

## 第3章 相關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關情報を挙げ、これを抽出する学習として相關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み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。

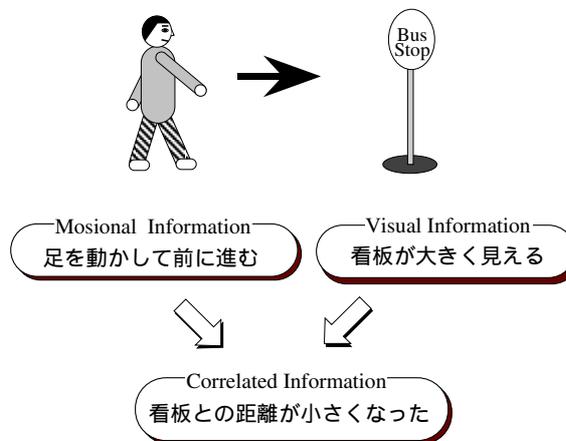


図3.1 運動と視覚の相關情報としての空間情報

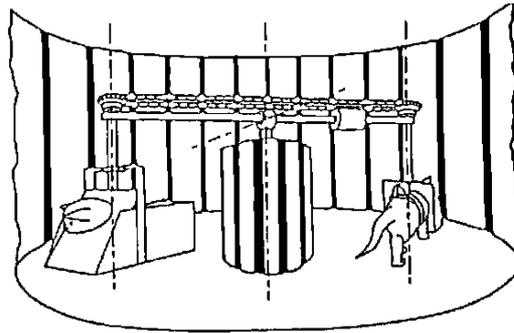


図3.2 Held & Hein の実験 ([Held 63] 津本忠治著「脳と発達」朝倉書店より転載)

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伸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入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]。

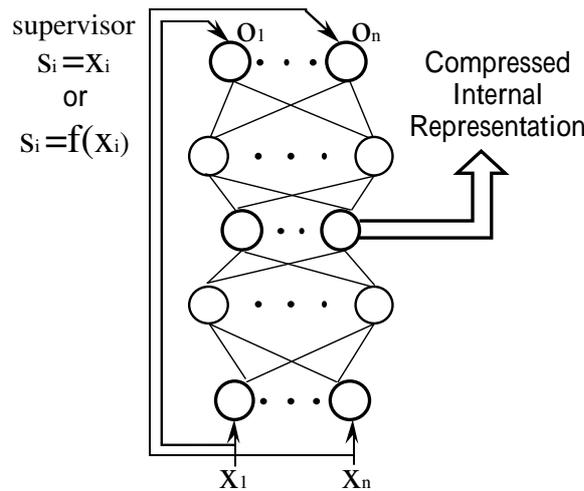


図 3.3 恒等写像ニューラルネットワーク (砂時計型ニューラルネット)

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、複数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[Becker 89]。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

複数種類の情報の中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

$$\mathbf{x}(t) = [x_1(t), x_2(t), \dots, x_m(t)]$$

$$\mathbf{y}(t) = [y_1(t), y_2(t), \dots, y_n(t)] \quad (3.1)$$

とし、この  $x, y$  から次式で表される情報ベクトル

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

$f, g$  : 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、 $x$  の関数と  $y$  の関数が  $t$  の変化によらず等しくなるような関数の組  $(f_1(x(t)), g_1(y(t))), (f_2(x(t)), g_2(y(t))) \dots$  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  $x(t)$ 、各網膜細胞の出力を  $y(t)$  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。

