部位置によらない物体位置認識を学習することが確認できた。

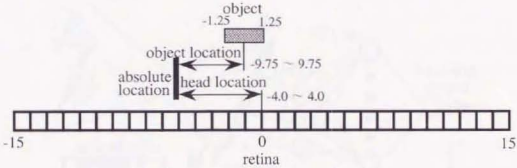


図５.２　物体と網膜との位置関係

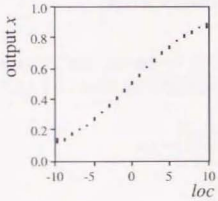


図５.３　学習後の物体の位置に対する出力

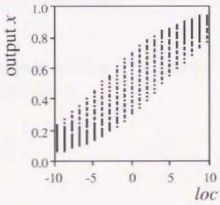


図５.４　頭部位置の入力値を０にした場合の物体の位置に対する出力

５.２　前庭動眼反射の学習モデル

我々生物が行っている頭部位置の補償には、前節のように頭の中で補償するという機能の他に、前庭動眼反射（VOR Vestibulo-Ocular Reflex）という機能があることがよく知られている。これは、頭部の位置が変化した場合、それと逆の方向に目が移動することによって、網膜上に映る物体がぶれないようにする機能である。また、この前庭動眼反射では、環境を操作してやることにより、頭部の動きと目の動きの間のゲインが適応的に変化することも知られている[Gonshor 76]。

そこで、図５.５のように、前章のシステムに、さらに目の動きを学習するニューラルネット（eye movement net）を設け、そこに頭の位置を入力し、その出力に従って目を動かすというシステムを構成した。この時の物体と網膜の位置関係は、図５.６のように、頭部の位置、目の位置、そして物体の位置を用いて表すことができる。物体の位置を抽出するニューラルネット（ＬＳＳＩネットと呼ぶ）は前章と同様な学習を行い、目の動きを決定するニューラルネットに関しては、ＬＳＳＩネットの時間軸の滑らかさを強化信号として、より滑らかになるように学習を行った。

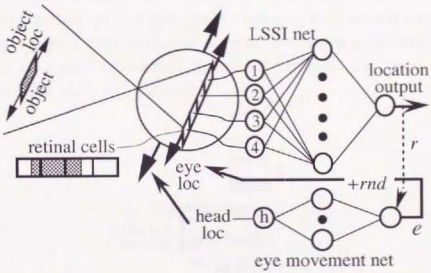


図５.５　前庭動眼反射の機能を学習するシステムの構成

具体的には、まず、目の動きのニューラルネットの出力に乱数 rnd を加えて目を動かす。そして、ＬＳＳＩネットの出力の凹凸の度合い v およびその時間平均 $\hat{v}$ を

$$v(t) = \left\{ \frac{x(t-1) + x(t+1)}{2} - x(t) \right\}^2 \qquad (5.1)$$

$$\tau \frac{d\hat{v}}{dt} = -\hat{v} + v \qquad (5.2)$$

x：ＬＳＳＩネットの出力

と計算する。この時、時間平均と比較した現在の出力の時間変化の滑らかさ、つまり、出力の時間変化の凹凸の小ささを強化信号とするため、強化信号 r を

$$r = \hat{v} - v \qquad (5.3)$$

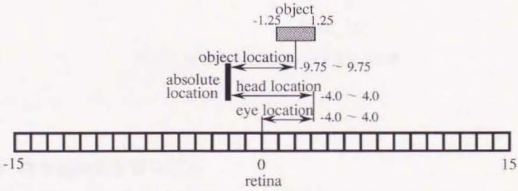
と計算する。そして、目の動きを決定するニューラルネットに対し、２.２節で述べたように

$$s_2 = x_2 + \eta \cdot rnd \cdot r \qquad (5.4)$$

x_2：目の動きのニューラルネットの出力
η：学習のための定数

という教師信号を内部生成して学習を行う。具体的に用いた数値は、ほぼ前節と同様である。ただし、目の動きに加えた乱数 rnd は±0.04 の範囲の一様乱数とし、η は試行錯誤から 100 とした。

学習した結果、頭の位置が動いているにもかかわらず、物体の位置に対するＬＳＳＩネットの出力は、図５.７のようにほぼ１対１の対応がとれるようになった。また、目の位置を中央（0.0）に固定すると、図５.８のように１対１の対応がとれなくなる。また、頭の動きと目の動きの関係を調べると、図５.９のようにちょうど符号が反対になっていることがわかる。このことから、目の動きによって頭の動きが補償されていることがわかる。



object location : simply oscillated
head location : randomly chosen
eye location : calculated by the eye motion net

図５.６　物体と網膜との位置関係

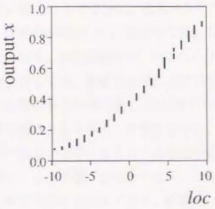


図５.７　学習後の物体の位置に対する出力

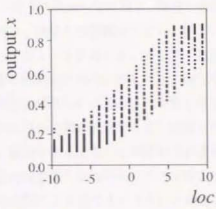


図５.８　目の位置を固定した場合の物体の位置に対する出力

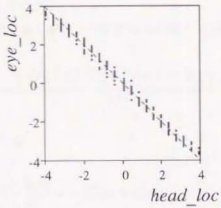


図５.９　学習後の頭の位置に対する目の位置

５.３　物体追跡の学習モデル

我々生物は、常にできるだけ多くの外界の情報を取り込み、それに対して適切な対応をとらなければならない。ところが、もしセンサからの入力が時間によって変化しなければ、これは新しい情報が得られていないということになる。従って、我々生物は、例えば、物体が見えなくなったらそれを追跡するというように、より多くの情報を得るように行動する必要がある。つまり、我々の行動が、センサを通じてより多くの外界の情報を得るように学習されているのではないかと考えることができる。ここでは、得られる情報の多さを、局所センサ信号統合化学習によって統合された空間情報の時間微分値の絶対値で表されるものと考え、強化学習を用いて、この時間微分値を最大化するように行動を決定する方法を考えた。そして、これによって、物体を追跡するという動作が学習できるかどうかを確認した。

まず、図５.１０のように、視覚センサからの信号と目の位置の情報を受け取り、物体の位置（出力１、物体の位置出力）、物体の位置出力の変化量（出力２、物体の位置微分出力）、新しい目の位置（出力３、目の位置出力）の３つの出力を出すニューラルネットを考える。そして、物体の位置出力に対しては、前章と同様の方法で学習を行い、物体の位置微分出力に対しては、(1)１番目の物体の位置の出力の変化量と (2)２番目の出力自身の時間変化が滑らかになるような値を足し合わせて教師信号として与え、学習をさせる。そして、３番目の目の位置を決定する出力に対しては、２番目の出力が大きくなるように強化学習を行うと共に、振動を防止するため時間変化が滑らかなほど良いという学習も合わせて行う。これによって、物体の位置出力が時間的に大きく変化する場合は２番目の出力が大きくなり、結果的に物体の位置が時間的により変化するように目が動くという仕組みになっている。直接物体の位置の出力の微分値を学習に用いると、その変化が急峻になってしまい、目の動きをうまく学習することができなくなる。そこで、２番目の出力ニューロンを設け、ニューラルネットの出力が、入力の変化に対して滑らかに変化する特徴と、上記のように時間に対して滑らかに変化するようにという項による学習によって、物体が視野から消えそうになるとその値がだんだん小さくなるようにする。

具体的な式および数値を示す。ある時点での位置出力の傾き $slope$ は、

$$slope(t) = \frac{\left|x(t) - x(t-1)\right| + \left|x(t+1) - x(t)\right|}{2.0} \qquad (5.5)$$

と計算し、その時間平均 $\overline{slope}$ を

$$\tau\frac{d\overline{slope}}{dt} = -\overline{slope} + slope \qquad (5.6)$$

と求める。ただし、時定数は1000とする。そして、位置微分出力 s_2 を、出力が滑らかになるための教師信号 $s_{2,smooth}$ と位置出力の微分値の教師信号 $s_{2,slope}$ を混ぜ合わせた

$$s_2 = \alpha s_{2,smooth} + (1-\alpha)s_{2,slope} \qquad (5.7)$$

とした。ここでは、α を0.9とし、それぞれの教師信号は

$$s_{2,smooth} = \frac{x_2(t-1) + x_2(t+1)}{2.0} \qquad (5.8)$$

$$s_{2,slope} = \frac{1.6}{1 + \exp(-slope / \overline{slope})} \qquad (5.9)$$

によって求めた。さらに目の位置出力に対しては、

$$s_3 = \frac{x_3(t-1) + x_3(t+1)}{2.0} \qquad (5.10)$$

を教師信号として、時間に対して滑らかになる学習を行うと共に、さらに、

$$s_3 = x_3 + \zeta \frac{\partial x_2}{\partial sensor} \qquad (5.11)$$

$sensor$：目の位置、ζ：学習の定数（ここでは、0.01）

によって、位置微分出力が大きくなるように強化学習を行った。この式における偏微分は、

$$sensor = x_2 + \Delta$$

$$sensor = x_2 - \Delta \qquad (5.12)$$

の２点（ここでは、Δ=0.01）における x_2 の値を計算し、その差に $1/\Delta$ をかけたものとした。また、このシミュレーションでは、図５．１１のように目の最大振幅を±15、物体の単振動運動の振幅を25.75とし、目を動かさなければ物体が視野からはみ出るように物体を動かして学習を行った。例えば、物体が最大振幅の位置にある場合は、目を反対側に12動かさなければ、物体の一部または全部が見えなくなる。

図５．１２に学習後の物体の位置に対するニューラルネットの３つの出力値の値をプロットしたものを示す。この時、物体は学習時と同様に単振動をしている。これより出力１が物体の位置を学習でき、さらに出力２がその変化量をほぼ学習していることがわかる。また、出力２の値が、物体の位置の絶対値が大きい時に小さくなっている。これは、物体の位置の絶対値がさらに大きくなって物体が視野から消えると出力１は一定値となり、出力２が０に近づくような学習が働き、さらに、前述のように出力２の時間変化が滑らかになるように学習されているため、物体が見えている場合でも端に行くほど出力が０に近づくことになる。そして、出力３は、出力が２が大きくなるように強化学習を施しているため、物体がはみ出さないように動作するような出力となっている。また、出力３は物体の位置に対して単調に増加せず、物体が中心から離れた場所では最大でフラットにな

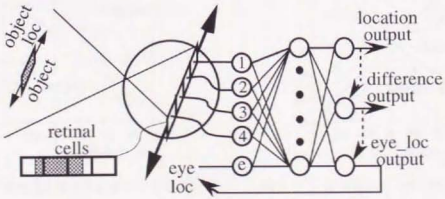


図５．１０　物体追跡を学習するシステムの構成

っている。これは、物体をより真ん中で捉えようとしていることになる。これは、出力２の出力が滑らかになるように学習されているため、真ん中付近の値が最大になるため、これをもとに学習した出力３はより真ん中にセンサを持っていこうとすると考えられる。また、物体の位置に対するセンサ及び物体とセンサの相対位置関係を図５．１３に示す。ここで、物体と目の相対位置の絶対値16.25を越えると物体が視野から完全に消えることになるが、目が物体の方へ動くことによって物体が視野から消えないようになっていることがわかる。また、本学習では、現在の目の位置がニューラルネットの入力となり、次の時刻の目の位置がニューラルネットの出力として得られるというフィードバックループを有する上、物体の位置出力の時間変化量の絶対値が大きいほどよいということになっているため、目の位置が大きく振動してしまうという不安定な解に陥りやすい。ここでは、目の位置出力に対しても時間変化が滑らかなほどよいという拘束を設けて学習をさせているが、より明確な学習アルゴリズムが必要であると考える。

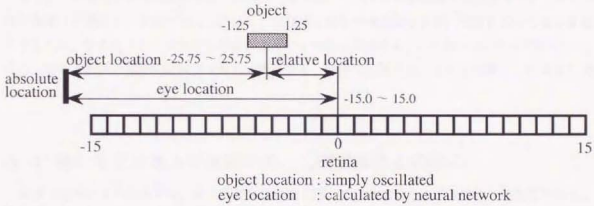


図５．１１　物体と網膜との位置関係

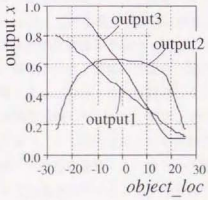


図５．１２　学習後の物体の位置に対する各出力

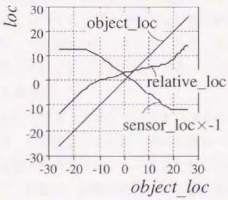


図５．１３　学習後のセンサの絶対値と物体との相対位置

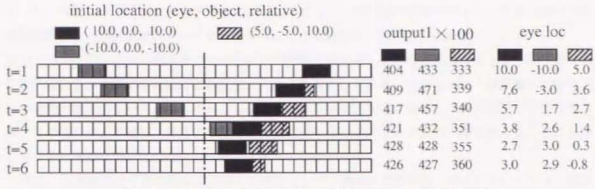


図５．１４　学習後に物体を提示した時の網膜上での物体の動き

図５．１４に物体をある位置に固定した時のセンサの動きの様子を表す。この図では、物体の位置と初期センサの位置を変えた３つの場合が示して

ある。この場合、物体の位置を０に固定した場合は、センサの初期位置を変化させてもセンサは同じ場所に収束することがわかる。ただし、収束点は真ん中付近になるが、学習によって多少変動をするため、学習をストップさせる時刻によってこの値は変動する。この場合は、右にずれている。また、物体と網膜の相対位置関係が等しい場合でも、物体の位置を違うものと認識し、収束点も違っていることがわかる。

５．４　他のモデルおよび生理学的、心理学知見との対応

本章で述べた３つのモデル、特に前庭動眼反射に関しては、他にもモデルや多くの知見がある。本節では、これらとの比較をおよび対応付けを行う。

頭部位置によらない認識に関しては、頭を大きく左右、上下に動かしても、まわりの空間が動いているように感じないという我々の感覚に基づくものである。また、これに関連した話しとして、Andersen らは、サルの後部側頭葉の皮質部において、網膜上の物体の位置だけでなく、物体の位置と眼球の位置の両方によって興奮の度合いが変化するニューロンが存在することが確認されている[Andersen 85]。この実験では頭部は固定して行われているが、いずれにせよ、頭部および眼球の位置情報と物体の位置情報を脳内で統合して空間の情報を得ている可能性は高いと考えられる。また、頭部、眼球の位置と物体の位置の正確な関係を生得的に持っていることは不自然であり、この機能を学習によって獲得、または調節している可能性は高いと考える。

前庭動眼反射に関しては、特に、その適応性が小脳における学習と関連しているということが言われており、よく研究されている分野である。この機能は、三半規管による頭の回転の検知→前庭神経核→外眼筋の運動ニューロン→外眼筋による眼球の運動という３つのニューロンを介した経路によって実現されていることが知られている。これに対し、伊藤は、前庭神経核への小脳片葉からの投射に注目し、小脳における学習が前庭動眼反射における適応の機能を担っているというモデルを提案し[Ito 70]、これを裏付ける実験も行った[Ito 82]。さらに、川人らは、フィードバック誤差から制御系の逆ダイナミクスモデルを階層型ニューラルネットで学習し、フィードフォワード制御を

行うというフィードバック誤差学習[Kawato 87]によってこの機能の説明が付くことを示し[川人 94]、前庭動眼反射の適応実験のデータが再現できることを確かめている[Gomi 92]。

川人らのモデルでは、網膜像におけるぶれの速度を検出し、これを誤差信号として小脳片葉における学習によって適応的な前庭動眼反射が実現されるとしている。しかし、この網膜像におけるぶれの速度の検出については触れられていない。これに対し、本章で提案した前庭動眼反射のモデルは、このぶれを、視覚センサ信号を統合した出力値の滑らかさというもので置き換えることによってよりその機能を明確化したものであると言うことができる。

次に、目が物体や背景の動きを追跡する運動として２種類の運動が知られている。１つは、追従性眼球運動（OFR : Ocular Following Response）、もう一つが平滑性追跡眼球運動（Smooth Pursuit Eye Movement）である。前者は、背景などの広い視野を占めるものがゆっくりと動作すると、目がそれにつれて反射的に同じ方向に動くというものである。後者は、興味ある物体を目で追いかける随意的な運動である。本章の物体追跡のモデルは、何か一つの物体に注意している場合を想定しており、平滑性追跡眼球運動に近い。しかし、随意的という点を本モデルではあまり考慮しておらず、また、生理学的、心理学的知見とのすりあわせもあまり行っていないため、今後検討していく必要がある。

5.5　まとめおよび考察

本章では、前章で提案した時間軸スムージング学習を用いた局所センサ信号統合化学習を拡張することにより、視覚システムの（１）頭部位置によらない認識（２）前庭動眼反射（３）物体追跡といった機能が学習によって獲得できることがある程度説明できることを示した。（１）および（２）では、物体の位置を抽出する出力が時間に対して滑らかに変化するようにという学習によって、頭部の位置を（１）ではニューラルネット内で補償し、（２）では目の動きをニューラルネットで学習させることで補償した。（３）では、外界の情報をより多く獲得するようにという学習によって実現した。ここでは、局所センサ信号統合化学習を行っているニューラルネットの出力の時間変化量の絶対値を得られる外界の情報量として用いた。

ただ、前章で述べた局所センサ信号統合化学習自体の問題点や、前庭動眼反射のモデルで頭部の位置を各単位時間毎に乱数で決定するなど不自然な点が多い。実際は、頭部の位置も何らかの運動指令に基づいて動いているはずであり、それをどう扱うかは大きな問題として残る。また、物体追跡の学習モデルでも出力の時間変化量の絶対値を大きくしようとしているため、目の動きが振動を起こしやすいという本質的な問題を抱えている。本論文では、目の動きを決定する出力も合わせた核出力が時間と共に滑らかに変化するようにという学習によって振動を抑えている。しかし、それでも目が小さな振動を起こすことがある等モデルとしてまだまだ不十分な点が多い。また、生理学的、心理学的な裏付けもあまり得られていない。これらの問題点を検討し、より妥当性のあるモデルを作っていくことは今後の課題である。

ただ、これらを通して、局所センサ信号統合化学習、時間軸スムージング学習および空間情報の時間的連続性が我々の生体の認識機構、特にその学習において大きな役割を果たしているのではないかという可能性をわずかながらも示すことができたのではないかと考える。

第6章　強化学習に基づく能動認識

6.1 背景

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、それに基づいて目的に沿った適切な行動を行うことができる。また、外界の情報を効率的に取り込むために、より良い認識ができる位置にセンサを動かしたり、その場の状況に即したセンサ情報を選択するといった、いわゆる能動認識[山崎 92]の機能も有する。図6.1に示したように、センサを移動させることによって、より良い認識ができ、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い、より効率的に目的を達成することができる。例えば、カエルが目の前を飛んでいる物体を発見したとすると、目を動かして目の前の物体がエサかどうかを認識し、エサだと認識した場合は、飛びついて食べることによってエサという報酬を得る。つまり、生物にとっては、認識自体も、そして、そのためにセンサを移動させるという動作も、食欲などの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可能である。このことは、認識や認識のための動作も、食物を食べるために食物に近づくといった目的に直接結び付くような動作と同様に捉えることができることを意味している。

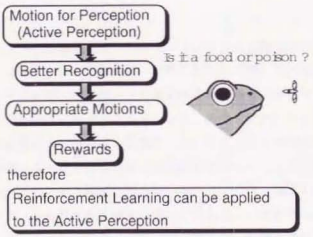


図6.1　なぜ能動認識に強化学習を適用するか

一方、第1章および前章で述べた強化学習は、試行錯誤を基により良い解を学習するという特徴を持ち、教師あり学習（ここでは、直接理想出力である教師信号を与える学習を教師あり学習と定義する）と比較してより自律的な学習である。さらに、通常、与える情報量が教師あり学習に比べ