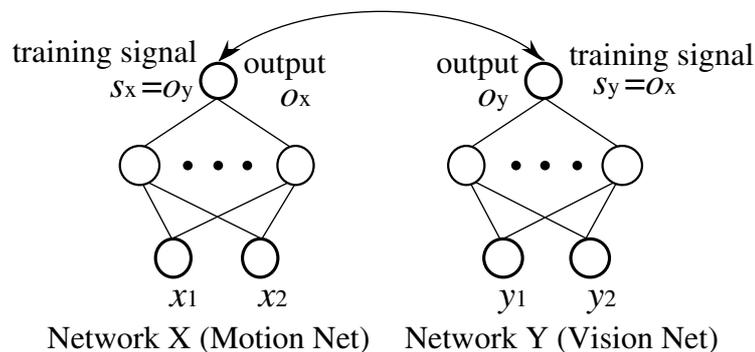


図3.4 相関情報抽出ネットワークの構成と学習方法

しかし、実際に学習を行なうと、

$$f(x(t)) = g(y(t)) = \text{const.} \quad (3.3)$$

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ふ。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

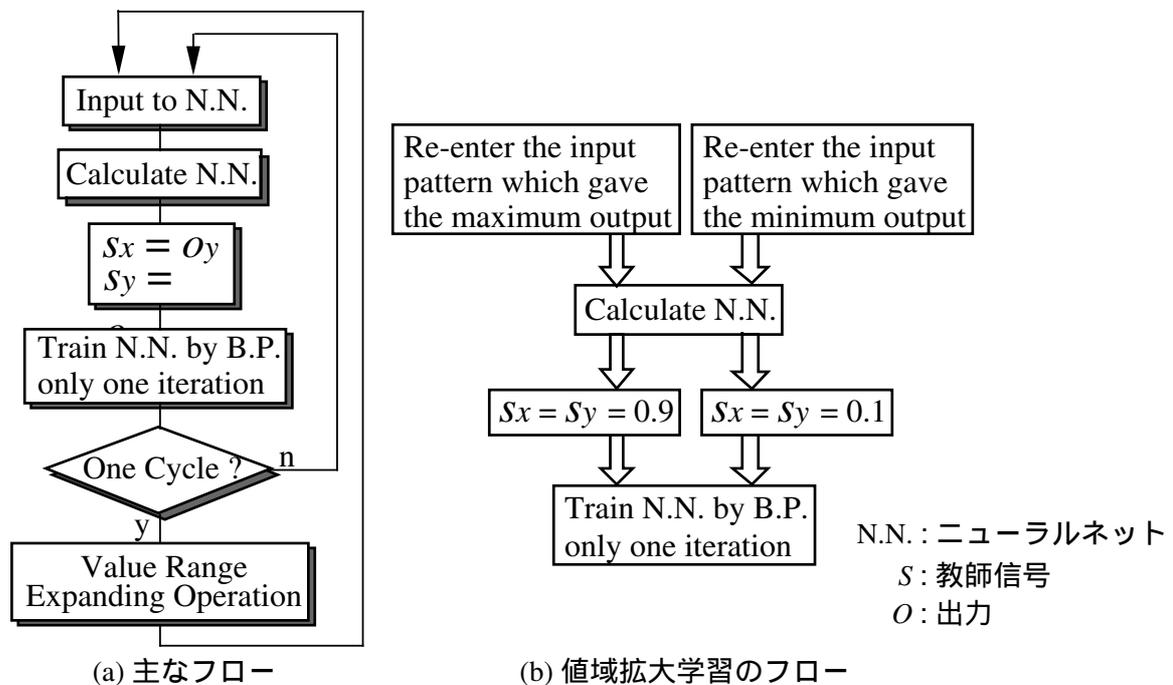


図 3.5 相関情報抽出学習のフロー

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つに分けることができる。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関連情報を抽出する際の個々の関連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用している[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

関連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、関連情報を

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

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、誤差も大きいですが、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。

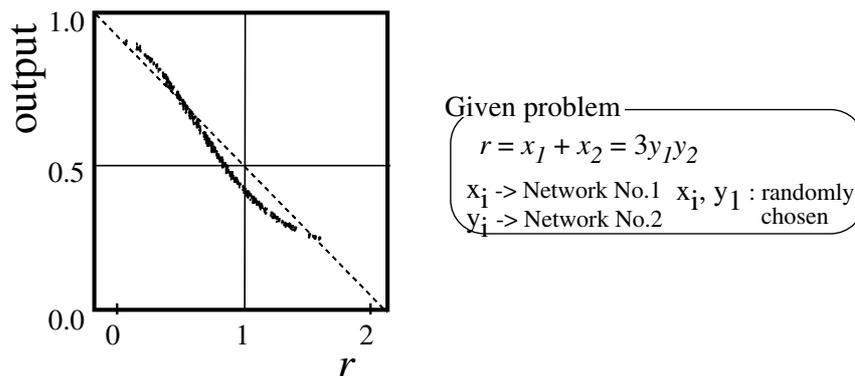


図 3.6 基本実験の結果

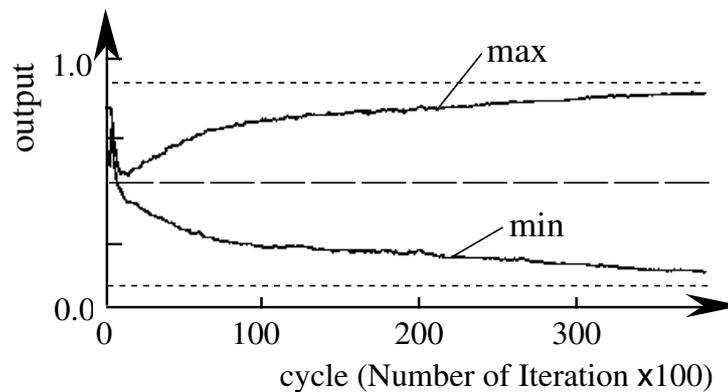


図 3.7 学習による出力の最大値と最小値の変化

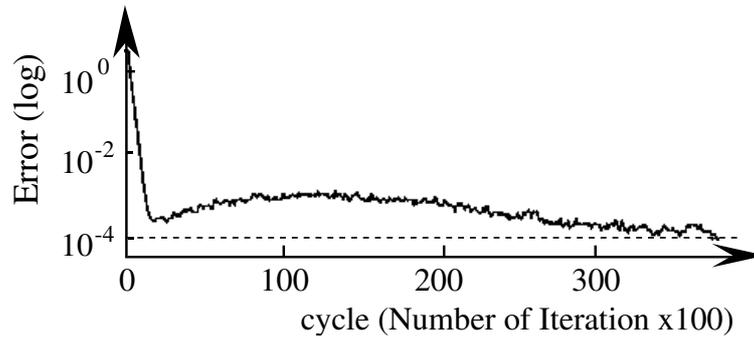


図 3.8 学習による誤差（両ネットワークの出力の差）の変化

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個として次の式のような相関関係を持たせて学習を行なった。

$$r_1 = x = 0.1 / y$$

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

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 $r_2$ の対称性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

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

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の時の出力値に近づいている。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。