て少ないことから、学習に時間がかかるが、その分環境に適した柔軟な学習ができる。従来、強化学習は主として、機械（ロボット）に与えられた目的を達成するための動作を学習させるために使われてきた。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、ニューラルネットの学習、汎化能力が強化学習に有効であることが確認されている。

前述のように、能動認識におけるセンサの移動も、また認識自体も、目的達成動作の一つと捉えることができる。そこで、我々は、強化学習で能動認識の機能を学習させることを考え、さらに、ニューラルネットを用いることで、ニューラルネットの学習、汎化能力を利用することを考えた。通常の強化学習では、強化信号は一連の動作を行った後に得られる、いわゆる遅延強化信号となるため、過去の動作をいかにさかのぼって学習するかが問題となる。しかしここでは、簡単のため、認識結果に対して外部から逐次強化信号が得られるという仮定をおいた。そして、その強化信号からセンサの移動と物体の認識を同時にニューラルネットに学習させるためのシステムの構成法と学習法を提案する。

従来、能動認識は、外部から全く情報を得られない場合について、阪口らによって、観測によるエントロピー減少量の期待値を最大化するという観点から定式化され、触知覚におけるセンサ情報の選択に適用されている[阪口 91][阪口 93]。

Whitl 行る評価を学習によって獲得し、その評価が最大になる注視点を選択するという方法をとっており、センサの移動を連続的な動作として捉えて学習させていない。また、ニューラルネットは用いていない。

6.2 学習アルゴリズム

6.2.1 全体構成

図6.2のような、センサと階層型ニューラルネットよりなる能動認識のための学習システムを提案する。本論文では、センサとして可動視覚センサを例として用いる。視覚センサは複数のセンサセルよりなり、各センサセルは、映し出された物体の面積が各センサセルの全体の面積に対して占める割合を出力するものとする。ニューラルネットは、各センサセルから信号を受取り、層に従って出力を計算する。出力層ニューロンは、認識用ニューロンとセンサ移動用ニューロンの2種類よりなる。そして、認識用ニューロンの出力値が認識結果を示し、センサ移動用ニューロンの出力値がセンサの移動量を決定するという構成になっている。ここでは、簡単のため、認識用ニューロンとセンサ移動用ニューロンをそれぞれ1個ずつ描いてあるが、いずれのニューロンも複数個あってもよい。また、ニューロンの出力関数はシグモイド関数とし、出力値は0から1の値域を持つ連続値とする。

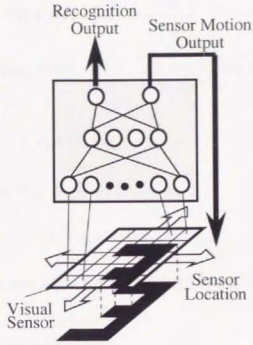


図6.2 ニューラルネットを用いた能動認識システムの構成

6.2.2 認識の学習

始めに、認識出力の学習に関して述べる。前述のように、認識自体も目的達成のための動作の一つと捉えることができるため、強化学習によって学習を行う。図6.3に認識の学習の際の信号の流れを示す。 まず、認識用の出力 o_p に対し、平均0の微小な乱数 rnd_p を

$$a_p = o_p + rnd_p \qquad (p=1,2,\cdots,\mathrm{P}) \qquad (6.1)$$

P:認識出力の数

のように付加して認識の信号 a_p とする。この時、出力が複数ある場合は、各出力に対し別々の独立した乱数を加える。そして、ここでは、理想的な認識パターン $\boldsymbol{ideal}$ が存在するものとして、外部の認識結果評価部 (Recognition Evaluator) において、

$$\phi = -\sum_{p=1}^{P} (ideal_p - a_p)^2 \qquad (6.2)$$

ϕ : 認識出力の評価値

というスカラー量で認識出力が評価される。そして、現時点 T での認識結果 a_p が前の時刻 T-1 と比べて良くなったか悪くなったかを表わすスカラー量 r が

$$r(\mathrm{T}) = \phi(\mathrm{T}) - \phi(\mathrm{T}-1) \approx \frac{d\phi(\mathrm{T})}{dt} \qquad (6.3)$$

と計算され、強化信号として出力される。そして、システムはこれを受け取り、自乗誤差 e_p および誤差信号 ε_p を

$$e_p = (r \cdot rnd_p)^2 = \varepsilon_p^2 \qquad (6.4)$$

として、つまり、教師信号 s_p を

$$s_p = o_p + r \cdot rnd_p \qquad (6.5)$$

としてＢＰ法[Rumelhart 86]に従いニューラルネットを１回だけ学習させる。

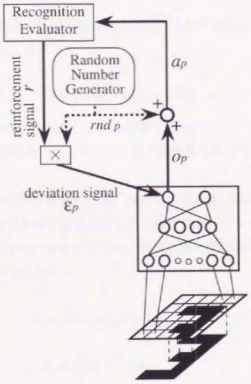


図６．３　認識の学習

この学習によって、ニューラルネットは現在見えるセンサの画像に対して正しい認識を行おうとする。その結果、センサが、与えられたパターンをうまく認識できる位置にある時には、学習によって次第に正しく認識できるようになるが、物理的にうまく認識できないような位置にセンサがある場合は、いくら学習しても正しく認識できるようにはならない。これに対して、次節で述べるセ

ンサ移動の学習により、システムはより正しく認識できる方向へセンサが移動していくことを学習する。

認識の学習に関しては、理想的な認識パターンを用いて、直接教師あり学習を適用する方が簡単であり、収束も速い。しかし、ここでは、強化学習による認識の学習の可能性を検討する意味で、あえて教師信号から強化信号を生成して学習を試みた。これは、サルに文字を見せて、対応する正しい札を挙げた場合にエサがもらえるという状況に近い。

6.2.3 センサ移動の学習

次に、センサ移動の学習について述べる。図６.４にセンサ移動の学習の際の信号の流れを示す。まず、センサ移動用の出力 o_m に対し、平均０の微小な乱数 rnd_m を

$$a_m = o_m + rnd_m \qquad (m=0,1,\cdots,\mathrm{M}) \qquad (6.6)$$

M:センサ移動用出力の数

のように付加した値 a_m を求める。認識の場合と同様に、出力が複数ある場合は、それぞれに別々の乱数を加える。そして、この a_m を用いて、

$$x_m(\mathrm{T}+1) = x_m(\mathrm{T}) + \alpha(a_m - 0.5) \qquad (6.7)$$

$x_m(\mathrm{T})$:時刻Ｔの視覚センサの位置

α : 単位時間あたりのセンサの最大移動量を決定する定数

のように視覚センサを移動させる。移動した後、視覚センサが再び画像を取り込み、ニューラルネットは新しいセンサからの情報を受け取る。そして、再びニューラルネットの計算を行う。こうして得られた新しい認識用出力から、前節のように、認識結果評価部から前の認識用出力よりも良くなったかどうかを示す強化信号 r をもらい、

$$e_m = (r \cdot rnd_m)^2 = \varepsilon_m^2 \qquad (6.8)$$

$$s_m = o_m + r \cdot rnd_m \qquad (6.9)$$

の値からＢＰ法に従ってニューラルネットを１回だけ学習する。この学習によって、ニューラルネットはより正しい認識ができる方向へのセンサの動作を獲得することができる。

上記の認識の学習およびセンサ移動の学習は並列に行う。つまり、センサ移動用出力に乱数を足したものにしたがってセンサを動かし、新しい画像を得る。そして、その得られた画像からニューラルネットの計算を行い、得られたセンサ移動用の出力は次のセンサの移動に用いる一方、認識用の出力はさらに認識出力に対する乱数を加えた後に評価される。従って、ここで生成される強化信

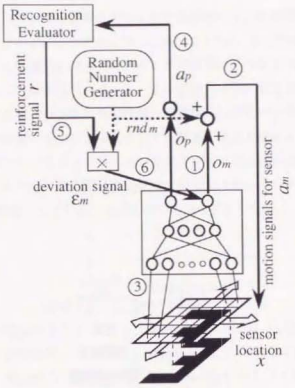


図６.４　センサ移動の学習

号の計算に使われる認識結果 a_p は、認識用出力に加えた乱数 rnd_p とセンサ移動用出力に加えた乱数 rnd_m の両者の影響を受けている。また、認識用出力、センサ移動用出力が複数個ある場合は、それだけの乱数の影響を受けることになる。そして、得られた強化信号 r は、認識用出力の学習と１単位時間前のセンサ移動用出力の両方の学習に用いられる。この時、加えた乱数間の干渉が考えられるが、加える乱数は平均が０で、それぞれ互いに独立であるため、他の乱数の影響の期待値は０になる。よって、多数の試行を繰り返すことにより、認識、センサ移動共に正しい方向に学習が進む。

また、認識およびセンサ移動の学習は、共に強化学習に基づいているため、式（６.３）と式（６.６）、式（６.４）式（６.５）と式（６.８）式（６.９）といったようにほとんど同一の式で学習方法を記述することができる。これは本学習の特徴である。ただし、加える乱数 rnd の値域については認識と動作に分けて調節する必要がある。

６.３　シミュレーション

６.３.１　１次元センサ動作

図６.５に示すように、１次元に配置された６個のセンサセルを持つ左右に移動可能な視覚センサが、大中小（それぞれ視覚センサのセンサセル２個、４個、６個分に相当）の物体を認識するという問題を考える。システム全体は図６.６のようになる。図６.７のように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、正しく認識するためには、物体がある程度視

野の真ん中に見えるようにセンサを移動させることが必要となる。認識用の出力ニューロン数は3個とし、理想出力を大中小の物体それぞれに対し、(0.1, 0.1, 0.9), (0.1, 0.9, 0.1), (0.9, 0.1, 0.1) とした。視覚センサは、1次元の運動を行うため、センサ移動用ニューロンは1個とした。中間層ニューロンは3個設計、ニューラルネットは6－3－4の3層とした。学習の具体的なフローは以下のようになる。まず、物体と視覚センサの相対位置の初期値を、視覚センサからの出力の合計が1以上となる範囲内で、乱数を用いて決定する。そして、物体を見せてから60単位時間経過するかまたは物体が消え失せるまで、つまり、6個のセンサ出力がすべて0になるまで続け、その後次のパターンを見せるということを繰り返し行った。各センサセルは1×1の大きさとし、センサの1回の動作は-0.2から0.2の範囲、つまり式（6.7）の α を0.4とした。

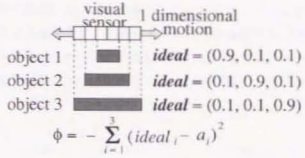


図6.5　シミュレーションで用いたパターン

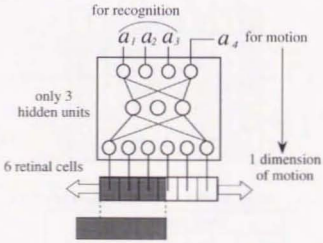


図6.6　シミュレーションのシステム構成

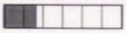


図6.7　どの物体が提示されたか区別が付かない視覚イメージ

学習の結果を図6.8に示す。それぞれ、各物体を見せた時の物体の中心とセンサの中心の相対位置に対する認識用出力、センサ移動用出力をプロットしている。センサ移動用出力を見ると、学習によって、いずれの大きさの物体をいずれの場所に見せた時も、物体を中心近くに捉えるようにセンサが移動するようになっていることがわかる。また、その中心付近での認識用出力を見ると、対応するニューロンの出力値が高い値となっており、正しい認識ができていることがわかる。そして、各パターンとも物体が端の方へいくほど、他のパターンとの区別ができなくなっていることもわかる。このように、区別不可能または区別が困難な場合は、区別ができる方向にセンサを動かし、正しく認識を行っていることから、効率的な認識を実現できたと言える。

中間層ニューロンは前述のように3個で学習することができた。表6.1に学習後の各重み値の値を示す。これより、それぞれの中間層ニューロンの役割は、おおむね、(1)両サイドに物体が見えている場合を検出するニューロン、(2)物体が中心に見え、かつ、物体が小さいものを検出するニューロン、(3)物体が主に視野の右側にあることを検出しているニューロンとなっていることがわかった。つまり、(1)最も大きい物体を識別する、(2)最も小さい物体を識別する、(3)センサを右に動かすべきか左に動かすべきかを決定するためにそれぞれ用いられていることがわかった。そして、中間の大きさの物体の認識出力は、最初の2つの中間層ニューロンが発火しない時に発火するようになっている。

表6.1　学習後のニューラルネットワークの重み値

(a) Input layer → Hidden layer

		hidden layer		
		1	2	3
input layer	1	5.88	-6.12	2.12
	2	2.02	-0.91	2.78
	3	0.44	-2.80	0.17
	4	0.40	-2.78	-0.54
	5	2.01	-1.15	-0.88
	6	5.77	-5.81	-2.90
	bias	-8.46	6.49	-0.23

(b) Hidden layer → Output layer

		output layer			
		1	2	3	4
hidden layer	1	0.56	-5.40	4.88	-0.56
	2	6.49	-6.94	0.60	-0.32
	3	-0.23	0.12	-0.18	6.34
	bias	-2.50	3.03	-2.52	-2.82

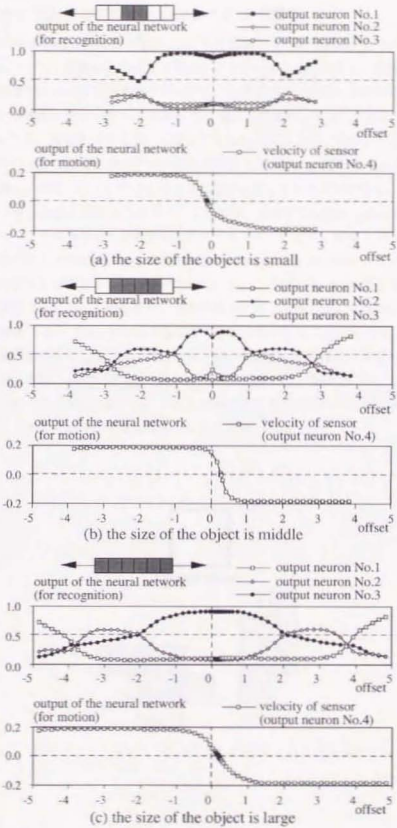


図6.8　シミュレーションの結果（物体とセンサの相対位置に対する各出力）

6.3.2 2次元センサ動作および小さいセンサによる識別

従来、文字認識をする際に、文字の重心にセンサを移動させてから認識を行うという方法がよく用いられてきた。そこで、最後に、認識する物体より小さいセンサを用意し、物体の重心にセンサを持ってきても認識できないような場合についても、上記の方法が適用するかどうかを確認した。視覚センサおよび2種類の提示パターンを図6.9に示す。視覚センサは、2 x 2とした。そして、図6.10(a) のように、視覚センサを重心に持ってくると、どちらのパターンを提示しても、4個のセンサ出力がすべて1になり、パターンの区別ができない。また、図6.10(b) のような場合も区別ができず、逆に、図6.11(a) や (b) のような場合は区別することができる。

これで、学習させた結果を図6.12に示す。ここでは、認識用の出力を1個とし、パターン1の場合は1を、パターン2の場合は0を理想出力として与える。そして、図6.12(a)では、パターン1を提示した場合のセンサの位置に対する出力の値を、図6.12(b)はパターン0を提示した場合の出力の値を1から引いた値を白から黒の色で示している。この図より、図6.11で示したように正しく認識できる場所では、正しく認識をし、物体の重心にセンサ中心がある時のように正しく認識できない場所（図6.10）では、どちらともつかない出力値であることがわかる。また、白抜きの矢印でセンサの動作方向を示している。これから、センサは正しく認識ができるところへ移動していることがわかる。また、白の太線は視覚センサ中心の軌道の例を描いている。このことから、視覚センサは徐々に色の黒い方向に向かい、正しく認識できる場所で止まることがわかる。また、その途中では、一旦認識の評価値が下がる場合もあることがわかる。

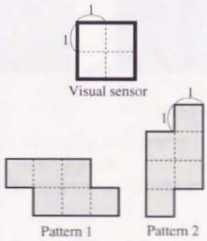


図6.9 シミュレーションで用いたパターンと視覚センサ

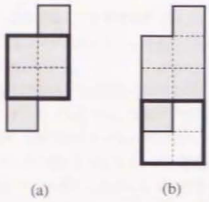


図6.10 識別不可能なイメージ

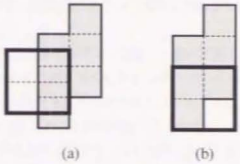


図6.11 識別可能なイメージ

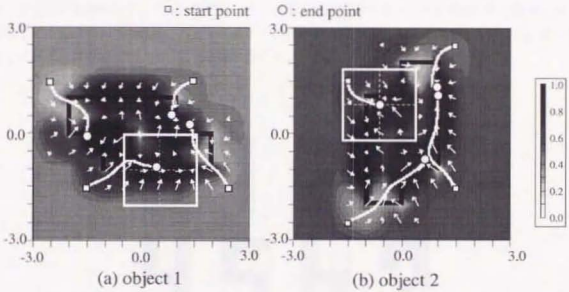


図6.12　学習後のセンサの位置に対する評価値（色）とセンサの動作ベクトル
およびセンサの動作軌跡の例
（黒い太い線が提示パターン、白い四角がセンサを表わす）

6.3.3 簡単な数字認識と能動認識による効率的認識の検証

能動認識を利用したパターンの識別問題の学習において、識別するパターン数がセンサ動作の次元に比べて大きい場合を考える。するとまず、識別パターン数が多いことでパターン間の識別が難しくなる。さらに、認識用出力の数がセンサ移動用出力の数に比べて多くなるが、識別が難しいところではセンサの移動を学習させればよいことから、効率的な認識が実現できる。よって、センサを移動させないであらゆる見え方に対して認識用の出力を学習させる場合と比較して中間層のニュ

ーロン数が少なくてすむ可能性が高いと考えられる。そこで、０から９までの数字を認識させる問題をシミュレーションした。そして、あわせてセンサに２次元動作をさせた場合に問題がないかを確認した。

提示したパターンと視覚センサを図６.１３に示した。視覚センサは、５個×５個の２次元に配列されたセンサセルを持ち、３個×５個の各マス目を０か１に塗りつぶした１０個のパターンを識別させた。１個のセンサセルの受容野および提示パターンのマス目の大きさは共に１×１とし、各センサセルは、投影されたパターンが受容野に占める割合を０から１の間の値で出力するものとした。例えば、"８"を提示した時にセンサ位置をx、y 共に中心から0.5ずつずらした時にセンサから獲得されるイメージは、図６.１４　のようになる。また、認識用出力の数は提示パターン数と同じ１０個とし、理想出力は、各提示パターンに対応する出力を0.9、その他を0.1とした。センサ移動用の出力数は２個とし、それぞれセンサの x 方向 y 方向の移動に用いた。視覚センサは単位時間あたり x 方向 y 方向共に最大0.05移動できることとした。ニューラルネットの中間層ニューロン数は１０個とした。そして、１つのパターンを提示した時の初期センサ位置は、２５個のセンサ出力の合計が１を越える場所からランダムに選択し、２００単位時間経過するか、センサ出力の合計が１を下回るまでの間、センサの動作、ニュラルネットの計算および学習を各単位時間毎に行った。その後新しいパターンを提示し、同じことを繰り返していった。