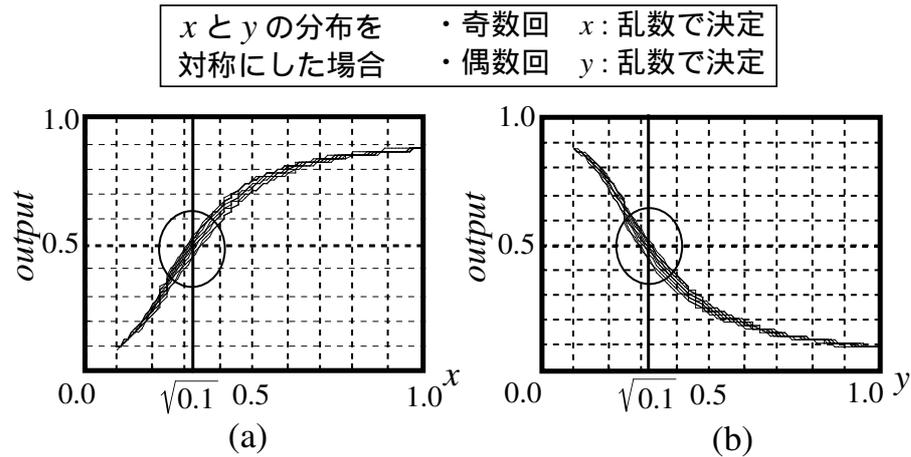


図 3.9 入力データに対称性を持たせた場合の出力の分布

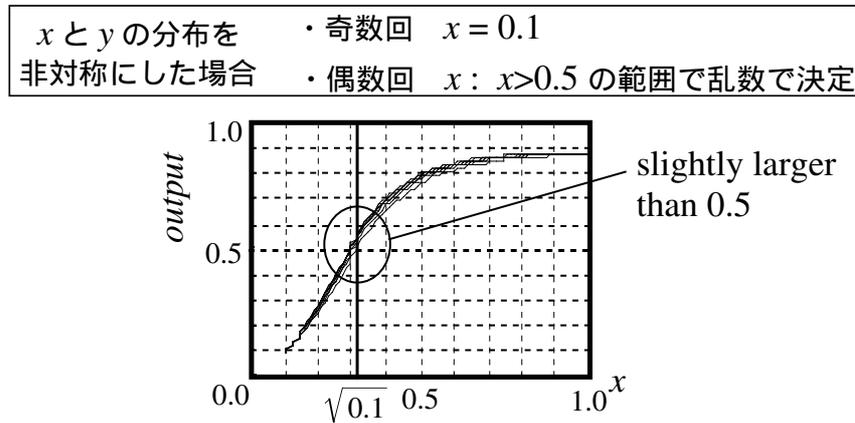


図 3.10 入力データを非対称にした場合の出力の分布

### 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  $r$  が

$$r(t) = f(x(t)) = g(y(t)) \quad (3.8)$$

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  $r$  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う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

$$ind_i(t) = \frac{(out_i(t) - mid_i)^2}{\sum_{j \neq i} (out_j(t) - mid_j)^2 + \theta} \quad (3.9)$$

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

$o_{n,i}$ :  $n$  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、

$\theta$ : 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、 $out_j(t) - mid_j$  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、

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

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

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

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

$in_{n,i}$ :  $n$  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。

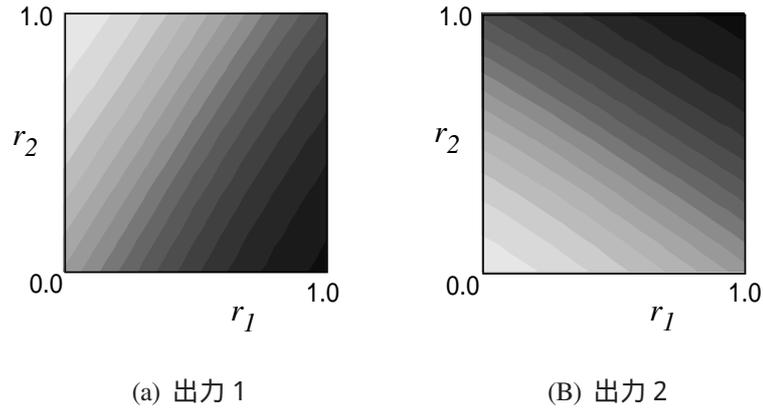


図3.1.1 相関情報  $r_1$ 、 $r_2$  に対する2つの出力の分布  
(それぞれ、2つのネットワークの出力の平均で、色の濃淡が値の大きさを示す)