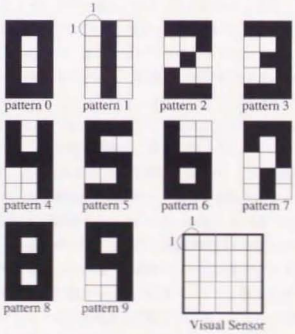


図６.１３　シミュレーションで用いたパターンと視覚センサ

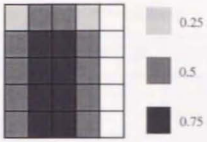


図６．１４　サンプル視覚イメージ（８を提示）

これで、学習させた結果を表６．２および図６．１５に示す。表６．２　では、学習終了後に、各パターンを、初期位置を変化させて1000回提示し、２００単位時間経過した後に、各提示パターンに対応する認識出力が全認識出力中で最大値となった回数を示している。全パターンの平均をとると、99.3%の認識率になり、残りの 0.7%は、見失った、つまり、視覚センサからの信号の総和が１を下回ってしまったことを示す。図６．１５では、０から３までの４パターンを提示した場合について、センサとパターンの相対的な位置に対する認識結果の評価値 ϕ（式（６．２））を濃淡で示し、さらに、その時の動作出力（ベクトル）を矢印で表わしている。ただし、各格子点の内、センサ信号の合計が１を下回る場合は矢印を示していないため、矢印が表示してある部分がほぼパターンを見失ってないセンサ位置となっている。この結果、提示パターン毎に認識しやすいセンサ位置があり、センサをそこに移動させるように学習が進んでいることがわかる。また、全パターンについて最終的にセンサが到達した位置を見ると、３×５の提示パターンの中心$(x,y)=(0,0)$と提示パターンの重心が異なるものについては、全般的に、重心に近いことがわかった。

比較のために、センサを初期位置に固定し、認識だけ学習をさせた場合のシミュレーションも行った。ただし、初期位置を前のシミュレーションと同じにすると物理的に識別不可能なパターンも含まれてしまうため、$-1.0<x<1.0$、$-0.5<y<1.5$ という前のシミュレーションと比較してかなり狭い範囲の中から乱数によってセンサ位置を選択し、学習を行った。

結果を、表６．３に示す。この表は表６．２と同様に、各パターンを 1000回提示した時に、各提示パターンに対応する認識出力が全認識出力の内最大であった回数を示す。これを見ると、誤認識がかなりあり、総平均で 93.2%の認識率となった。これに対し、中間層のニューロンを２０個に増やすと、認識率が98.9%、３０個に増やすと 99.5%と認識率の向上が見られた。これより、中間層ニューロン数１０個で認識率が低かったのは、中間層ニューロン数が少なすぎたことが原因と考えられる。センサの移動まで学習させた場合には、中間層ニューロン数１０個で 99%以上の認識率を実現できたことと比較すると、センサの移動まで学習させることによって効率的な認識が実現でき、より少ない中間層ニューロン数で学習ができたと言える。

表6.2 数字の認識結果

		Number of being maximum output unit (1000 presentations/pattern)										
		0	1	2	3	4	5	6	7	8	9	disappear
Presented Pattern	0	996	0	0	0	0	0	0	0	0	0	4
	1	0	985	0	0	0	0	0	0	0	0	15
	2	0	0	995	0	0	0	0	0	0	0	5
	3	0	0	0	999	0	0	0	0	0	0	1
	4	0	0	0	0	991	0	0	0	0	0	9
	5	0	0	0	0	0	982	0	0	0	0	18
	6	0	0	0	0	0	0	995	0	0	0	5
	7	0	0	0	0	0	0	0	989	0	0	11
	8	0	0	0	0	0	0	0	0	996	0	4
	9	0	0	0	0	0	0	0	0	0	994	6

correct	mistake	disappear
9922	0	78

表6.3 センサを固定した場合の数字の認識結果

		Number of being maximum output unit (1000 presentations/pattern)									
		0	1	2	3	4	5	6	7	8	9
Presented Pattern	0	732	0	0	0	0	193	0	0	75	0
	1	0	1000	0	0	0	0	0	0	0	0
	2	0	0	883	2	0	49	0	66	0	0
	3	0	0	0	974	0	24	0	2	0	0
	4	0	0	0	0	912	2	73	10	0	3
	5	0	0	0	0	0	1000	0	0	0	0
	6	0	0	0	0	10	0	965	17	8	0
	7	0	0	0	0	0	0	0	1000	0	0
	8	0	0	0	0	0	0	0	0	999	1
	9	10	0	0	0	80	0	9	4	42	855

correct	mistake
9320	680

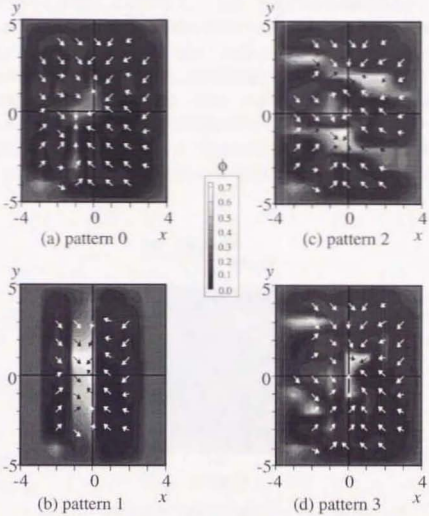


図6.15 学習後のセンサと物体の相対位置に対する評価値とセンサの動作ベクトル

6.4 考察および今後の課題

上記のシミュレーションを通して気づいた点、問題点を述べる。

まず始めに、数字の認識の問題において、センサの移動まで学習させた場合、上記のように、少ない中間層ニューロン数で効率的な学習が実現できたが、さらに学習を進めていくと、逆に認識率が落ちるという現象が見られた。0のパターンを提示した場合の結果を図6.16に示す。この図のセンサの動作ベクトルと評価の値を見ると、図の右上の○で囲んだ部分のように、センサが、提示パターンを見失う方向に動いていってしまったり、その少し下の○で囲んだ部分のように、局所的に見てまわりより正しく認識ができるところにトラップされて、最も認識が正しくできる所まで移動することができず、結果的に誤認識をしてしまっていることがわかった。これは、学習の初期には、全体としての大まかなセンサの動きを学習するが、学習が進むにつれて、評価の曲面のより細かい凹凸に適応したより細かい動作を学習することができるようになり、結果的にローカルなとこ

ろにトラップされてしまうためと考えられる。これは、非常に大きな問題であるが、遅延強化学習において使われるような、後に得られる報酬を事前に評価するような手法を使って解決できるのではないかと考える。ただし、この場合は、最終ゴールをどう設定するかという問題が残る。

また、1回の認識動作の完了をどうするかという問題もある。本論文中のシミュレーションでは、視野内をある程度以上の速度でセンサが動いた時に、センサが端から端まで十分に移動できるということから時間を設定し、その時間の経過を認識動作の完了とした。その他、センサが動作しなくなった時や最大認識出力と2番目に大きい出力の差が一定以上になった時等を認識動作の完了とする規準が考えられる。しかし、我々人間が行っている認識は、文脈や置かれている状況等によって

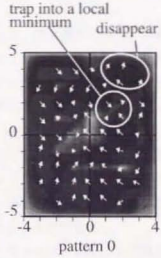


図6.16　学習を続けた場合の評価値とセンサの動作ベクトル

変化してくるため、完了の規準もより複雑なものと考えられる。

本研究は、元々能動認識を学習することによってより適切な動作を行い、より報酬が得られるという発想に基づいたものである。従って、この能動認識の学習を通常の強化学習といかに結びつけていくかという点も今後の大きな課題である。また、前述のような遅延強化学習を適用する場合の最終ゴールや1回の認識動作の完了をどうするかといった問題も、本来の目的指向の強化学習の中に組み込むことによって、特に意識することなく、自ずと学習によって獲得されてくるものではないかと考える。

また、ここでは、認識のコーディングを、従来文字認識等でよく使われるように、各パターンに対して出力を1つ割り振って、そのパターンが提示された時はその出力の理想値を0.9、他を0.1とした。しかし、これも効率的なパターンコーディングとは言えない。そこで、これに関しては、強化学習と統合することで自律的にコーディングまで決定されることが期待される。

次に、行ったシミュレーションにおいて、学習の初期には、センサは、その位置によらず直線的にある方向に動き、その結果視野から物体が見えなくなった。そこで、物体が見えなくなったらペナルティを与えた、つまり、式（2）で求めた評価値からさらにある値（ここでは、0.2）を引くと

いう設定を行ったところ、物体が視野の外に消えないようにするセンサの移動が早期に学習でき、学習の加速が見られた。我々人間も、物体が視野から消える状態を他から区別して認識しているように見える。このことから、我々も視野から物体が消えた状態を検出し、何らかのペナルティの信号を生成する機構を持っている可能性も考えられる。また、前章の物体追跡のモデルのところで述べたように、外界の情報の変化がなくならないようにという学習によって物体の追跡動作を学習させることが可能である。このモデルとの融合も有効であると考える。

また、上記で用いたニューラルネットは、リカレント構造になっていないため、過去の履歴を認識や動作に利用することができない。しかし、我々は、例えば物体を見失っても、消えていった方に目を動して物体を再び見えるようにすることができる。このように、過去の履歴を反映した認識および認識のための動作を実現するためには、リカレント型のニューラルネットの導入が今後必要であると考える。

我々が物体を認識する時を考えると、通常視野の中心に物体を捉えるが、本論文で提案したような学習によって説明が付くのではないかと考えられる。数字の認識の実験では、動作後の視覚センサの中心が提示パターンの中心よりも提示パターンの重心に近くなる傾向があることを述べた。このような傾向は、単純な図形を人間に見せた時の視点の留まる位置に関する Kaufman ら実験結果[Kaufman69]でも見られる。また、小さい視覚センサを用意し、提示パターンの重心にセンサを移動すると逆に認識できないような設定で学習を行ったところ、重心ではなく、認識しやすい位置にセンサを動かして正しい認識を行うこともできた。また、視野の中心でパターンを捉えることが多いことから、われわれの目の網膜上の視細胞の密度が中心ほど大きいという不均一性は、能動認識を行うための効率的な形であると考えることができる。

また、本論文では、1つのニューラルネットの出力を認識用とセンサ移動用に分けるという形をとったが、認識用とセンサ移動用にニューラルネットを分割することも可能である。最初のシミュレーションにおいては、中間層のニューロンが、(1)(2)は認識用に、(3)はセンサ移動用に使われていることから、ニューラルネットを一体化したことによる中間層ニューロン数の減少の効果は見られなかった。しかし、数字の認識のシミュレーションでは、認識用ニューロンとセンサ移動用ニューロンの両方と大きな結合を持つ中間層ニューロンが存在したことから、両ニューラルネットを一体化したことによって中間層ニューロンの効率的な利用に結びついた可能性が大きい。また、それ以外には、ニューラルネットを一体化することによって、与えられた問題に対し、中間層ニューロンを認識用とセンサ移動用に柔軟に配分できるという利点が考えられる一方、学習の収束が遅くなるという欠点が予測される。

最後に、我々生物との比較を行う。我々人間は、通常、サッケード、つまり、跳躍的な視点の移動を行うことによってパターンを認識していると言われている[乾 93]。本学習では、視覚センサの動きは連続的であり、そこが大きく違うところである。サッケードを行う場合でも、跳躍先をどこに持っていくかは学習が必要であり、こう考えることによって本学習が適用できると考える。また、サッケードを行う場合、サッケード中には認識を行わなくても良いため、移動中の認識のための時間が節約されるという考え方もできる。今後、両者を比較検討し、生物の良い点を取り入れていく必要がある。

6.5 まとめ

認識や認識のための動作を目的達成のための動作の一つであると捉えることにより、強化学習をによって両者を学習することを提案した。そして、能動認識におけるセンサの動きと認識自身を強化学習によって並列に学習させるためのシステムの構成と学習方法を提案した。認識も認識のための動作も共に強化学習を用いているため、ほとんど同じ学習方法で両者の学習を実現することができた。

この学習システムを用いて、簡単な認識問題のシミュレーション行ったところ、このシステムは、センサの動かし方に関する情報を一切与えていないにもかかわらず、物理的にパターンを識別できないような位置にセンサがあっても、センサを動かして正しい認識を行うように学習ができた。ただし、学習をさらに進めると、認識に対する評価値のローカルミニマにトラップされたり、視野からはずれる方法へセンサが動いてしまうという状況が見られた。

また、うまく認識できた時の最終的なセンサの位置は、提示パターンの重心近くにいく傾向が見られたが、単純に重心にセンサを移動させると区別ができないようなパターンを用意したところ、区別ができるところにセンサを移動させる学習を行うことができた。

また、0から9の１０個のパターンについて簡単な文字認識の問題を学習させたところ、全てのパターンについて、センサの動作後ほぼ正しい認識ができるようになった。そして、センサを動かさないであらゆる見え方に対して認識を学習した場合と比較して、ニューラルネットの中間層のニューロン数が少なくても高い認識率を示すことがわかった。また、学習後のニューラルネットを解析したところ、中間層ニューロンでは、認識、センサ動作を決定するために、センサ信号である入力空間を効率的にコーディングしていることもわかった。

第７章　時間軸スムージング学習に基づく遅延強化学習

第１章で述べたように、自律学習の範疇に入る代表的な学習アルゴリズムに強化学習がある。本章では、報酬や罰（強化信号）が一連の動作の後に得られる場合に、そこまでの動作をいかに学習するかという遅延強化学習の問題に時間軸スムージング学習が適用できることを示す。そして、ニューラルネットを用いることにより効率的な学習が実現できること、視覚センサ信号を直接入力としても学習を行うことができ、その時、ニューラルネットの中間層に、空間の情報がどのようにコーディングされるかといった点を移動ロボットのシミュレーションを通して示す。そして、これらから、自律学習とニューラルネットの有効性について述べる。

７.１　背景

１.２節で述べたように、遅延強化学習の問題に関しては、1983年の倒立振子の制御の学習を例題として扱った Barto らの critic-actor アーキテクチャ（ＴＤ学習）[Barto 83] が有名である。ここでは、現在のセンサからの入力を元に、現在から将来にわたって得られる強化信号 r に、現在からその強化信号が得られるまでの時間 t によってディスカウントファクタ γ^t $(0 \le \gamma \le 1)$を掛けたものの総和

$$\bar{r} = \sum_{t=0}^{\infty} \gamma^t r(t+1) \qquad (7.1)$$

を予測するように critic 部が学習し、この総和を最大化するように actor 部で動作を学習する。この critic 部の $\bar{r}$ の予測は、ＴＤ(Temporal Difference)学習[Sutton 88]を適用し、逐次的に学習を行う。また、このＴＤ学習を用いて $\bar{r}$ を学習すること、またはこれを用いて強化学習の問題を解くことを単にＴＤ学習と呼ぶ場合もある。元々のＴＤ学習は、予測問題の逐次的学習方法である。例えば、月曜日に次の日曜日に晴れる確率を予測することを学習によって獲得するという問題を考えた時に、実際に日曜日が来るまで学習しないのではなく、火曜日の予測値は月曜日の予測値よりも信頼性が高いという観点から、月曜日の予測値に対し火曜日の予測値を教師信号として学習し、火曜日の予測値を水曜日の予測値から学習するといったように、逐次的に予測の学習を行う方法である。これを $\bar{r}$ の予測値 $P(t)$ の学習に適用すると、（７.１）式から

$$
\begin{aligned}
P(t) &= \sum_{i=0}^{\infty} \gamma^i r(t+1) \\
&= r(t+1) + \gamma \sum_{i=0}^{\infty} \gamma^i r(t+2) \\
&= r(t+1) + \gamma P(t+1)
\end{aligned}
\qquad (7.2)
$$

を満たすようになればよいことになる。そこで、$P(t)$ に対して１単位時間経過後に得られる（７．２）式の右辺の値を教師信号として学習を行うことによって実現できる。これによって、倒立振子が倒れた時に罰を与えるだけで、倒立振子が倒れないような制御を獲得することができる。Barto らは、critic 部、actor 部共に、センサの入力を人間が予め決めた方法で場合分けし、それに対してテーブルルックアップ方式で $\bar{r}$ や動作を決定している。その後、Anderson らは、この学習に多層ニューラルネットを用い、学習にＢＰ法を用いることによって倒立振子が倒れるまでの回数が飛躍的に伸びることを示している[Anderson 89]。

本章では、この遅延強化学習の問題に対し、時間という概念を陽に捉えることによって、前述の時間軸スムージング学習の適用を試みた。この時間軸スムージング学習は、前述のように、センサ信号の統合化による空間情報の抽出にも用いることができ、より汎用的な学習則である。そして、これによって、前述の Barto らの方法とほぼ同様な機能を実現できることを示す。さらに、この学習アルゴリズムの問題点を整理し、より適応的な強化学習のアルゴリズムを提案すると共に、視覚センサ信号を直接入力した場合について検討する。

７．２ 学習アルゴリズム

システム（ロボット）の構成は、図７．１のように、強化信号検出部、状態評価部、動作生成部の３つの部分から構成され、状態評価部、動作生成部はニューラルネットによって形成する。そして、システムの状態を評価する状態評価部（評価関数）を遅延強化信号を用いて学習し、さらに、その

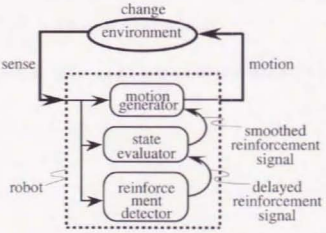


図７．１ 遅延強化学習システムの構成

評価値を用いて動作を学習する。この構成自体は、Barto らと同じものであり、critic 部がここでは状態評価部 (State Evaluator)、acotor 部が動作生成部 (Motion Generator) に相当する。

7.2.1 動作の学習

ここで述べる動作の学習は、基本的に Williams らの方法 [Williams 88] と同様である。彼らは、確率的動作という観点から説明を行っているが、ここでは、試行錯誤という観点から説明する。簡単のため、ここでは、システムの状態が２つの連続的な変数で表されるものとする。そして、図７．２のように、２個の状態変数（x、y 軸）と評価関数値（z 軸）から形成される３次元空間を考える。ロボットは、A 点から動作生成ニューラルネットの出力である指令動作（速度）ベクトル $\boldsymbol{m}$ に乱数ベクトル $\boldsymbol{rnd}$ を加えたベクトルに従って実際の動作を行う。この乱数ベクトル $\boldsymbol{rnd}$ は $\boldsymbol{m}$ に比べて微小でかつ確率分布は状態変数空間上で原点対称とする。これは、状態変数の時間変化を

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{f}(\boldsymbol{m}) \qquad (7.3)$$

と表すことができ、この $\boldsymbol{f}$ が $\boldsymbol{m}$ によって微分可能であると考えればよい。そして、ある状態Aから微小な単位時間で動作できる領域を、簡単のため図中の斜線で塗りつぶされた円のように表わされるものとする。ここで、動作生成ニューラルネットの出力値に対する教師信号ベクトル $\boldsymbol{m}_s$ を、

$$\boldsymbol{m}_s = \boldsymbol{m} + \zeta \boldsymbol{rnd}\, \Delta\Phi \qquad (7.4)$$

$\triangle\Phi = \Phi(\boldsymbol{x}(t+1)) - \Phi(\boldsymbol{x}(t))$: 単位動作による評価値 Φ の変化量
$\boldsymbol{x}(t)$: 時刻 t でのシステムの状態、　ζ : 定数

とロボット内部で自動生成し、ニューラルネットを学習させる方法を考える。こうすれば、$\Delta\Phi$ が大きい時の $\boldsymbol{rnd}$ の方向に動作がより強化されることになる。ここで、単位時間は十分に短いとし、$\Delta\Phi$ が微小であったとすると、

$$\Delta\Phi = (\boldsymbol{m} + \boldsymbol{rnd})\nabla\,\Phi(\boldsymbol{x}) \qquad (7.5)$$