また、入力と相関情報の間に非線形な関係を設定すると、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( (3.1.0) 式)とし、2つめのニューラルネットの入力を

$$in_{2,1} = r_1 \cdot r_2$$

$$in_{2,2} = r_1 / (r_2 + 0.1) / 10.0 \quad (3.1.2)$$

とすると、図3.1.2のように、出力分布はほぼ直交していることがわかる。

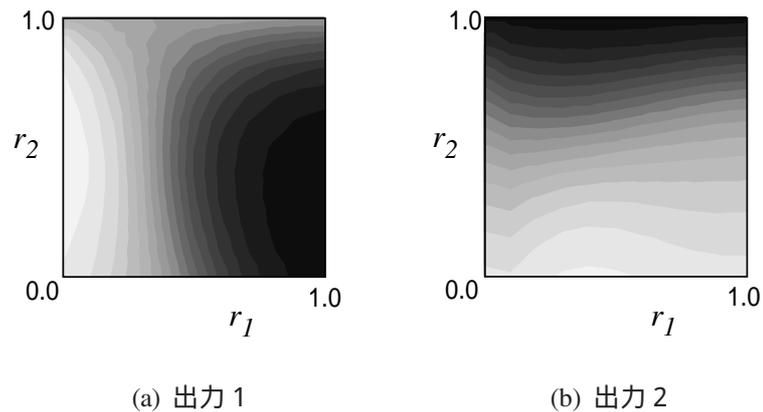


図3.1.2 相関情報と入力の関係が非線形である場合の相関情報  $r_1$ 、 $r_2$  に対する2つの出力の分布

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、入力が関連情報の線形結合（(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関連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。

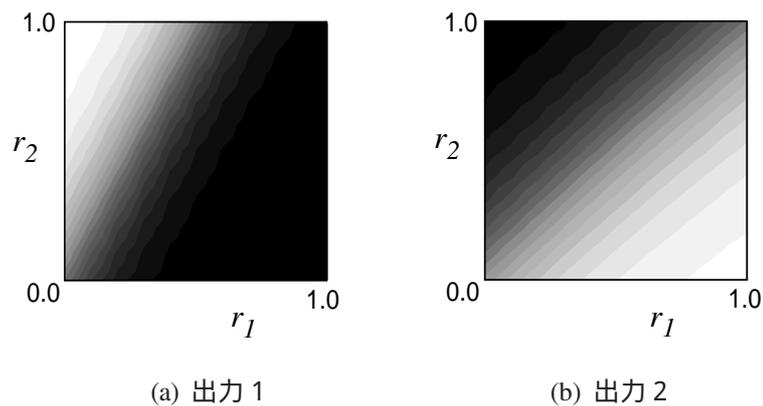


図3.13  $r_1$ 、 $r_2$ のどちらかを0.5に固定した場合の2つの出力の分布

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。

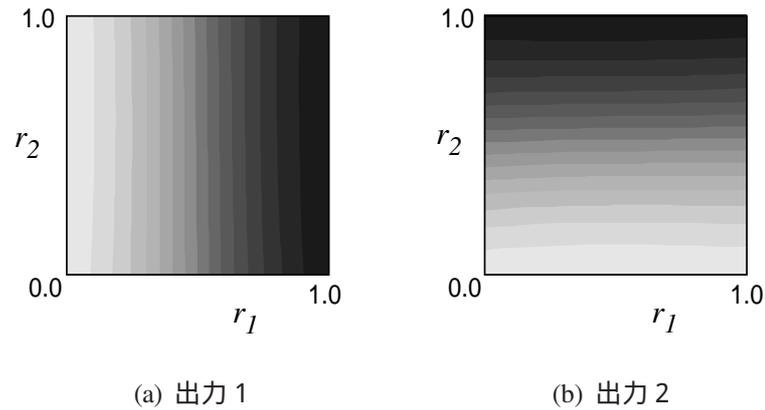


図3.14 出力間直交化学習の後、各出力の最大値、最小値を出力した時の入力値に対し、それぞれ独立度の最大値、最小値をとった時の出力を教師信号とした学習を行った場合

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。

$$\begin{aligned}
 in_{1,1} &= (2r_1 + r_2 + r_3) / 4.0 \\
 in_{1,2} &= (r_1 + 2r_2 + r_3) / 4.0 \\
 in_{1,3} &= (r_1 + r_2 + 2r_3) / 4.0 && (3.13) \\
 \\
 in_{2,1} &= (r_1 + r_2 + r_3) / 3.0 \\
 in_{2,2} &= (r_1 + r_2 - r_3 + 1.0) / 3.0 \\
 in_{2,3} &= (r_1 - r_2 + r_3 + 1.0) / 3.0 && (3.14)
 \end{aligned}$$