となる。そして、微小な乱数ベクトル $\boldsymbol{rnd}$ を変化させた時の教師信号ベクトル $\boldsymbol{m}_s$ の期待値 $\overline{\boldsymbol{m}_s}$ を求めると、

$$\bar{\boldsymbol{m}}_s = \boldsymbol{m} + \zeta \frac{\iint_{|\boldsymbol{rnd}|<rnd_{max}} \boldsymbol{rnd}\{(\boldsymbol{m}+\boldsymbol{rnd})\nabla\Phi(\boldsymbol{x})\}\, d\,\boldsymbol{rnd}}{\iint_{|\boldsymbol{rnd}|<rnd_{max}} d\,\boldsymbol{rnd}}$$

$$= \boldsymbol{m} + \zeta \frac{\iint_{|\boldsymbol{rnd}|<rnd_{max}} (\boldsymbol{rnd}\cdot\boldsymbol{m}+|\boldsymbol{rnd}|^2)\, d\,\boldsymbol{rnd}}{\pi \cdot rnd_{max}^2} \nabla\Phi(\boldsymbol{x}) \qquad (7.6)$$

となり、$\boldsymbol{rnd}$ が原点対象であることから、

$$\bar{\boldsymbol{m}}_s = \boldsymbol{m} + \zeta \frac{\iint_{|\boldsymbol{rnd}|<rnd_{max}} |\boldsymbol{rnd}|^2\, d\,\boldsymbol{rnd}}{\pi \cdot rnd_{max}^2} \nabla\Phi(\boldsymbol{x}) = \boldsymbol{m} + k\nabla\Phi(\boldsymbol{x}) \qquad (7.7)$$

$$k = \zeta\, rnd_{max}^2 / 2$$

$\boldsymbol{m}$ から評価関数曲面の最急勾配方向に変化したものであることがわかる。これによって、常に動作生成と学習を区別せずに行うことができる上、動作ベクトルが評価関数の最大傾斜方向に確率的に変化していくことがわかる。ただし、動作可能範囲が状態変数空間上で凹部を持つと、動作がローカルマキシマムに陥る可能性があるが、それがなければ、与えられた評価関数に関して、最適動作へ近づいていくことができる。ただし、この学習には、評価関数曲面が微分可能であることが必要である。また、本論文中では、通常は小さい乱数で、たまに大きな乱数が発生させるように、一様乱数の３乗を用いた。

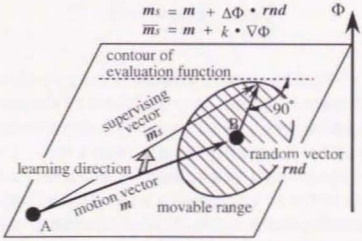


図７.２　試行錯誤に基づく動作の学習

7.2.2 評価関数の学習

最初に、ロボットの状態が目的達成にいかに近いかを評価する時の規準について考える。まず、図７.３(a)のような２次元空間上の位置関係に、ロボットとターゲットがあるとする。まず、両者間の距離による評価が容易に考えつくが、この２つの次元の間でどう正規化すべきかを考えることは困難であるし、そもそも、距離を知る方法すらロボットは知らない。さらに、図のようにロボットとターゲットとの間に障害物がある等の場合は、目的達成度を単純に距離だけでは表わすことはできない。そこで、目的達成までに要する時間によって現在の状態を評価することを考える。こうすれば、図７.３(b)のように各状態は時間という１次元空間に投射され、正規化の問題も、障害物の問題も解決される。ところが、目的達成までに要する時間は、目的達成後しかわからない。しかし、自分が通った経路すべての場合のセンサ入力等の情報を保持し、経路をさかのぼって評価を学習することは、記憶容量などの面から大変に無駄が多い。

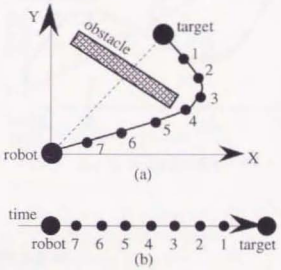


図７.３　所要時間による評価

図７.４は、時間の変化に対する状態の評価値の変化の様子を示したものである。図中の太い点線のように評価値が初期状態から目的達成状態までの間、時間に対して直線的に単調増加すれば、目的達成までの所要時間で評価を行ったことになる。ところが、実際の評価関数の時間変化が実線のように変化したとすると、状態Ａの方が、状態Ｂより評価が良いことになり、所要時間による状態評価に反する。このような評価の逆転は、評価値の時間変化に凹凸がある場合に起こる。時間変化に対する凹凸は、時間の２階微分値で表わされるため、これを０を近づけるという時間軸スムージング学習を行えば理想の評価関数に近づいていく。ただし、初期状態は評価値を低く、目的達成状態では評価値を高くなるように学習する。そこで、評価値の値域を０以上１以下とした場合、図中の矢印のように、初期状態の時に0.1、目的を達成した時に0.9を教師信号として学習し、かつ、その他の部分では常に教師信号Φ_sを

$$\Phi_x(x(t)) = \Phi(x(t)) + \xi \frac{d^2\Phi(x(t))}{dt^2} = \frac{\Phi(x(t-1)) + \Phi(x(t+1))}{2} \qquad (7.8)$$

ξ: 定数

と内部で自動生成して学習させる。実際は、１単位時間さかのぼって学習を行うことになる。評価値の時間に対する２階微分値は、動作しながらその都度知るができるため、過去にさかのぼることなくリアルタイムの学習が実現できる。

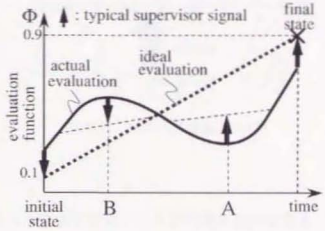


図７.４　評価関数の学習

7.2.3　２点間経路最適化の原理

本節では、この評価の学習と動作の学習を行うことによって、経路が最適化されることを示す。ある時、ロボットが、図７.５の経路(b) のように大回りをして点Ａから点Ｂまで行っていたものとする。この場合、点Ａから点Ｂまで経路(a) のように真っ直ぐに進めば８単位時間で到達できるが、経路(b) では１１単位時間かかる。ここで、ロボットは自分の通った経路に対し、評価関数の時間変化を一定になるように学習する。また、評価の値は、ニューラルネットの出力であるため、その汎化能力によって自分の通った経路の近傍も滑らかになる。点Ａと点Ｂの評価値は、どちらの経路を通っても同じ値となるため、この図の領域が微小な領域であるとすると、$d\Phi/dt$ の値は経路Ａの方が大きくなる。すると、7.2.1 で説明したように、動作は学習により $d\Phi/dt$ の値が大きい方向へ変化していく。従って、経路は矢印のように、徐々に最適化されていく。

これによって、評価関数曲面に無意味な凹凸ができ、凹のところは避けて通り、凸のところに引き寄せられるようになった場合でも、凸の部分を行ったり来たりしているうちに、凸の部分は次第にへこみ、凹の部分を避けて尾根の部分を通るように学習した場合も、乱数成分によって尾根からはずれ、また尾根に引き戻されるという動作によって凹の部分は徐々に盛り上がり、前述のように

動作の最適化ができることがわかる。ただし、評価関数がターゲットから離れる方向に単調に増加しているようになった場合は、このロボットは単にターゲットから離れていくだけで、うまく学習できない。また、学習の初期では、乱数成分のみによる動作となるため、偶然目的を達成するまで待つことは非常に多くの時間を必要とする。従って、最初は簡単な学習から始め、徐々に難しくしていくという方法によって学習を加速させる。このような方法は逐次接近法と呼ばれ、実際の生物でもその効果が観察される[東 69]。

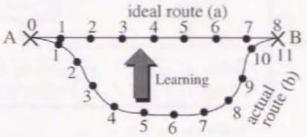


図７.５　２点間経路最適化の原理

7.2.4 ニューラルネットによる学習システムの構成と学習方法

上記のような原理によって動作するニューラルネットの構成と学習方法を示す。図７.６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（状態評価ネット）と、同じ入力から動作信号を出力するニューラルネット（動作生成ネット）の２つのネットワークからなる構成とする。ここでは、それぞれのネットワークは通常の階層型とし、出力は０から１の間の値で表わし、それぞれ自動生成された教師信号に近付くようにＢＰ法によって学習を行う。動作ネットの方は、評価ネットの時間変化量、つまり、時間に関する１次微分値を、評価ネット自体には２次微分値を用いて学習する。また、この時の学習は収束するまで繰り返すのではなく、１単位動作につき１回のみ繰り返す。そして、動作生成ネットの出力に乱数成分を加えた値に従って単位動作を行う。動作信号が複数ある場合は、それぞれに対して別々の乱数成分を加える。そして、この動作と学習を目的達成まで繰り返し行う。ただし、両ネットとも、学習を全く行っていない状態では出力層ニューロンへの結合の重み値を全て０とし、入力にかかわらず常に中間の値である 0.5 を出力するようにした。そして、この時の動作を０とし、当初は乱数のみの動作になるようにした。

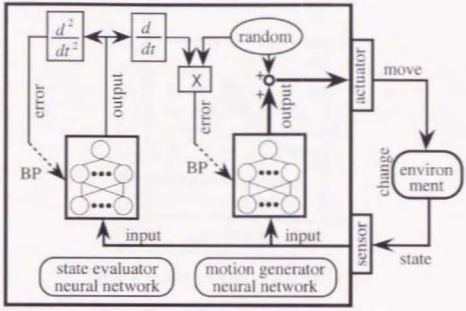


図7.6 学習システムの構成

7.3 シミュレーション

7.3.1 経路最適化に関する基本シミュレーション

まず始めに、経路が最適化されるかどうかを確かめるための基本的なシミュレーションを典型的な2つの場合について行った。一つは、評価関数に最終目的点以外に凸部ができてそこにトラップされてしまった場合、もう一つは、評価関数に無意味な凹部ができて、そこを避けるように学習してしまった場合である。凸部にトラップされてしまった場合は、図7.7のように、試行錯誤の乱数成分の影響で、凸部を行ったり来たりすることになる。すると、時間軸スムージング学習によって、凸部は徐々にへこみ、最終的には凸部から抜け出すことができると考えられる。

そこで、まず、正方形の平面を考え、この上をロボットが動くような環境を考える。そして、ロボットは、この平面の x 座標と y 座標をニューラルネットに入力し、評価の出力と2つの動作の出力を計算するものとする。そして、2つの動作出力にしたがって、それぞれ x 方向と y 方向に動作するものとする。この時、まず始めに、平面上の中心の評価値が大きく、周辺に行くほど小さくなるように、平面上のいくつかの点に対して評価値の教師信号を与え、単純な教師あり学習で評価の学習を行うと共に、動作は7.2.1で述べた動作の学習にしたがって、より評価値が高くなる方向への動作を学習させる。この学習を行わせたものを図7.8(1)に示す。中心に評価関数のピークができ、さらに動作ベクトルはその中心に向かっていることがわかる。この状態でロボットを平面の中心に置き、前述の学習を行う。すると、かなり時間が掛かるが、最終的にロボットは尾根から脱出することができる。脱出した時の評価関数と動作の様子を図7.8(2)に示す。評価関数の尾根が低くなり、動作ベクトルの向きからロボットが抜け出せていることがわかる。この時の、学習の進行による評価値の最大値と最小値の変化の様子を図7.9に示す。この図からも、最大値と最小値の差が徐々に小さくなっていることがわかる。

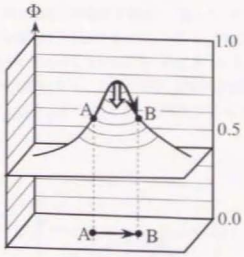


図７．７　評価関数の凸部にトラップされた場合

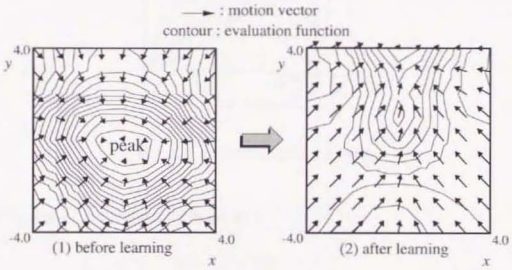


図７．８　評価関数の凸部にトラップされた状態から学習を積んだ場合

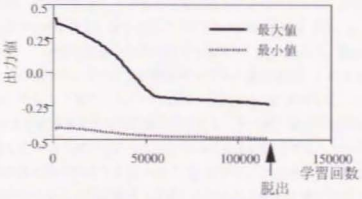


図７．９　学習による評価値の最大値・最小値の変化

次に、図７．１０のように、評価関数に無意味な凹部ができ、そこを避けるような経路を学習してしまった場合、経路の最適化が進むかどうかをシミュレーションした。この場合も、やはり試行錯誤の乱数成分により、ロボットが軌道からはずれるが、軌道からはずれて再び軌道に戻る際に、時間の変化に対して評価値の変化は凹になる。したがって、凹部の評価値は徐々に盛り上がり、最終的には評価関数曲面はフラットに近づき、経路は7.2.3で見たように、最適な経路に近づいていくと期待される。

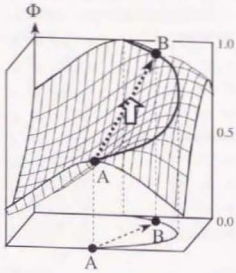


図７．１０　評価関数に凹部ができた場合

そこで、今度は、まず評価関数の教師信号 Φ_s を

$$\Phi_s = 0.4\exp\left(-\frac{(y-\frac{3}{16}x^2+3.0)^2}{4}\right)+0.05x-0.2 \qquad (7.9)$$

として、ニューラルネットに学習させると同時に、動作もニューラルネットの評価出力にしたがって学習させた。この時、単位時間あたりの動作可能範囲を円形にするため、動作出力に対する教師信号ベクトルの大きさが0.5を越えた場合には、ベクトルの方向を変えずに、大きさを0.5に正規化してから学習に用いた。この時の評価と動作の様子を図７．１１(1) に示す。評価関数の尾根が放物線状になっていることがわかる。また、太い線がサンプルの軌跡を示しているが、ほぼ尾根に沿っていることもわかる。そして、次に、スタート地点を$(x, y)=(-4.0, 0.0)$とし、これを（７．９）式に代入した値である -0.2 を教師信号として評価の学習をし、その後、前述の学習則で評価と動作を学習し、$x=4.0$ に到達した時点で再び（７．９）式で求めた値を教師信号として評価を学習させた。これを繰り返した後の評価と動作の様子を図７．１１(2) に示す。この図より、学習を積むことによって、評価関数の凹部がなくなり、ほぼ最適（直進）と思われる軌跡を通ってゴールに到達するよう

になっていることがわかる。また、その際、評価関数の等高線は、経路（図中の太線）に関してほぼ等間隔になっていることがわかる。ただ、経路以外に関しては評価関数が学習されていないので、特に、評価関数が y が正の場合と負の場合で対称になっていない等、所要時間を表すようになっていない。この学習によるスタートからゴールまでの到達時間の変化を図７．１２に示す。所要時間８０がこの場合最適値であるが、学習によって、経路の最適化が行われていることがわかる。

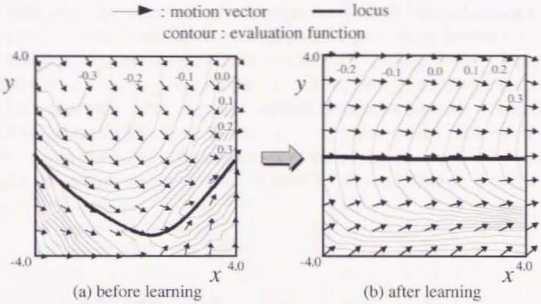


図７．１１　評価関数の凹部ができた状態から学習を積んだ場合

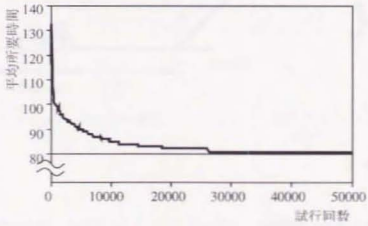


図７．１２　試行回数による平均所要時間の変化

7.3.2 非対称動作特性を持つ移動ロボットのシミュレーション

上記のアルゴリズムを検証するために、図７．１３のような環境で移動ロボットが目標物を取り込むという問題をシミュレーションした。ロボットは、図の位置に置き、目標物(Target) を図中の太線上にランダムに置く。そして、ロボットが動いていき、目標物に到達するまで、動作と学習を繰り返す。ここで、目標物に到達するとは、目標物の中心がロボットに接触した時とする。ただし、目標物が自分より後ろに行ってしまった場合は、失敗として教師信号 0.1 を与えて学習させる。そして、到達するか失敗したら、そこまでを一試行とし、再びエサの位置をランダムに設定する。ロボットが得られる入力としては、ロボットから目標物を見た時の前方向と横方向の相対距離とし、動作生成ネットからの２つの出力にそれぞれ乱数成分を足した値で示された角度だけロボットの左右の車輪を回す。また、図７．１４のようにロボットの動作特性に非対称性を持たせた。具体的には、右側の車輪は、ニューラルネットの出力と乱数を足したものを３倍にした数に従って回転させ、左側は１倍して回転させた。また、このロボットは学習前は試行錯誤の乱数成分だけでしか動作をできないため、初めから目標物を遠くにおいておくと、いつまで待っても目標物に到達することができない。そこで、前述のように、簡単な問題から徐々に問題を難しくするために、ロボットがある時間経過しても目標物に到達できない場合は、目標物をロボットの近くに移動させるということを行う。

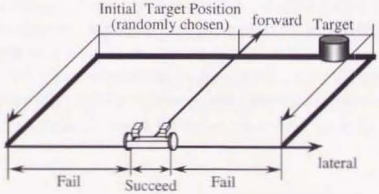


図７．１３　シミュレーションの環境

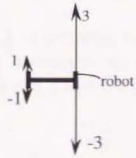


図７．１４　シミュレーションで仮定したロボットにおける非対称な動作特性

学習後の評価と動作の様子を図７．１５に示す。この図は、評価関数の値を表す等高線と、目標物を５カ所に置いたそれぞれの場合のロボットの軌跡を１０単位時間毎にプロットしたものである。ただし、この図は、ロボットを中心にした座標で描いてあるため、目標物の位置によって評価が決まると共に、ロボットの動作によって、相対的に目標物がロボットへ近づいてくる。これより、学習後は、形成された評価関数曲面の尾根が、ロボットの近くから徐々に左前方に伸びていることが

わかる。そして、ロボットは、目標物が遠くにある時には、まず自分の左前方に目標物が見えるように回転している。これは、前述のように、このロボットが右側の車輪をより速く回すことができるため、たとえ目標物がロボットの右遠方にあっても、回転の遅い左側の車輪をめいっぱい回転させて右前方に移動するより、一気に回転して左前方に目標物が見えてから、右の車輪を使って左前方に前進した方が有利であることを学習した結果であると考えられる。さらに、目標物が近くに来ると、逆に、目標物を自分の右側に持ってくる。これも、このロボットの動作特性によるもので、右の車輪をたくさん回転させることができるため、自分の近くにある時は、右側に目標物があった方が早く捕らえられるためであると考えられる。それに対し、比較のため、近い程良いという評価関数を予め与え、動作のみを学習させた場合を図７.１６に示す。この場合は、回転して自分の前方に目標物が見えるようにし、真っ直ぐ前進するという動作を学習している。以上の結果を、絶対座標に直したものを図７.１７に示す。評価関数を学習せずに動作だけ学習した場合は、評価関数と動作を学習させた場合よりも経路が直線に近く、一見よりよい経路に見える。しかし、ロボットが目標物に到達するのに要する時間は、後者の方が短い。また、図７.１８に5000試行後と30000試行後のロボットの経路を示す。学習を積むことにより経路が最適化が進んでいることがわかる。

図７.１９に試行の数による目標物到達までの所要時間の変化をプロットしたものを示す。縦軸は目標物を１２カ所に置いた時の平均所要時間を示している。また、1000単位時間経過しても目標物に到達できない場合は、所要時間を1000とした。また、図中には、評価関数を与えた時と学習させた時それぞれについてニューラルネットの初期値を変化させた２つの場合をプロットしている。これを見ると、学習の初期は、評価関数を与えた方が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。ところが、試行を積んでいくと、評価関数を学習する方は、環境に適応し、経路の最適化が進むため、最終的には、重み値の初期値によらず、評価関数を学習させた方が速く目標物まで到達できるようになっていることがわかる。