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。

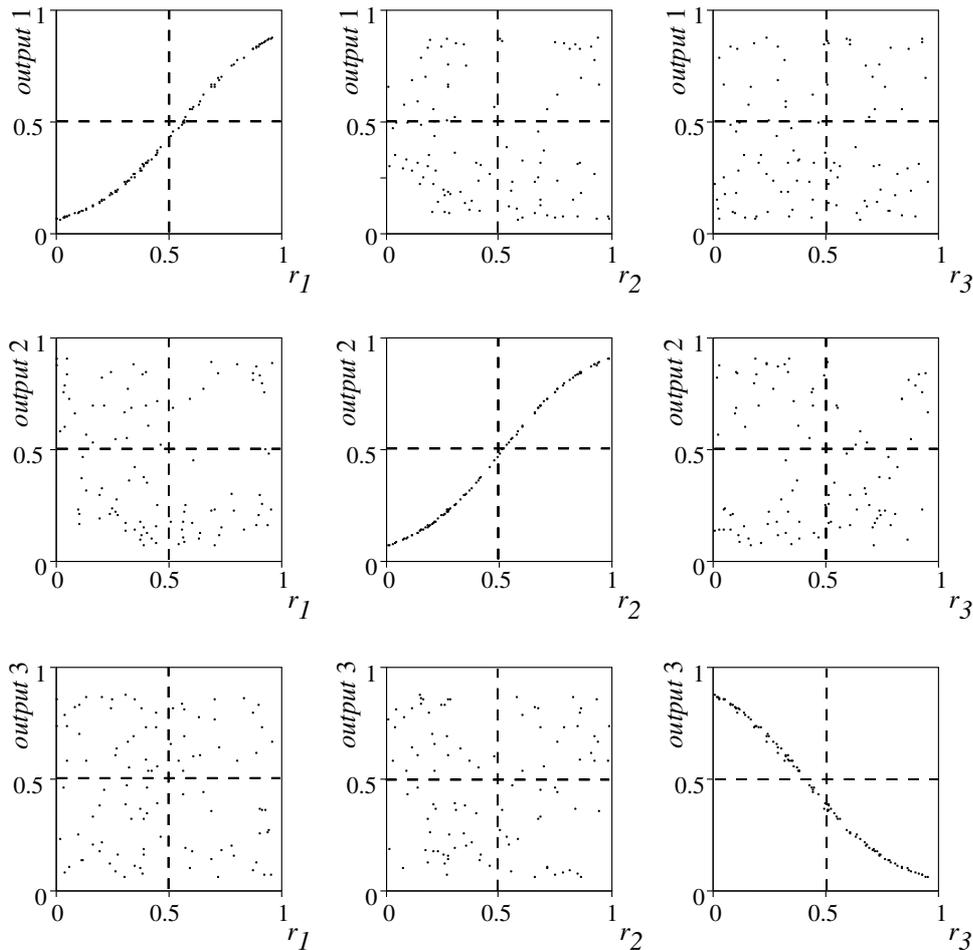


図 3.15 相関情報が 3 次元の場合の学習結果

### 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する。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、 $X, \Phi$  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  $X, \Phi$  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、 $X, \Phi, z$  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

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

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。

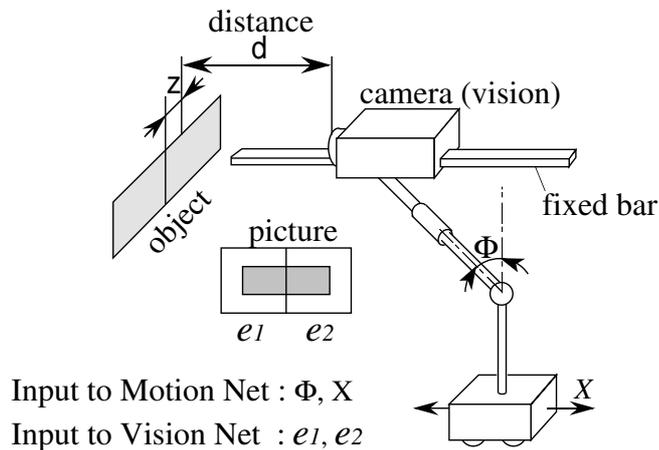


図3.16 シミュレーションで仮定した視覚センサ付きロボット

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  $X, \Phi$  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、 $X, \Phi$  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。

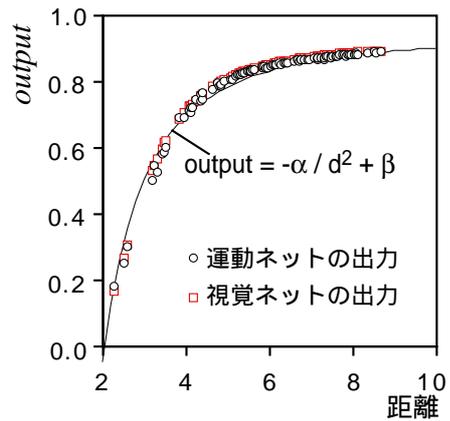
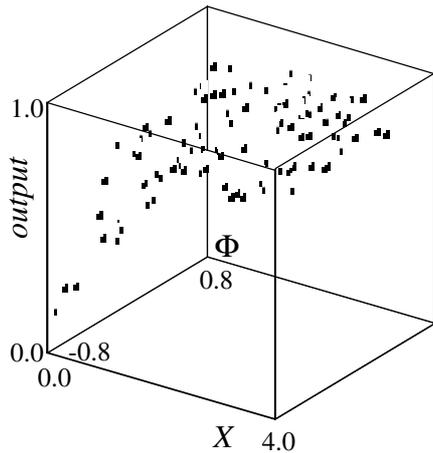


図 3.17 学習後の  $X$ ,  $\Phi$  に対する出力の分布 図 3.18 学習後の距離  $d$  に対する出力の分布

### 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  $\Phi$ ,  $X$  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、 $\Phi$ 、 $X$ 、 $l$ 、 $t$ 、 $z$  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  $\Phi$ 、 $X$  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。 $\Phi$ 、 $X$ 、 $l$ 、 $t$ 、 $z$  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、 $\Phi$  が  $0 \sim \pi/4$ 、 $X$  が  $5.0 \sim 8.0$ 、 $l$  が  $2.0 \sim 4.0$ 、 $t$  が  $0.5 \sim 1.0$ 、 $z$  が  $-1.0 \sim 1.0$  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。



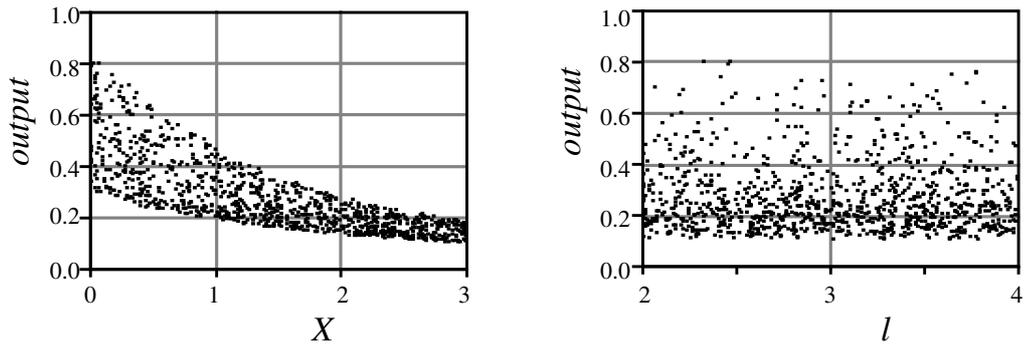


図 3.2 1 ロボットの駆動部  $X$  および物体の長さ  $l$  と両ニューラルネットの出力

### 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 3.2 2 のように、 $x, y, \Phi$  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。