このように、評価関数と動作を共に学習することにより、具体的な知識を与えることがなくても、学習を積むことによって最適に近い動作を獲得できることがわかった。そしてその動作は、時として我々が想像していないものであるが、それがロボットにとっては良い動作であることがわかった。

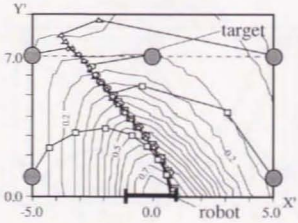


図７.１５　学習後の評価関数とロボットの経路（ロボット固定の座標）

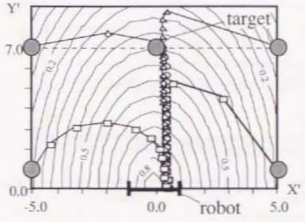


図７.１６　比較実験（評価関数固定）における評価関数とロボットの経路（ロボット固定座標）

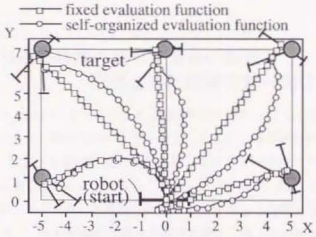


図７.１７　評価関数を学習させた場合と与えた場合の絶対座標におけるロボットの経路の比較
（絶対座標）

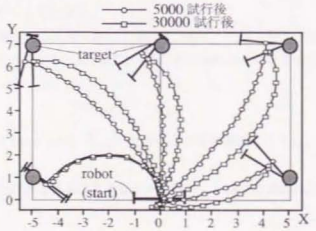


図７.１８　学習によるロボットの経路の変化（絶対座標）

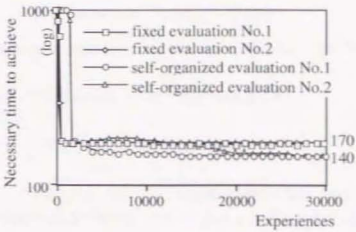


図７.１９　試行による target 到達までの所要時間の変化

７．４　試行により所要時間の異なる場合の評価と評価値の時間変化量一定化学習

前述のシミュレーションにおいて、ロボットの動作特性を非対称にする代わりに、ロボットと対象物との相対位置関係によって動作特性が変化するようにした。具体的には、ロボットから見た対象物の横方向の距離 x' によって動作特性を変化させ、ロボットの左側の方ではゆっくりしか進むことができず、右側では速く進むことができるようにした。すると、図７.２０に示すようにロボットの経路は最適化されず、ゆっくりしか進めない場所を通るような経路となった。この原因を探ってみた。図７.２１のように、左右対称の位置に対象物を置き、全く対称的な経路を通っている場合を想定する。この時、route(a) は時間が掛かり、route(b) は短時間で対象物に到達することができる。すると、評価の時間変化を滑らかにするだけでは、図７.２２のように、時間当たりの評価値の変化量が一定にならない。従って、所要時間による評価が正しくできないため、点Ａと点Ｂの評価値を比較すると点Ａの評価が高いという逆転現象が起こる。従って、ロボットの経路はより遅くなる方へ学習によって進んでいくことになる。

そこで、評価を時間軸に対して滑らかにするだけでなく、時間に対する評価の傾き $d\Phi/dt$ が一定になるように学習を行うこととした。具体的には、単位時間あたりの理想評価値変化量 $\Delta\Phi_{ideal}$ を

$$\Delta\Phi_{ideal} = V / N_{max} \qquad (7.10)$$

V: 理想値域、ここでは 0.9-0.1=0.8、N_{max}: 過去の最大所要時間（ただし、１次遅れで減衰させる）

によって求め、これと現在の評価値の変化量とを比較し、１単位時間前の評価値に対し、

$$\Phi_s(t-1) = \Phi(t-1) - \eta(\Delta\Phi_{ideal} - \Delta\Phi(t)) \qquad (7.11)$$

Φ_s：評価値に対する教師信号、　$\Delta\Phi(t) = \Phi(t)-\Phi(t-1)$、　η：学習のための定数

という教師信号を生成して学習を行い、さらに、現在の評価値に対し

$$\Phi_s(t) = \Phi(t) + \eta(\Delta\Phi_{ideal} - \Delta\Phi(t)) \qquad (7.12)$$

という教師信号によって学習を行うことによって評価値の時間変化量を一定化することにした。具体的には、現在の評価値の変化量が理想値より大きい場合は、１単位時間前の評価値を上げて現在の評価値を下げ、逆に、現在の評価値の変化量の方が小さい場合は、前の評価値を下げて現在の評価値を上げる学習を行う。

これによって学習した結果を図７．２３に示す。環境は、前述のように、ロボットから見た対象物の位置である X' 座標が大きいほど速く進めるように設定した。この図を見ると、ロボットが速く進める場所を通る経路を学習によって獲得できていることがわかる。

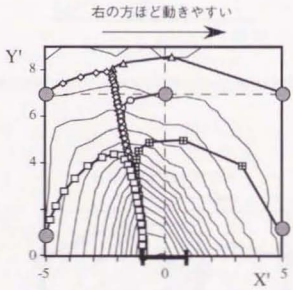


図７．２０　非対称動作環境での学習後の評価関数とロボットの経路

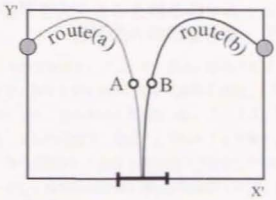


図７．２１　非対称動作環境での対称的な経路

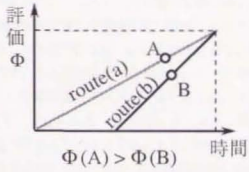


図７．２２　非対称動作環境で、対称的な経路をとった場合の評価関数の時間変化

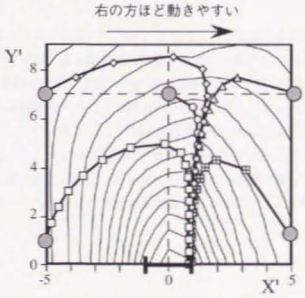


図７．２３　評価値の時間変化量一定化学習の際の非対称動作環境での学習後の評価関数と経路

７．５ 視覚センサ信号を入力とした場合のシミュレーションと評価値の時間変化量一定化学習の学習方法

より現実に近いロボットを想定するということで、前述の移動ロボットが持つセンサとして視覚センサを用いることを考える。視覚センサを強化学習に適用した例としては、１.２節で述べたように、浅田らが行ったサッカーロボットの例がある[浅田 95]。ここでは、視覚センサの信号からボールやゴールの見える大きさ、左右の位置等を検出するプログラムを書くことによって、人間が予め状態空間を分割してやり、その分割された状態と動作のペアに対してQ学習を適用してシュートの動作を学習させている。しかし、遅延強化学習の考え方を利用すれば、視覚センサの出力から直接評価を学習することが可能であると考えられるし、ニューラルネットを用いて学習を行うことによって、より適応的に、また滑らかな状態の評価を行うことができると考えられる。

そこで、図7.２４のように、センサを視覚センサに置き換えて学習を行わせた。すると、図７．２５に示したように、ロボットが目標物にたどりつく直前にループに陥ってしまい対象物までなかなか到達しないという現象が見られた。このループに陥った時には、ロボットへの視覚入力が常に一定になっており、評価値の時間変化量を理想値に近づけようとしても物理的に不可能なってしまう。従って、所要時間での正しい評価が行えないという状況になり、経路の学習がうまく進まないということが判明した。そこで、これを解決するためには、ループやそれに近い状況に陥った時には、周辺の評価値を下げるという学習を行うべきであると考えた。そして、傾き一定化学習を、式（7.１１）および式（7.１２）（図7.２６ の左の図）のように、理想の傾きと実際の傾きの差を求め、１単位時間前の評価値からそれだけ小さくなるように学習し、現在の評価値に対し、その分だけ大きくなるように学習を行う代わりに、図7.２６の右図のように、１単位時間前だけ式（７．１１）に従って学習を行い、現在の評価値に対しては学習を行わないようにした。これによって、経路がループになってしまうと評価の時間変化量が０に近づいていくため、経路全体で評価値が下がるという学習が行われる。また、完全にループにならない場合でもそれに近い状態になると評価値が下がるため、動作の学習によって、経路はループに近づかないように学習によって変化していくことになる。

そこで、この学習アルゴリズムを用いて、図7.２４のような視覚付き移動ロボットのシミュレーションを行った。視覚センサは、２４個の網膜細胞を１次元に配列した視覚センサを２個用意し、各網膜細胞は7.5度の受容野を持ち、全体で180度の視野を持つものとする。また、各網膜細胞は、受容野中で対象物が占める割合を０から１の範囲の連続値で出力するものとする。また、動作の非対称性はここでは用いていない。図7.２７に700回試行後と10000回試行後の評価関数とロボットの経路を示す。これを見ると、学習の初期には、個々の視覚センサが小さな受容野しか持っていない影響で、評価関数が放射状に広がっていることがわかるが、学習が進むことによって滑らかな評価関数が形成され、ロボットの経路もほぼ最適な経路になっていることがわかった。

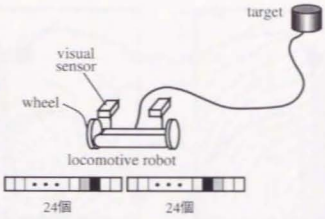


図７．２４　視覚センサを持った移動ロボット

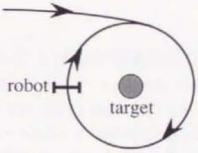


図７．２５　ループに陥った場合

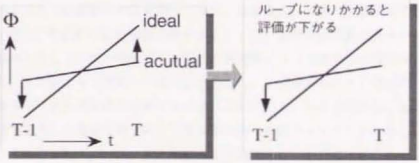


図７．２６　評価関数の時間軸方向傾き一定化学習の実現方法

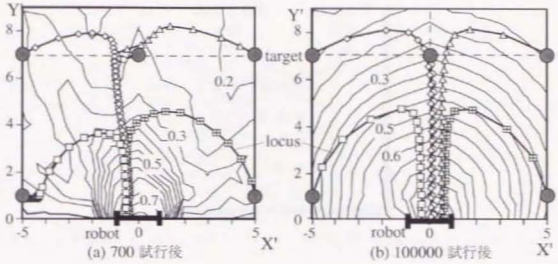


図７．２７　視覚センサ付きロボットが学習した評価関数と経路

７．６　中間層ニューロンにおける空間情報のコーディング

次に、このようにして学習したニューラルネットの中間層に意味のある情報がコーディングされているかどうかを調べる実験を行った。図７．２８のように、本来の強化学習を行うニューラルネットにテスト用の出力を加え、ニューラルネットの中の結合の重み値を微小な乱数で決定する。そして、まず、強化学習用の出力（評価用１個、動作用２個）に対し、前述の視覚センサ付き移動ロボット上で強化学習を行わせた。この時、テスト用出力は学習を行わない。そして、その後、今度はテスト用出力に図７．２９に示すような６つの点に関して、そこに物体を置いた時の視覚センサ信号をニューラルネットに入力し、その点の X'座標によって、X'座標が-5.0 の時に 0.1、5.0 の時に 0.9 という教師信号を与えて教師あり学習を行ってみた。学習後の、物体の位置に対する出力の分布を図７．２９(a) に示す。色が黒い部分が出力値が小さく、白い部分は出力値が大きいことを表す。これより、教師信号を与えていない場所でも、出力が X'座標によって滑らかに変化するという傾向が見られた。また、入力層から中間層への重み値を固定し、中間層からテスト用出力への重み値だけを学習させた場合もこれとほぼ同じ結果となった。これに対し、比較のために、強化学習を行わないで、強化学習を適用した場合と同じ結合の重み値を持つ初期ネットワークに対し、教師あり学習だけを行った。すると、図７．２９(b) のように、出力値のきれいな分布は見られず、出力値の尾根と谷がロボットの位置から放射状に広がるようになった。これは、視覚センサの各センサセルの受容野が放射状に広がるため、その受容野の中に物体がある状態で学習をすると、その受容野の中に物体がある場合全てに対して学習が影響を及ぼすためと考えることができる。このことから、強化学習を行うことで、中間層で知識の抽象化が行われ、X' をコーディングするニューロンが生まれていたと考えられる。このことは、知識の抽象化の能力を示しているのみならず、知識の共有の可能性を示しているものであり、非常に興味深い結果であると言える。

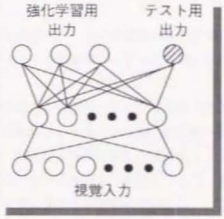


図７.２８　中間層ニューロンにおけるコーディングのテストのためのニューラルネットワーク

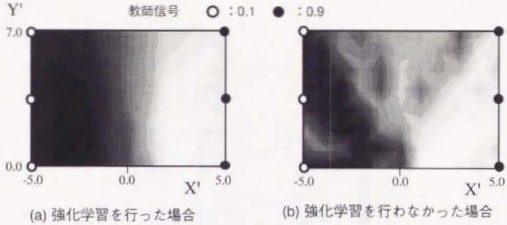


図７.２９　中間層ニューロンにおけるコーディング

次に、中間層ニューロンが実際に空間の情報をどのようにコーディングしているかを調べた。図７.３０のように、５層のニューラルネットを用意し、恒等写像ニューラルネット（第３章参照）のように、真ん中の中間層ニューロンを２個にしぼり、このニューラルネットに視覚情報を入力し、強化学習を行うことによってこの２つの中間層ニューロンの中で空間の情報がどのように分布しているかを調べた。

そして、何回か学習した後に目標物を図７.３１の格子上に置いた時の視覚信号をニューラルネットに入力した時の中間層ニューロンの値を図７.３２に示す。この図は、横軸に中間層ニューロン１の値、縦軸に中間層ニューロン２の値をとり、図７.３１の格子が中間層ニューロン空間でどのように表されているかをプロットしている。４つの図は、強化学習の進行によってこのコーディングがどのように変化しているかを表している。図７.３２中の１から５までの番号は、図７.３１中の特徴的な目標物の位置５つにつけた番号に相当している。この図より、

（１）学習が進むにつれ、空間の情報を表すために中間層ニューロン空間の広い範囲を使用するよ

うになってくる

（２）空間の情報の変化と共に中間層ニューロンの値が滑らかに変化するようになる

（３）ロボットにとって重要なところ（目標物がロボットに近い状態にいる時）を拡大して表現している

（４）目標物を置いた領域が長方形で、中間層ニューロン空間も正方形の領域であるにもかかわらず、４５度回転した形で中間層ニューロンが目標物の位置を表している。

ということがわかる。（２）は、評価値を時間に対して滑らかにしようとする学習によって、空間情報も滑らかに表現するようになっていると考えられる。これは、第４章での局所センサ信号の統合の際と同じである。ただ、第４章で見た空間情報の表現とは、（３）のように、目的達成のために重要なところが拡大されるという点が大きく違う。また、（４）は、（３）の影響で、ロボットの近くが重要であり、そこを拡大する力が働いた結果であると考えることができる。

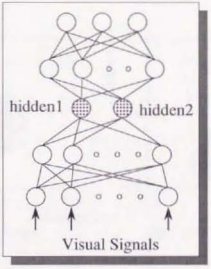


図７.３０　中間層ニューロンのコーディングを調べるためのニューラルネット

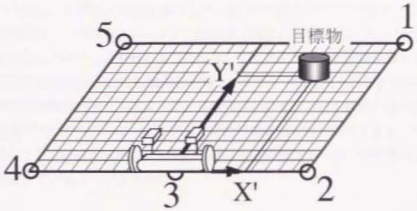


図７．３１　目標物を置いた格子

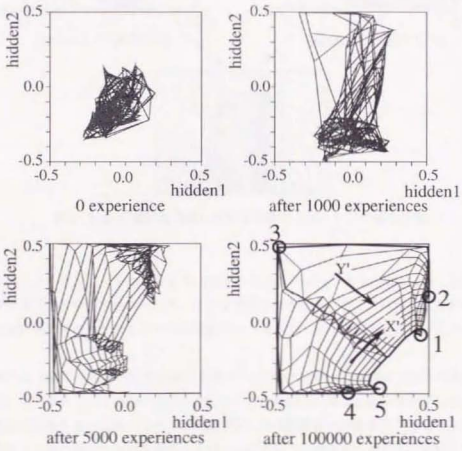


図７．３２　強化学習の試行回数による中間層ニューロンにおける空間情報コーディングの変化

さらに、これを逆に、目標物の位置を軸として、各中間層ニューロンの値の分布を示すと図７．３３に示す。この図より、中間層ニューロン１がX'=1のあたりで、ニューロン２がX'=-1のあたりで値の分布が密になっていることがわかる。これは、このロボットが目標物を捕らえるかどうかの境界がX'=-1およびX'=1にあり、この境界の内側ではロボットは直進すれば良いのに対し、外側では回転して目標物が境界の内側に入るようにしなければならないという動作の違いを必要とする。そのため、中間層の２つのニューロンが分担してその境界の内と外を検出するように学習されたと考えられる。また、２つの図を重ねたものが図７．３３の下の図であるが、中間層ニューロン１と２の分布がほぼ直交していることがわかる。

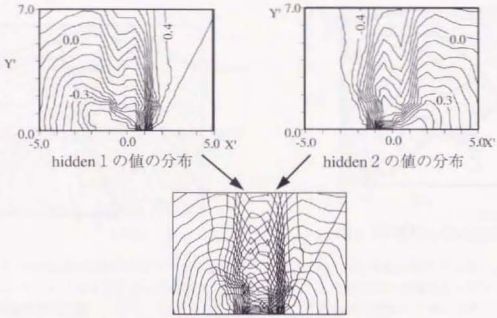


図７．３３　目標物の位置に対する中間層ニューロンの値の分布

次に、このニューラルネットの入力を７．３節のシミュレーションのように、ロボットと目標物との相対位置を入力して学習させたところ、うまく学習ができなかった。これは、入力が連続値になったことで逆に空間の切りわけを２つの中間層ニューロンでうまく表現できなかったためと考えられる。

次に、図７．１４のような非対称動作特性を持つロボットについて、同様に３層目の中間層が２個の５層のなニューラルネットを用いて学習を行った。この場合、学習があまり安定しないが、学習中で比較的良い経路を通った時（７８０００試行後）の評価関数と経路を図７．３４に示す。５層にした場合でも、３層の場合とほぼ同様に、ロボットは左前方に物体見えるように回転した後に左前方に前進していることがわかる。また、３層目の２つのニューロンにおける物体の位置のコーディングを　図７．３５に示す。この図でロボットの正面を表している線（３から延びている太めの薄い線）を見ると、中間層での表現では、物体がロボットの近くにある場合は物体が右側にある場合を、

物体がロボットから遠くにある時には左前方を拡大されていることがわかる。この表現法は、ロボットに非対称動作特性を持たせない場合と比べて変化しており、かつ、右に動くか左に動くかまたは前進するかの境界のあたりが拡大して表現されていることがわかる。このことから、中間層の表現が、単に視覚センサの受容野が放射状に広がっているという特性だけでなく、動作特性、または動作系列によって適応的に変化することわかった。また、ここで学習が不安定なのは、中間層での物体の位置の表現が非対称になるため、２つの中間層ニューロンの空間（正方形）の中にきれいに収まらないということが一つの要因として考えられる。また、５層の場合でも中間層のニューロン数を増やせば学習は安定した。

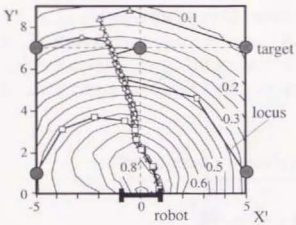


図７.３４　非対称動作特性のロボットが５層のニューラルネットで学習した場合の評価関数と経路

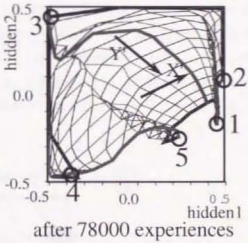


図７.３５　非対称動作特性のロボットが学習した場合の中間層ニューロンにおける物体の位置のコーディング

７．７　障害物回避のシミュレーション

最後に、ここまでのアルゴリズムで障害物回避ができるかどうかをシミュレーションで確認した。ここでの問題は、障害物の前後で状態が大きく変わってしまうが、その不連続さをどのようにニューラルネットで学習させるかにある。例えば、目標物および障害物との相対位置（合計４入力）を入力とし、６.２で述べた時間軸スムージング学習によるアルゴリズムで学習させると、障害物に当たると、前に進まなくなったり、一旦障害物の右側を通って目標物に到達することを学習すると、目標物が障害物より左側にあっても障害物の右側を通ろうとする等学習がうまく進まなかった。これの原因として、入力が障害物との相対距離で表しているため、障害物の前後で入力は連続となってしまう。しかし、ニューラルネットはシグモイド関数という入力に対して出力が滑らかにしか変化しない関数を用いて非線形変換を行うため、入力値の差が小さい時に、出力値の差を大きくする

ことは困難である。ところが、視覚センサを導入することにより、障害物の手前では、視覚センサに障害物が映るが、一旦障害物を越えてしまえばセンサに障害物が映らないということで、障害物の前と後ろで視覚センサの入力が大きく変化するため、ニューラルネットを用いても障害物の前後での空間の分離が可能になることが期待できる。

まず始めに、障害物回避に近い問題として、ロボットから見た目標物がある一定の領域に入った時にロボットの速度が 1/5 に落ちるという状況でシミュレーションを行った。ただし、センサ入力は、物体との前後および左右の相対位置である。その結果を図７.３６に示す。この図も、ロボットを中心にした座標で描いてあり、斜線で囲んだ円形の領域に目標物が入るとロボットの速度が 1/5 になるという設定になっている。外部から何も情報を与えていないにもかかわらず、学習の結果、斜線の領域のあたりで評価値が落ち込んでいることがわかる。また、ロボットの正面に目標物がある場合でも、わざわざロボットが回転して、斜線の領域を避けるような経路をとっていることがわかる。また、目標物が斜線部より右にある場合は、斜線部の右を通って、左にある場合は、左側を通って目標物を捕らえていることがわかる。このことから、空間の分離がうまくできているということができる。さらに、一旦、斜線部を避けたロボットは、目標物をロボットの真ん中で捕らえずに、端の方で捕らえていることもわかる。これは、無駄な動作を避けているということが言える。

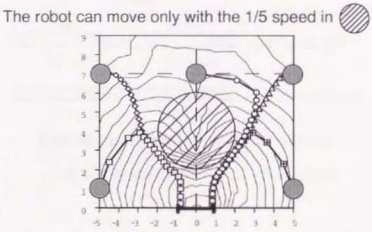


図７.３６　ロボットから見た目標物がある領域にある時にロボットの速度が遅くなるという設定の場合のシミュレーション

次に、実際の障害物を想定したシミュレーションを行った。シミュレーションの環境を図７.３７に示す。ここでは、少し不自然であるが、視覚センサを２種類用意し、片方は目標物のみ見え、もう片方は障害物のみ見えるという設定にした。そして、それぞれ左右１２個のセンサセルを１次元に並べ、合計４８個の信号をニューラルネットに入力した。また、目標物と障害物の位置は、各試行毎に乱数を用いて決定するが、目標物と障害物は最低１以上離すようにした。

結果を図７.３８、図７.３９に示す。両図とも、絶対座標で描いてあり、目標物や障害物は固定でロボットが動いていく。図７.３８は、障害物が視野に現れない場合に、目標物の位置を変化させ

た場合のロボットの経路を表している。ここでは、ロボットは最初に目標物の方に向きを変えて、その後ほぼ直進するという最適に近い経路を通っていることがわかる。一方、図７．３９では、ロボットのスタート地点の目の前に障害物がある場合の各目標物の位置に対するロボットの軌跡を表している。この場合には、図７．３８の場合とは明らかに違う経路をとり、障害物を避けて目標物にたどり着いていることがわかる。ただし、この学習は不安定であり、障害物の右側に目標物が見える時も障害物の左から回っていったりすることもある。また、右回りと左回りのちょうど境界あたりでは、ロボットは前に進めず立ちすくんでしまう。

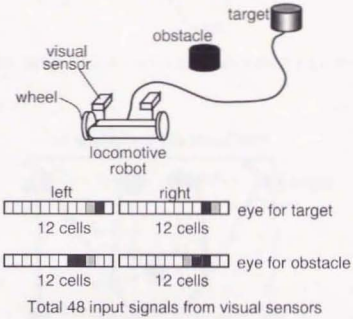


図７．３７　障害物回避のシミュレーションの環境

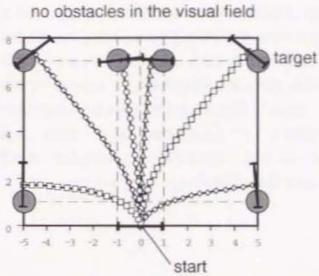


図７.３８　障害物が見えない場合の各目標物の位置に対するロボットの経路

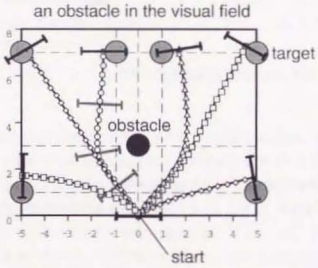
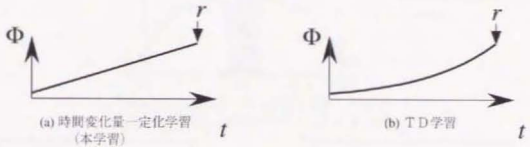


図７.３９　目の前に障害物が見える場合の各目標物の位置に対するロボットの経路

７．８　ＴＤ学習との比較

ここで、遅延強化学習の評価の学習として従来一般的に用いられてきたＴＤ学習（７.１節参照）との比較を行う。本章では遅延強化学習を目的達成のための所要時間の最短化という問題として捉え、時間軸スムージング学習を拡張した評価値の時間変化量一定化学習を提案した。この学習は、評価値の時間変化量を適応的に変化させるという点を除くと、結果的にＴＤ学習と非常に近い学習則となった。ＴＤ学習は、将来の重み付き報酬和を最大化するという定式化であったが、これを、

単一報酬の場合に適用すると、評価値（（7.2）式の$P(t)$）の時間変化は図7.40(b) のようになり、図7.40(a) の本学習の場合と比較すると、ＴＤ学習は直線の代わりに指数関数曲線で所要時間を表現しているという見方ができることがわかる。逆に、本学習を複数の報酬源がある場合に適用しようととすると、傾きが一定のため、近い方を選択する等の機能が働かないことになる。これに関しては、我々人間を振り返ると、何らかの方法で予め目標を一つに絞って行動を起こしているように考えられることから、目標を一つに絞る機構を加えることにより解決できると考えられる。また、本学習とＴＤ学習を別の視点からみると、本学習は式（7.1）に示したＴＤ学習の強化信号の重みづけ総和のディスカウントファクタγを１とし、常時微小な罰を与えていると捉えることも可能である。



(a) 時間変化量一定化学習
（本学習）

(b) ＴＤ学習

図7.40　ＴＤ学習との評価値の時間変化の比較

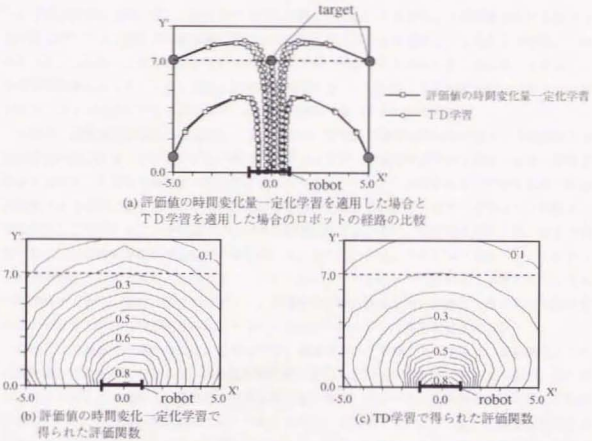
ＴＤ学習を指数関数による所要時間の評価と捉えると、所要時間が正しく学習できたとすれば、ＴＤ学習での評価値と本学習での評価値には１対１の関係付けをすることができる。また、微小線形近似を行うことにより、どちらの場合も評価値の時間に対する勾配が最急な方向に動作を学習させていくことにより経路の所要時間の最短化が実現できる。つまり、学習装置の能力が十分にあり、ローカルミニマにトラップされることなく十分な時間学習したとすれば、その解は等しくなるはずである。ただし、評価関数の形状は、ＴＤ学習ではゴールに近いほど急勾配に、本学習では一定勾配になるため、それが学習経過にどのように影響してくるか、また学習装置の能力が限定されていることがどのように影響するかが問題となる。そこで、前述の視覚センサ付き移動ロボットの問題のシミュレーションを行って両者の比較を行った。

まず、想定した環境は、前述のような図7.13のような環境で、移動ロボットは図7.24のような視覚センサを持ち、動作特性に非対称性は持たせなかった。そして、ＴＤ学習のほうも、条件をそろえるため、評価値の最大値を0.9、最小値を0.1となるように、式（7.10）の代わりに

$$\gamma = \left(\frac{0.1}{0.9}\right)^{\frac{1}{N_{max}}} \qquad (7.13)$$

からγを求めて学習させた。また、動作の学習は全く同様な方法を用いた。学習の結果を図7.41

に示す。獲得されたロボットの経路は、図７．４１(a) のように、本学習とＴＤ学習でほとんど差は見られない。一方、評価関数の形状は、図７．４１(b)、(c) を比較してわかるように、ＴＤ学習では、目標物がロボットの近くにある場合は勾配が急に、離れるに従って勾配がゆるやかになっている。図７．４２にロボットが目標物に到達するまでの所要時間の学習による変化を、それぞれの学習を初期値を変えたものを２つずつ示す。ここからも、両者にほとんど差がないことがわかる。



(a) 評価値の時間変化量一定化学習を適用した場合とＴＤ学習を適用した場合のロボットの経路の比較

(b) 評価値の時間変化一定化学習で得られた評価関数

(c) TD学習で得られた評価関数

図７．４１　評価値の時間変化量一定化学習とＴＤ学習による学習結果の比較

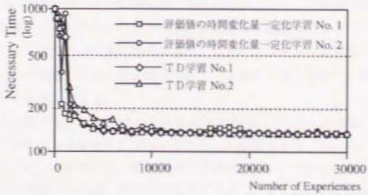


図７．４２　評価値の時間変化量一定化学習とＴＤ学習の学習曲線

７．９　考察および今後の課題

7.6節では、時間軸スムージング学習に基づく強化学習により、４章で述べた局所センサ信号の統合に近いことがニューラルネットの中間層において実現できることがわかった。ただし、強化学習を行った場合は、評価や動作の変化が大きい状態を拡大して表現しているという特徴が見られる。また、４章の方法では、時間軸スムージング学習に値域拡大学習を併用しなければ、出力の値域に幅をもたせることができない上、第４章の結論でも述べたように、値域拡大学習を適用するためには、時間軸スムージング学習による連続的な学習を中止しなければならず、現時点ではあまり好ましい方法とは言えない。また、第４章のような学習を行うということは、強化学習におけるセンサ信号から動作に至る過程での中間層に対しての学習と捉えることもできる。このような中で、４章のような学習は敢えてする必要がないのか、それとも、学習を高速化させるためにも、４章のような学習が必要なのかという点も検討を要する課題である。それから、ＴＤ学習を行った場合の中間層のコーディングがどのようになっているかも今後調べるべきところである。

さらに、遅延強化学習をさせる時に、探索のために非常に学習に時間がかかるという問題がある。本章で述べたシミュレーションでも、物体に到達するという簡単な学習を行うのに、０から学習させるとは言え、１０００回以上の試行が必要であった。これをさらに効率化させるためには、単純な乱数による探索ではなく、より効率的な探索が必要であると考える。また、そのような探索は、学習を通して実現することが可能ではないかと筆者は考える。また、効率的な探索とは、過去の履歴を踏まえた系統的な探索が必要であると考える。そのためには、リカレント型のニューラルネットの適用がポイントであり、ここでも、リカレントニューラルネットの効率的な学習アルゴリズムへの期待が大きい。また、探索だけでなく、評価および動作自体も過去の履歴を踏まえたものにすることが望ましい。この意味からもリカレントニューラルネットの適用が望まれる。

今後はより複雑な問題への対応が必要となる。複雑なタスクを解こうとすると、目的達成までの所要時間が増大することにより、評価値の時間変化がどんどん小さくなっていってしまう。これには、我々がタスクに慣れると達成時の快感を感じなくなり、さらにその前段階の状態に対して快感を得るといったことが有効に働いていると考えられる。１.２節で述べたように、大脳基底核でこれに近い発火状況をするニューロンが発見されている。また、これは、ＴＤ学習の際の誤差に相当するのではといった見解も出ているが、今後さらに検討していく必要がある。

また、複雑な問題に対応していくためには、障害物を避けるという問題が解けることも重要である。障害物の問題は、単にロボットの行動における問題だけでなく、問題解決に対するあらゆる障害の回避に適用できると考えられる。本論文でもある程度障害回避ができるようになったが、障害物の前で立ち止まってしまうという動作も見られた。ニューラルネットを使うと、状態と動作の関係を不連続にすることが困難であるためと考えられる。このようなことから、状態をある程度シンボリックに表すことも必要と考える。

最後に、これも１.２節で述べたが、このシステムを実用化しようとした時に、所望の行動をさせるためにどのように強化信号を設定すればよいかといった問題がある。例えば、部屋の中のそうじをロボットにさせようとした場合でも、強化信号の設定は難しい。それをしないとロボット自身の生存にかかわるような問題であれば、遺伝的アルゴリズム（ＧＡ）を用いて強化信号を生成させることができるかもしれないが、いずれにせよ難しい問題である。

７．１０　まとめ

本章では、時間軸スムージング学習を基本にして遅延強化学習の評価を学習する方法を提案した。この学習では、所要時間の違う複数の経路での評価を平等に行うためには、時間軸スムージング学習を単純に適用するだけでなく、評価の時間変化量を一定化させることが必要であることがわかった。また、評価の時間変化量一定化学習を行う際に、より将来の評価値の方が正しいという考え方に基づき、将来の評価値から過去の評価値を学習することが重要であることがわかった。

さらに、この学習を用いることによって、ロボットの動作特性を変化させた場合や状態によって動作特性を変化させた場合、予めそれに関する知識を与えることなく状況に適した動作を学習することを示した。そして、時には設計者自身が意図していないような動作を身につけることもわかった。そして、予め知識を与えた場合と比較して、学習に時間はかかるものの、本学習では、経路の最適化が行われることによって最終的には目的達成までの所要時間が短くなった。

また、入力を視覚センサ信号とした場合でもうまく学習ができ、中間層ニューロンに空間の情報を滑らかにコーディングしていることがわかった。この時、評価や動作が大きく変化する状態を中間層では拡大して表現していることがわかった。また、この時の入力層から中間層の部分を他の学習に用いることによって、強化学習で獲得された空間情報を利用できることがわかった。さらに、障害物を設けたシミュレーションでも、ある程度障害物を避けて通る様な経路を学習することができた。

第8章　結論

8.1 まとめ

本論文では、知識付与型知能システムからの質的変革を目指して、生物を手本とする自律学習という観点、つまり、外界とのフィードバックループを用いる等していかに自律的に学習するか、そして、いかに少ない情報から多くの機能を学習によって獲得するかといった観点からシステムがとるべき学習について考えてきた。ここでは、最初からシステムに与えれば良いと考えられる機能についても、敢えてできるだけ学習によって獲得するということにこだわった。これは、システムがあらゆる状態において学習ができるためには、「無からの学習」もできなくてはならないと考えたからである。また、そうすることによって、個々の機能に捕らわれることなく、学習がどうあるべきかの本質をつかみ、その結果、汎用的な学習則の実現に結びつくと考えたからである。

この意味から、本論文では、大きく２つの基礎となる学習則を提案した。１つは、「相関情報抽出学習」、もう一つは「時間軸スムージング学習」である。**相関情報抽出学習**は、異種情報源からの信号に共通に存在する情報（相関情報）が我々生物にとって重要な情報であるという仮説の下で、それを教師なしで抽出することを学習するというものであった。これは、第２章においてその概要を示すと共に、第３章において、値域拡大学習と組み合わせてニューラルネットを用いて学習させる方法を示した。さらに、抽出する情報がベクトルとなる場合、その各成分を直交化させる複数出力の直交化法を値域拡大学習の拡張として示した。そして、さらに、視覚の信号と運動の信号から空間認識能力が形成されるという例を取り上げて、シミュレーションを行った。しかし、この学習アルゴリズムは、特定のセンサ信号や運動の信号が与えられた場合にのみ有効となる特殊なアルゴリズムではなく、非常に汎用的な学習アルゴリズムである。Aitken らの実験で実際の子どもで確認されたように[Aitken 82]、この視覚の信号を、聴覚の信号と置き換えても基本的に同様な学習を行うことが可能である。また、実際に正常な視覚センサを有している場合でも、聴覚からも空間的な情報を得ている場合は多いと考えられるが、ここでは、情報源が３つの場合についてもその学習方法を示しており、これを用いることによって、このような場合もうまく説明することができる。さらに、第３章の終わりでも述べたように、運動の信号が入らない場合でも異種のセンサ信号間の相関情報をとることによって、文字や言葉の認識、概念の形成といった高度な機能へと結びついていく可能性を秘めていると考える。

本論文で、もう一つの基礎となる学習アルゴリズムである**時間軸スムージング学習**は、空間的な広がりを持つセンサ信号の時間的な位置づけを獲得することが、時空間上に存在するシステムとして状況を認識し、適切な動作をする上で重要であるという観点から生まれた汎用的な学習アルゴリズムである。本学習は、第４章で述べたように、空間情報が時間的に滑らかにしか変化しないとい

う仮説の下で、値域拡大学習と組み合わせることにより、局所的な受容野しか持たないセンサセルからの信号を統合し、空間情報をアナログ値として抽出することの学習に用いることができ、実際に３０個のセンサセルが１次元に配列された視覚センサを仮定し、その前を左右に単振動している物体の位置を教師なしで学習させることができた。そして、この出力が時間の変化と共に滑らかに変化するべきであるということから、頭部位置の補償や、前庭動眼反射のような眼の動きを学習させることができることを示した。また、さらに、その出力の変化が滑らかかつ時間変化が大きくなるようにという評価を与えることにより、物体を追跡するという眼の動きも学習できることを示した。一方、第７章では、遅延強化学習を目的達成までの所要時間の最適化問題と捉えることにより、時間軸スムージング学習が、センサ信号から所要時間の評価を学習することに用いることができることも示した。また、この遅延強化学習を階層型ニューラルネットを用いて行うことにより、中間層に空間の情報がきれいにコーディングされることがわかった。このことから、前述のようなセンサ信号の統合、認識という過程は、強化学習を進めるに当たって自ずと形成される機能と考えられるかも知れない。

もう一つ、値域拡大学習および複数出力の直交化学習というものを提案してきた。ここでは、出力の時間変化の中で、平均値からの偏差が大きいものに対し、強制的により偏差が大きくなるような学習を行わせるものであり、センサ信号を統合する際に、相関情報抽出学習または時間軸スムージング学習によって出力に拘束を設けつつ、出力の値域を確保することができた。しかし、これに関しては、ある一定間隔、または偏差が大きい時といった条件を満たした時に他の学習を中断して適用特別な学習を行わなければならない等、その学習アルゴリズムがスマートでないこと、さらに、自律学習という面から考えれば、ＢＰ法を適用した際に中間層ニューロンの値域が有効に使われるといった具合にならなければならないと考える。また、強化学習との統合によって解決されるものとも考えられる。

第６章では、認識や認識のための動作を目的達成の動作の一部とみなすことによって強化学習を適用することを提案し、簡単な文字認識などに適用できることを示した。ここでは、学習則自体は特にオリジナリティはないが、認識や認識のための動作を通常の目的達成のための動作として位置づけられるということを示した。これは、強化学習の能力をさらに引き出すものであり、自律学習の可能性をさらにおおきくしたものと考える。

また、本論文は、センサ信号の統合、センサ動作の学習、強化学習という切りわけをすることもできる。センサ信号の統合という観点から見ると、相関情報抽出学習を利用したものと時間軸スムージング学習を用いたものを提案したが、学習方法こそ違うものの、空間的に連続的な情報を連続値として獲得しようとしたという意味で非常に近いものであるということができる。従って、両者の融合は比較的簡単に実現できるのではないかと考える。ただ、入力の分布に対する出力の分布を見ると、相関情報抽出学習を使ったものは入力を分類する方向に、時間軸スムージング学習を使ったものは、入力の分布に従って出力を配分するというようにその性質は全く逆である。これによって両者をうまく使い分けることが可能ではないかと考えられる。

センサ動作の学習に関しては、時間軸スムージング学習によるセンサ信号の統合を効率的に獲得するものと強化学習に基づいた目的志向のものという２つの観点からの方法を提案した。前者においても、センサ動作を学習させる際にも、センサ信号を統合した出力が滑らかになるようにとかよ

り情報を得るようにという指針を元に強化学習を行っている。つまり、強化信号をセンサ信号の統合における汎用的な規準におくか、システムに設定された強化信号とするかという位置づけもできる。

強化学習については、与える情報量が少ないが故に得られる柔軟性、適応性の能力の大きさを再認識した。また、これによって、センサ信号の統合やセンサの動作まで学習できるということがわかったことは大きな収穫であった。強化学習を、センサ信号の統合、センサ動作の学習にも適用していくことは必要であるが、全て強化学習によって行うべきかどうかはわからない。

以上、自律学習システムの構築に向けて研究を進めてきたが、時間軸スムージング学習と相関情報抽出学習を提案できたことは、自律学習の研究の上で大きな進歩であったと考える。

8.2 今後の課題

本論文中の個々のテーマに関しての問題点は各章で述べてきたが、ここでは、自律学習の研究を進めるに当たって筆者が今後の課題と考えるところを述べる。

（1）強化学習による能動認識の学習と目的指向動作の学習の統合

前述のように、良いか悪いかという情報だけから認識および認識のための動作の学習ができることを確認した。しかし、ここでは、毎単位時間毎に良いか悪いかの評価が得られるという設定であった。そこで、評価は最後にしか得られないという設定で、おいしいかまずいかという最終的な評価から、その認識および認識のための動作、さらには、食べるため（目的指向）の動作の全てを学習する方法を探る。

（2）強化学習における効率的な試行法とその学習による獲得

強化学習では、学習にばく大な時間がかかることが一つの問題となっている。この理由の一つとして、従来は、単なる乱数によって探索を行ってきたことが挙げられる。しかし、我々生体は、系統的な探索、今までの知識を利用した探索およびそれを実現するための探索法自体を学習することができる。この機能の実現を目指したい。

（3）マルチエージェントシステムにおける強化学習およびコミュニケーションの発現

マルチエージェントシステムにおいて、個々のエージェントに強化学習を適用することにより、相互の協調の有効性と協調動作を学習によって獲得すると共に、協調のためのコミュニケーションの発現、さらにはコミュニケーションのための知識のシンボル化の必然性と発現の仕組みを探っていきたい。

（4）マルチエージェントシステムにおける強化学習およびコミュニケーションの発現

マルチエージェントシステムにおいて、個々のエージェントに強化学習を適用することにより、相互の協調の有効性と協調動作を学習によって獲得すると共に、協調のためのコミュニケーションの発現、さらにはコミュニケーションのための知識のシンボル化の必然性と発現の仕組みを探っていきたい。

（5）リカレントニューラルネットの学習アルゴリズム

知的システムの構成において、過去の必要な情報を記憶し、系統的な動作を行う必要がある。こ

のためには、リカレント構造のニューラルネットの適用が必要である。しかし、現状のリカレントネットの学習則は、理論的に求められた非常に複雑な学習アルゴリズムであり、実用に耐えがたい。これに関して、過去の学習則のアナロジーから大胆な学習アルゴリズムのシンプリファイが可能であるとの見通しを得ており、これの実現を目指したい。これは（２）を実現する上でも不可欠であり、自律学習システムの構築に向けて最重要課題と認識している。

（６）成長型ニューラルネットのアルゴリズム

ニューラルネットにおいて、その構造決定は非常に難しい問題である。従来、理論によって構造を決定する方法が試みられているが、この問題を根本的に解決し、適応的に構造を変化させるためには、ニューロン自身が必要に応じて成長し、ネットワークを形成していくという解が最も適していると考えられる。生体において、神経の成長を促進するＮＧＦという化学物質が発見されているという裏付けもある。これに関して、私は、ニューロンの成長と学習は共に環境に適応していく手段であり、実は同じ範疇で扱える、つまり、学習アルゴリズムの拡張で成長に関しても記述できると考えており、これによって環境に適した柔軟なニューラルネットの形成を目指したい。

以上の課題を克服していくことにより、自律学習システムを実用レベルまで引き上げ、また、生物の柔軟な学習能力の解明につながっていくことと考えている。

参考文献

{}内は主な参照元の章

[Aitken 82] {3} Aitken, S. and Bower T. G. R., "Intersensory substitution in the blind", *Jornal of Experimental Child Psychology*, Vol. 33, pp.309-323 (1982)

[Andersen 85] {5} Andersen, R. A., Essick, G. K., and Siegel, R. M., "Encoding spatial locationby posterior parietal neurons", *Science*, Vol. 230, pp.456-458 (1985)

[Anderson 89] {7} Anderson, C. W., "Learning to Control an Inverted Pendulum Using Neural Networks", *IEEE Control System Magazine*, Vol. 9 pp.31-37 (1989)

[Asada 94] {7} Asada, M., Uchibe, S., Noda, S. and Tawaratsumida, S., "Coodination Of Multiple Behaviors Acquired By Vision-Based Reinforcement Learning", *Proc. of IROS'94*, pp.917-924 (1994)

または、Asada, M., Noda, S. and Tawaratsumida, S., "Purposive Behavior Acquisition for a Robot by Vision-Based Reinforcement Learning", *Proc. of MLC-COLT Workshop on Robot Learning* (1994)

[浅田 95] {7} 浅田, 野田, 依積田, 細田, "視覚に基づく強化学習によるロボットの行動獲得", 日本ロボット学会誌, Vol. 13, No. 1, pp.68-74 (1995)

[浅野 95] {1} 浅野吉宏, "モデレーショニズムに基づくロボットアームの学習", 東京大学工学部卒業論文 (1995)

または、浅野吉宏, 岡部洋一, "モデレーショニズムに基づく反射弓学習のロボットアーム実験による検証", 日本神経回路学会第6回全国大会講演論文集, pp.167-168 (1995)

[麻生 87] {3} 麻生, 栗田, 大津, "正準相関分析および判別分析の非線形の定式化による解釈について", 行動計量学, Vo. 14, pp. 1-9 (1987)

[Asoh 94] {3} Asoh, H. and Takechi, O., "An Apploximation of Nonlinear Canonical Correlation Analysis by Multilayer Perceptrons", *Proc. of ICANN'94*, pp.713-716 (1994)

または、麻生, 武市, "非線形正準相関分析の近似的実現について－実験－", 日本神経回路学会第5回全国大会講演論文集, pp.259-259 (1994)

[東 69] {1} 東洋, 大山正, "学習と思考", 心理学入門講座, Vol.3, 大日本図書 (1969)

[Baldwin 1896] {1} Baldwin, J. M., "A new factor in evolution", *American Naturalist*, Vol. 30, pp. 441-451 (1896)

[Ballard 87] {3} Ballard, D. H., "Modular Learning in Neural Networks", *Proc. of AAAI'87*, Vol. 1, pp.279-284 (1987)

[Barto 83] {7} Barto、A. G., Sutton, R. S. and Anderson C. W., "Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems", *IEEE Trans. SMC-13*, pp.835-846 (1983)

[Barto 95-1] {1} Barto, A. G., Bradtke, S. J. and Singh S. P., "Learning to act using real-time dynamic programing", *Artificial Intelligence*, Vol. 72, pp.81-138 (1995)

[Barto 95-2] {1} Barto A. G., "Adaptive critics and the basal ganglia", *In Houk, J. C. et al. eds., Models of Information Processing in the Basal Ganglia*, pp.215-232, MIT Press, Cambridge, MA (1995)

[Becker 89] {3} Becker, S. and Hinton, G. E., "Spatial coherence as an internal teacher for a neural network, *Technical Report CRG-TR-89-7*, University of Toronto (1989)

または、 Becker, S. and Hinton, G. E., "A self-organizing network that discovers surfaces in random-dot stereograms, *Nature*, Vol. 355, pp.161-163 (1992)

または、 Hinton, G. E. and Becker, S., "An unsupervised learning procedure that discovers surfaces in random-dot stereograms", *Proc. of IJCNN Hillsdale*, NJ. Erlbaum, Vol. 1, pp.218-222 (1990)

または、 Zemel, R. S. and Hinton, G. E., "Discovering Viewpoint-Invariant Relationships That Characterize Objects", *In Advances In Neural Information Processing Systems*, Vol. 3, pp. 299-305, Morgan Kaufmann Publishers (1991)

[Bellman 57] {1} Bellman, R. E., "Dynamic Programming", Princeton University Press, NJ (1957)

[Blackmore 76] {1} Blackmore, C., Van Sluyters, R. C. and Movshon, J. A., "Synaptic competition in the kitten's visual cortex", *Cold Spring Harbor Symp. Quant. Biol.*, Vol.40, pp.601-609 (1976)

または、 津本忠治、"生後環境のよる視覚中枢の変化"、脳と発達（第７章）、朝倉書店 (1986)

[Charms 76] {1} De Charms, R. C., "Enhancing Motivation - change in the Classroom"

（訳本）佐伯（訳）, "やる気を育てる教室"、金子書房 (1980)

[銅谷 86] {1} 銅谷賢治、"運動パターンの自己組織化"、第 16 回 SICE 学術講演会予講集、pp.961-964 (1986)

[銅谷 95] {1} 銅谷賢治、"ＴＤ学習則の連続時間モデルへの拡張"、日本神経回路学会第６回全国大会講演論文集、pp.22-23 (1995)

[銅谷 96] {1} 銅谷賢治、"強化学習"、日本神経回路学会第７回全国大会講演論文集、pp.158-162 (1996)

[Elman 90] {1} Elman, J. L., "Finding Structure in Time", *Cognitive Science*, Vol. 14, pp. 179-211 (1990)

[Gomi 92] {5} Gomi, H. and Kawato, M., *Biological Cybernetics*, Vol. 68, pp.105- (1992)

[Gonshor 76] {5} Gonshor, A and Melvill-Jones, G, *J. of Physiol.*, Vol. 174, pp.417-488 (1976)

[畠中 92] {1} 畠中寛、"神経成長因子ものがたり"、羊土社 (1992)

[Hebb 75] {1} Hebb, D. O., "A Textbook of Psychology", W. B. Saunders Company

（訳本）白井ら共訳、"行動学入門　生物科学としての心理学"、紀伊國屋書店 (1975)

[Held 63] {3} Held, R. and Hein, A. "Movement-produced stimulation in the development of visually guided components" J.Comp. Physiol. Psychol., 56 : 872-876 (1963)

[Holland 87] {1} Holland, J. H. and Reightman, J. S., "Cognitive System Based on Adaptive Algorithms", Pattern-Directed Inference Systems, Waterman, D. A. and Hayes-Roth, F. ed., Academic Press (1987)

[Houk 95] {1} J. C. Houk, J. L. Adams, and A. G. Barto, "A Model of How the Basal Ganglia Generate and Use Neural Signals That Predict Reinforcement", *In Houk, J. C. et al. eds., Models of Information Processing in the Basal Ganglia*, pp.249-270, MIT Press, Cambridge, MA (1995)

[Hubel 72] {1} Hubel, D. H. and Wiesel, T. N., "The period of susceptibility to the physiological effects of unilateral eye closure in kittens", *Journal of Physiol.*, Vol. 206, pp.419-436 (1972)

[Hyvarinen 81] {3} Hyvā inen, J., Hyvā inen, L. and Linnankoski, I., " Modification of parietal association cortex and functional blindness after binocular deprivation in young monkeys", *Exp. Brain Res.*, Vol.42, pp.1-8 (1981)

[乾 93] {6} 乾敏郎、脳と視覚－人間からロボットまで－、サイエンス社 (1993)

[入江 90] {3} 入江文平、川人光男、"多層パーセプトロンのによる内部表現の獲得"、電子情報通信学会論文誌、J73-D-II, Vol. 8, pp.1173-1178 (1990)

[Ito 70] {5} Ito, M., "Neurophysiological aspects of the cerebellar motor control system", *International Journal of Neurology*, Vol. 7, pp.162-176 (1970)

[Ito 82] {5} Ito, M., Sakurai, M. and Tongroach, P., "Climbing fibre induced depression of both mossy fibre responsiveness and glutamate sensitivity of cerebellar Purkinje cells", *J. of Physiology London*, Vol. 324, pp.113-134 (1982)

[片山 90] {3} 片山正純、川人光男、"視覚、体性感覚と運動指令を統合する神経回路モデル"、日本ロボット学会誌、Vol. 8, No. 6, pp.757-765 (1990).

または、 片山正純、川人光男、"視覚情報と体性感覚情報を用いた対象物の３次元内部表現の学習"、電子情報通信学会春期全国大会講演予稿集、D-24 (1989)

[Kawato 87] {5} Kawato, M., Furukawa, K. and Suzuki, R., "A hierarchical neural-network model for control and learning of voluntary movement", *Biological Cybernetics*, Vol. 57, pp.169-185 (1987)

または 川人光男、宇野洋二、鈴木良次、"随意運動制御における適応と学習 II"、日本ロボット学会誌、Vol. 6, No. 3, pp.50-58 (1988)

[川人 94] {5} 川人光男、五味裕章、"脳の中の運動モデル"、科学、Vol. 64, No. 11, pp.720-729 (1994)

または、 川人光男、"脳の計算理論"、産業図書 (1996) 特に pp.203-210

[Kaufman 69] {6} Kaufman L. and Richards, W., "Spontanious fixation tendencies for visual forms", *Perception and Psychophysics*, Vol. 5, No. 2, pp.85-88 (1969)

[甲原 94] (1) 甲原隆矢、"モデレーションニズムに基づく神経回路の学習法"、東京大学大学院博士論文 (1994)

[Luo 89] {3} Luo, R. and Kay, M. "Multisensor Integration and Fusion in Intelligent System", *IEEE Trans. on SMC*, Vol. 19, No.5, pp.901-931 (1989)

[丸野 89] {1} 丸野俊一、"知能はいかに作られるか"、ブレーン出版 (1989)

[モンタルチニ 79] {1} R. レビーモンタルチニ、P. カリサーノ、(天野武彦 訳)、"神経成長因子"、サイエンス、Vol. 9, No. 8, pp.90-101 (1979)

0[中野 84] 中野馨、鋼谷賢治、"運動パターンを自己形成するシステム"、計測自動制御学会第２３回学術講演会、S 1-3 (1984)

[Okabe 88] {1} Okabe, Y., "Moderationism: Feedback learning of neural networks", *Proc. of IECON*, pp.1028-1033 (1988)

または、岡部洋一、"フィードバック学習"、数理科学, 338, pp.26-30, 1990

[奥野 71] {3} 奥野忠一 他、"多変量解析法"、日科技連出版 (1971)

[Onat 95] {1} Onat, A., Kita, H. and Nishikawa, Y., "Reinforcement Learning of Dynamic behavior by using Recurrent Neural Networks", *Proc. of WCNN'95*, Vol. 2, pp.II-342-345 (1995)

[Poggio 85] {4} Poggio T., Torre V. and Koch C., "Computational vision and regularization theory", Nature, 317, 6035, pp.314-319 (1985)

または、 横矢、坂上、"画像理解と最適化原理"、電子情報通信学会学会誌、Vol. 74, No. 4, pp.326-334 (1991)

[Portmann 44] {1} Portmann, A., Biologische Fragmente zu einer Lehre vom Menschen", Benno Schwabe (1944)
(訳本) 高木正孝訳、"人間はどこまで動物か"、岩波新書, G121 (1961)

[Rumelhart 86] {1} Rumelhart, D.E., Hinton, G.E. and Williams, R.J., "Learning representations by back-propagating errors", Nature, 323, 9, pp.533-536, 1986 又は

[Rumelhart 88] {3} Rumelhart, D. E., Hinton, G. E. and Williams, R. J., "Learning Internal Representations by Error Propagation", *Parrallel Distributed Processing*, Vol. 1, pp.318-362 (1988)

[阪口 91] {6} 阪口 豊, 中野 馨, "能動的認識の数理モデル", 第6回生体・生理工学シンポジウム論文集, pp.373-376 (1991)

[阪口 93] {6} 阪口 豊, "触知覚における感覚統合と能動認識", 電子情報通信学会学会誌, Vol. 76, No. 11, pp.1222-1227 (1993)

[酒田 76] {3} 酒田英夫, "頭頂連合野の機能", 現代の神経科学, Vol.3, 高次脳機能と中枢プログラミング（伊藤正男, 島津浩編）, 産業図書, pp.145-169 (1976)

[酒田 82] {3} 酒田英夫, "感覚の統合－連合的意識",脳と認識（伊藤正男編）, 平凡社, pp.167-194 (1982)

[Samuel 59] {1} Samuel, A. L., Some Studies in Machine Learning Using Game of Checkers, *IBM Journal on Research and Development*, Vol. 3, pp. 210-229 (1959)

[Schultz 93] {1} Schultz, W., Apicella, P. and Ljungberg T., "Responses of Monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task", *Journal of Neuroscience*, Vol. 13, pp.900-913 (1993)

または、 Schultz, W., Romo, R., Ljungberg, T., J Mirenowicz, Hollerman, J. R. and Dickinson A., "Reward-related Signals Carried by Dopamine Neurons", In Houk, J. C. et al. eds., *Models of Information Processing in the Basal Ganglia*, pp.233-248, MIT Press, Cambridge, MA (1995)

[柴田 89] {1} 柴田克成, "バックプロパゲーション法に基づくロボットの学習機能に関する研究", 東京大学大学院工学系研究科機械工学専攻修士論文 (1989)

または、柴田克成, 稲葉雅幸, 井上博允, "ニューラルネットによるロボットの運動学習", 第6回日本ロボット学会学術講演会予稿集, pp.141-142 (1988)

[Skinner 61] {1} Skinner B.F.,"Cumulative Record", *Appleton-Century-Crofts* (1961)

[Sutton 88] {7} Sutton, R. S., "Learning to Predict by the Methods of Tmporal Differences", *Machine Learning*, Vol. 3, pp.9-44 (1988)

[田中] {3} 田中 豊 他, "パソコン統計解析ハンドブック2 多変量解析編", 共立出版

[田中 90] {1} 田中みどり, "発達とは何か", 人間発達の心理学, 藤永保監修, サイエンス社 (1990)

[Tesauro 92] {1} Tesauro, G., "Practical issues in temporal difference learning", *Machine Learning*, Vol. 8, pp.257-277 (1992)

[津本 86] {1} 津本 忠治, "脳と発達", 朝倉書店 (1986)

[畝見 95] {1} 畝見達夫, "強化学習法とロボットへの応用", 日本ロボット学会誌, Vol.13, No.1, pp.51-56 (1995)

[Watkins 92] {1} Watkins, C. J. C. H. and Dayan, P., "Q-learning", *Machine Learning*, Vol. 8, pp.279-292 (1992)

[Werbos 90] (1) Werbos, P. J., "Overview of Designs and Capabilities", *Neural Networks for Control*, MIT Press, pp. 59-96 (1990)

[Werbos 95] (1) Werbos, P. J., "Optimal Neurocontrol : Practical Benefits, New Results and Biological Evidence", *Proc. of WCNN'95*, Vol. 2, pp. II-318-325 (1990)

[Whitehead 91] {6} Whitehead, S. D. and Ballard D. H., "Learning to Perceive and Act by Trial and Error", *Machine Learning*, Vol. 7, pp.45-83 (1991)

[Widrow 85] (1) Widrow, B. and Stearns, S. D., *Adaptive Signal Processing*, Englewood Cliffs, NJ:Prentice-Hall (1985)

または、Nguyen, D. and Widrow, B., "Neural Controls", *Proc. of IJCNN'92*, Vol. 1, pp.10-14 (1992)

[Williams 88] {7} Williams, R. J., "Toward a theory of reinforcement-learning connectionist systems", *Technical Report NU-CCS-88-3*, College of Computer Science, Northeastern University, Boston, MA. (1988).

または、Williams, R. J., "Simple Statistical Gradient-Following Algorithm for Connectionist Reinforcement Learning", *Machine Learning*, Vol. 8, pp.229-256 (1992).

[山内 95] {3} 山内康一郎, 石川直宏, "異種センサ情報の統合による概念の教師なし学習", 日本神経回路学会第6回全国大会講演論文集, pp.319-320 (1995)

[山内 95] {3} 山内康一郎, 太田幹也, 石川直宏, "異種センサ情報の統合によるクラスタの教師なし学習", 日本神経回路学会第7回全国大会講演論文集, pp.174-175 (1996)

[Yamamura 95] {1} 山村雅幸, 宮崎和光, 小林重信, "エージェントの学習", 人工知能学会誌, Vol.10, No.5, pp.23-29 (1995)

[山崎 90] {3} 山崎弘郎ら, "特集：センサ情報の統合", 日本ロボット学会学会誌, Vol. 8, pp. 721-774 (1990)

[山崎 92] {6} 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)

[横矢 91] {4} 横矢直和, 坂上勝彦, "画像理解と最適化原理", 電子情報通信学会誌, Vol. 74, No. 4, pp.326-334 (1991)

[依田 90] {1} 依田晴夫, 宮武孝文, 松島整："本能に基づいて運動系列を学習する運動モデル", 電子情報通信学会論文誌 J73-D-II, No. 7, pp.1027-1034 (1990)

[Zipser 86] {3} Zipser, D., "Programing Neuralnets to do Spatial Computations", *ICS Report 8608*, Inst. for Cognitive Science, Univ. of California, San Diego (1986)

発表文献等

＜投稿論文（筆頭）＞

[1] 柴田克成，岡部洋一：相関情報抽出ネットと空間認識能力の教師なし学習，
日本神経回路学会論文誌, Vol.3, No.2, pp.11-16, 1996.5

[2] 柴田克成，岡部洋一：時間軸スムージング学習による局所センサ信号の統合と空間情報の抽出学習，
日本神経回路学会論文誌, Vol.3, No.3, pp.98-105, 1996.9

[3] 柴田克成，岡部洋一：強化学習による能動認識能力の学習，
日本神経回路学会論文誌, Vol.3, No.4, 1996 掲載予定

[4] 柴田克成，岡部洋一：遅延強化学習における評価関数の時間軸スムージング学習，
日本神経回路学会論文誌（査読済み、修正中）

＜国際会議（筆頭）＞

[5] K. Shibata and Y. Okabe : Some Learning Models of Visual System based on Local Sensory Signals Integration Learning,
Proc. of ICNN '95 Perth, IV, pp.1986-1990, 1995.11

[6] K. Shibata, T. Nishino and Y. Okabe : Active Perception based on Reinforcement Learning,
Proc. of WCNN '95 Washington, II, pp. 170-173, 1995.7

[7] K. Shibata and Y. Okabe : Unsupervised Learning Method to Extract Object Locations from Local Visual Signals,
Proc. of ICNN '94 Orlando, Vol. 3, pp.1556-1559, 1994.7

[8] K. Shibata and Y. Okabe : A Robot that Learns an Evaluation Function for Acquiring of Appropriate Motions,
Proc. of WCNN '94 San Diego, Vol.2, pp. II-29 - II-34, 1994. 6

[9] Katsunari Shibata : A Neural-Network to get Correlated Information among multiple Inputs,
Proc. of IJCNN'93 Nagoya, Vol.3, pp. 2532-2535, 1993.10

[10] Katsunari Shibata : Spatial Recognition Model by Extracting Coorelated Information between Vision and Motion Information using Neural-Network,
Proc. of IJCNN'93 Nagoya, Vol.3, pp.2536-2539, 1993.10

＜研究会・全国大会（筆頭）＞

[11] 柴田克成，岡部洋一：視覚センサ信号を入力とした遅延強化学習，
日本神経回路学会第７回全国大会講演論文集，pp.144-145, 1996.9

[12] 柴田克成，岡部洋一：時間軸スムージング学習と局所センサ信号の統合，
日本神経回路学会第７回全国大会講演論文集，pp.178-179, 1996.9

[13] 柴田克成，岡部洋一：相関情報抽出ネットによるステレオ画像上物体の奥行き情報抽出の教師なし学習，
日本神経回路学会第６回全国大会講演論文集，pp. 231-232, 1995.10

[14] 柴田克成，西野哲男，岡部洋一：強化学習に基づく能動認識，
日本神経回路学会第５回全国大会講演論文集，pp. 82-83, 1994.11

[15] 柴田克成：ニューラルネットによる目的達成動作の能動的学習，
1991年電子情報通信学会秋季大会講演論文集，pp.6.304-305, 1991.9

[16] 柴田克成，他８名：高速学習型ニューロＷＳＩのシステム設計，
電子情報通信学会技術報告，CPSY90-71, ICD90-127, 1990.10

[17] 柴田克成，稲葉雅幸，井上博允：ニューラルネットによるロボットの運動学習，
第６回日本ロボット学会学術講演会予稿集，pp.141-142，1988.10

＜その他の口頭発表等（筆頭）＞

[18] 柴田克成：局所センサ情報統合化学習に基づく視覚システムの学習モデル，
科研費重点領域「脳高次処理」冬のワークショップ，1996. 1

[19] 柴田克成：遅延強化学習に基づくロボットの運動学習と局所センサ情報の統合化学習，
科研費重点領域「脳高次処理」冬のワークショップ，1995. 1

＜その他（筆頭）＞

[20] 柴田克成、岡部洋一：強化学習，
脳科学ハンドブック，朝倉書店，執筆中

[21] 柴田克成：バックプロパゲーション法に基づくロボットの学習機能に関する研究，
東京大学大学院工学系研究科機械工学専攻修士論文，1989.3

[22] 柴田克成：六軸力計および六分力検出テーブルを用いた自動加工システムの研究，
東京大学工学部機械工学科卒業論文，1987.3

＜投稿論文（共著）＞

[23] 安永守利，柴田克成，浅井光男，山田稔：ニューラルネットワーク集積回路の自律的な欠陥救済能力，
電子情報通信学会論文誌 D-I, Vol.J75-D-I, No.11, pp1099-1108, 1992.11

＜国際会議（共著）＞

[24] Yuji Sato et. al：Development of a High-Performance, General Purpose Neuro-Computer Composed of 512 Digital Neurons,
Proc. of IJCNN'93 Nagoya, Vol.2, pp.1967-1970, 1993.10

[25] M. Yasunaga et. al ：A Self-Learning Neural Network Composed of 1152 Digital Neurons in Wafer-Scale LSIs,
Proc. of IJCNN'91 Singapore, Vol.3, pp.1844-1849, 1991.11

＜研究会・全国大会（共著）＞

[26] 西野哲男，柴田克成，岡部洋一：強化学習による視点移動の学習
日本神経回路学会第５回全国大会講演論文集, pp. 84-85, 1994.11

[27] 安永守利，柴田克成，浅井光男，山田稔：ウェハースケールニューラルネット集積回路の自律的な欠陥救済能力，
1993 年電子情報通信学会春期大会講演論文集, SD-8-2, 1993.3

[28] 坂口隆宏，他７名：高速学習型デジタルニューロＷＳＩの設計評価，
1992 年電子情報通信学会春期大会講演論文集, pp.5-195, 1992.5

[29] 浅井光男，他５名：高速学習型ニューロＷＳＩ，
電子情報通信学会技術報告, NC90-12, 1990.5

＜その他の口頭発表等（共著）＞

[30] Y. Okabe et. al : Formation of Reflex Arc by Feedback Learning,
RWC conference '94, RWC Technical Report, TR-94001, pp.133-134, 1994. 6

[31] Y. Okabe et. al : Application of Moderationism to a Non-Linear Environment,
RWC conference '95, RWC Technical Report, 1995. 6

筆者の略歴および本研究の経緯

筆者は、平成元年３月東京大学大学院工学系研究科機械工学専攻修士課程修了後、同年４月（株）日立製作所中央研究所に就職した。その後、平成４年１０月同社を退社し、同年同月東京大学大学院工学系研究科先端学際工学専攻博士課程へ入学、平成５年９月同大学院中途退学、同年１０月東京大学先端科学技術研究センター協力研究員を経て、同年同月東京大学先端科学技術研究センター助手となり、現在に至る。現在３１才である。

修士課程では、井上・稲葉研究室に所属し、「バックプロパゲーション法に基づくロボットの学習機能の研究」と題して、当時発表されたばかりのバックプロパゲーション法の解析、学習誤差に従って中間層ニューロン数を増大させていく方法、制御への応用として、フィードバック＋フィードフォワード制御の学習による獲得、さらには、本論文の１つの柱である「相関情報抽出学習」に関して、その基本的アイデアを示した。

（株）日立製作所では、ニューラルネットのハードウエア化の研究を主に行い、ニューロＷＳＩ（Wafer Scale IC）の設計・製作に携わり、特に学習回路の設計を担当した。その後、銅谷ら、依田らの常に強化信号が与えられるタイプの強化学習に触れ[銅谷 86][依田 90]、遅れた強化信号に対応できるようにするために「時間軸スムージング学習」によって遅延強化学習を行う方法を考えついた。この時点では、筆者はBartoらの研究を知らなかった。また、この時点では、当初、視覚センサ信号を入力としていたがうまくいかず、物体との相対位置の情報をロボットに与えてシミュレーションを行っていた。

その後、これらの研究をさらに推進するため、東京大学大学院先端学際工学専攻の博士課程に入学し、東京大学先端科学技術研究センターの岡部・廣瀬研究室に所属し、現在まで身分は変わったものの４年以上にわたってここで研究を続けている。本研究室では、遅延強化学習の研究をさらに進め、非対称動作特性の場合のシミュレーションを行い、その有効性を確かめると共に、うまくいかなかった問題を解析することによって、評価値の時間変化量が一定であることが必要であることがわかり、評価値の時間変化量一定化学習を提案した。その後、この時間軸スムージング学習が局所センサ情報の統合に有効であることがわかり、そのための研究を行った。また、相関情報抽出学習について、視覚センサを２つ用意することによって物体の大きさが変化してもその位置情報を抽出できることを示した。さらに、抽出する情報が２次元以上の場合について、修士の際に提案したアイデアについてより深く検討を行った。その後、リカレントニューラルネットの学習が必要であると感じ、連続時間モデルのニューラルネットの学習アルゴリズムの研究に入った。ここでは、離散時間モデルとのアナロジーから大胆なシンプリファイを試みたが、うまくいきそうでいかなかった。結局、時間を掛けた割には本論文に掲載できるような成果は得られず、今後の大きな課題として残った。次に、当時４年生であった西野君（現在修士２年）と共に、かねてから考えていた強化学習による能動認識の研究を行った。そして、センサからの信号からある程度簡単なパターンを認識させると共にセンサの動作も学習できることができることがわかった。その後、時間軸スムージング学習による局所センサ信号の統合学習の拡張として、前庭動眼反射等の機能の説明を試みた。そして、最近、再び強化学習の問題に取り組み、視覚センサ信号を直接入力としても学習ができ、さらに中間層にうまく空間の情報がコーディングされることを確認し、現在に至っている。

謝辞

前述のように、本研究は、主に、東京大学先端科学技術研究センター岡部・廣瀬研究室にて行った。本研究をこのように推進し、論文としてまとめることができたのは、岡部洋一教授のおかげによるところが非常に大きい。筆者は、（株）日立製作所時代に通産省のＲＷＣプロジェクトに対する岡部教授の提案書を拝見し、この人なら私の研究を理解し、後押ししてくれると考えたのが最初である。その後、教授を訪ね、それ以来４年間すっかりお世話になってしまった。岡部教授は、自主性を重んじられ、また、どんな相手でも尊重して下さった。そのお陰で、筆者は自由に研究することができた。また、先生は、人をエンカレッジすることがうまく、私が落ち込んでいる時には勇気づけて下さり、私が自信なくしゃべることに対しても、その良い点を探して評価して下さった。また、そのお人柄から、明るく、楽しく研究生活を送ることができた。ここに、深く感謝の意を表する。

また、本研究の一番の発端は、私の大学院修士時代である。修士の１年の終わり頃からニューラルネットの研究を始めた。これに対し、様々な議論をし、応援して下さった当時の指導教官である井上博允教授、および稲葉雅幸助教授を始めとする井上・稲葉研究室の方々にも深く感謝する。

（株）日立製作所時代には、社内の発表会の内容として、本研究のきっかけとなるものを認めていただき、研究の時間を頂いた。それに対して暖かい眼で見ていただいた山田稔主任研究員（当時）、益田昇さん、安永守利（現筑波大）さん、浅井光男さんを始めとする回りの方々に感謝する。

それから、ＲＷＣでの議論も本研究を進めるに当たり大いに役立った。東大の吉沢先生、筑波大の平井先生、東邦大の古谷先生、日立時代にも一緒に仕事をし、公私ともにお世話になった佐藤裕二さん、岡部研のＯＢでもあり、強化学習等教えていただいた山川宏さんはじめ皆さんに感謝する。

また、学会その他でご指導、ご意見、議論を頂いた、ＡＴＲの銅谷さん、農工大の大森先生、電総研の仁木さん、麻生さん、野田さん、幸島さん、電通大の阪口先生、三菱電機の山田さん、東大の前田さん、名工大の山内さんはじめとする皆さんに感謝する。

岡部。廣瀬研究室においては、廣瀬助教授には、研究に関して鋭いご指摘、議論を頂き、大変勉強になりました。北川助手、宮尾技官には、特に技術的な面で御指導を頂きました。秘書の菅沼、白石さん、研究室関連のお仕事でお世話になり、また、研究の合間に楽しいお話を聞かせて頂きました。協力研究員である田宮さんは、研究室の母親的存在で、身の回りの様々なことについてお世話になりっぱなしでした。ここに感謝致します。

その他、既にＯＢとなりました、甲原さんには、研究や計算機管理およびその他いろいろとお世話になりました。それから、今はチリに帰られた Pable さんには、英語の論文チェックをしていただいたり、様々なことを教えていただきました。筆者と先端学際工学専攻の同期であった東芝からいらしていた林さんには、睡眠時間を削って仕事と研究を両立される姿に大変励まされました。現在博士３年の松浦君とは私が研究室に来て以来のおつきあいですが、打ち合わせでの議論および彼のプレゼンテーションには非常に勉強になりました。同じく博士３年の市瀬君とは、よく学会に一緒に行ったり、研究の議論をしたり、その他の話し相手としてと世話になりました。筆者と同時期に研究室に来た時松君には、そのパワフルさが非常に印象的でした。現在博士２年の五月女君には、計算機関連その他非常に幅広い知識を勉強させてもらいました。現在、私が多少計算機の管理の仕事ができるようになったのもほとんど彼のおかげによると言っても過言ではない。同じく博士２年の掛谷君には、研究に関して鋭い突っ込みをしてもらい、漫然とした研究生活に活を入れてもらいました。また、幅広い分野での議論が大変勉強になりました。また、一緒に研究をした西野君とは、自分の考えた研究を一緒にこなした同士であり、また、研究スタイルも筆者と近く、楽しく研究を

共にすることができました。また、修士を卒業した中尾君、成沢君には、研究の議論をしたり、また、研究の息抜き等共にしました。タイの留学生のワサンさんにも、研究、その他でお世話になりました。博士課程１年の小高君にも２年間同じ部屋で楽しい研究生活を送ることができました。また、お国の話しなどを聞かせていただき、英語の勉強もさせていただいた呉先生、Moore 先生、章先生にもお世話になりました。その他、日本人より日本人らしい金さん、あのウィットの富んだしゃべり方にあこがれた鈴木さん、良く研究室に泊まって研究をしていた宮崎さん、Mac のことや身の回りのことで御指導いただいた木村さん、先端学際工学専攻の先輩である松本さん、いつもにこにこしている大屋さん、明るい留学生であった Franck さん、韓国からいらした文さん、日立から先端学際工学専攻に入られた甲本さん、小林君、永久さん、藤井君、浅野君、大西君、乾君、渡部君、遠藤君、赤羽君、掛川君、神谷君、杉山君、村上君、酒井さん等々本当に多くの皆さんにお世話になりました。深く御礼を申し上げたい。

最後に、応援をしてくれた家族の皆に感謝します。本研究進行中に他界した父には本論文の慣性を報告したい。母には、わがままばかり言ってきたにもかかわらず、熱心に応援してくれました。また、２才の息子には、研究生活の励みになると共に、発達過程が時として研究のヒントとなりました。また、妻には、土日も大学に行くか寝てるかであまり家族サービスもできず、また何と言っても会社を辞めて大学に行く決心を付けさせてくれ、金銭的にも精神的にも支えてくれた。ここに、感謝の意を表する。

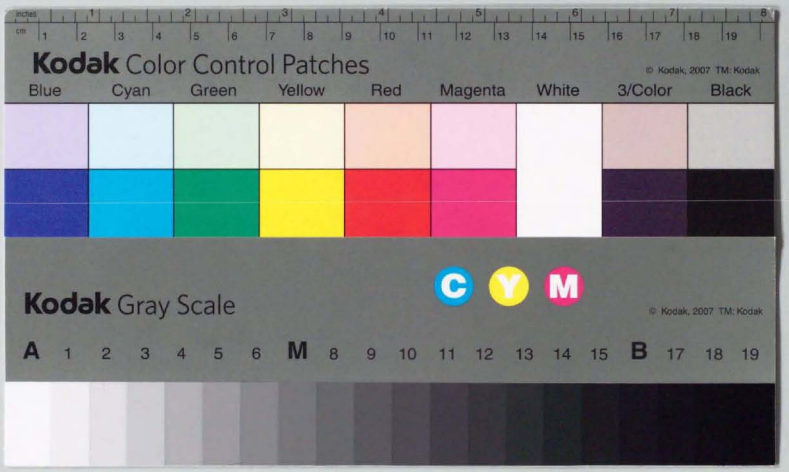
以上

１～１３８ページ　完

学位請求論文

平成８年１１月１日提出

柴田　克成


inches
cm
Kodak Color Control Patches
© Kodak, 2007 TM: Kodak
Blue
Cyan
Green
Yellow
Red
Magenta
White
3/Color
Black
C
Y
M
Kodak Gray Scale
© Kodak, 2007 TM: Kodak
A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19