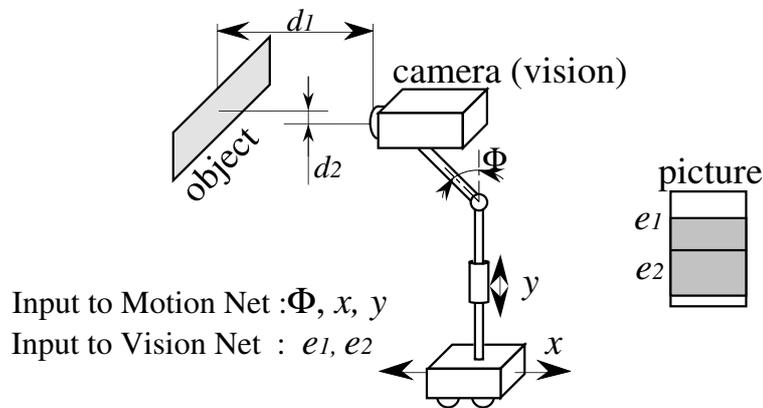


図 3.2 2 2次元の運動をする視覚センサ付きロボット

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 3.2 3 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眺めると、図 3.2 4 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。

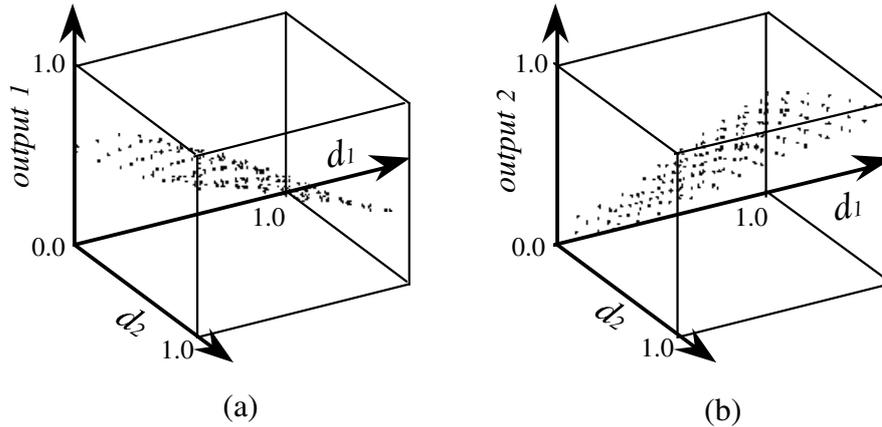
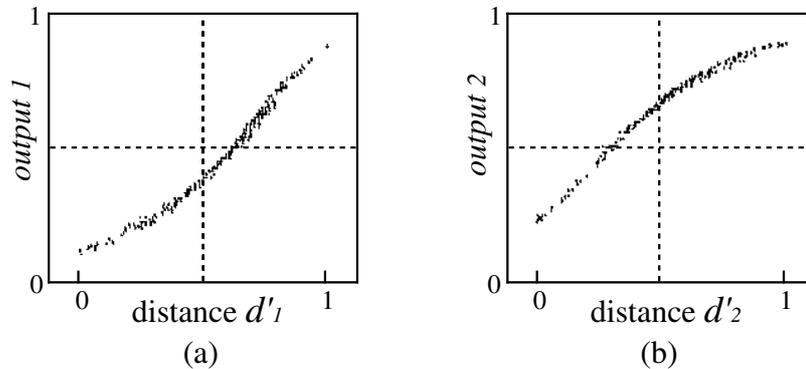


図 3.2.3 ロボットと物体の前後、上下の位置に対する各出力の値の分布

図 3.2.4  $d_1, d_2$  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 3.2.5 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、接触するかないかのあたりでは、出力が、0から1の値域に対して、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に似ていると考えられる。

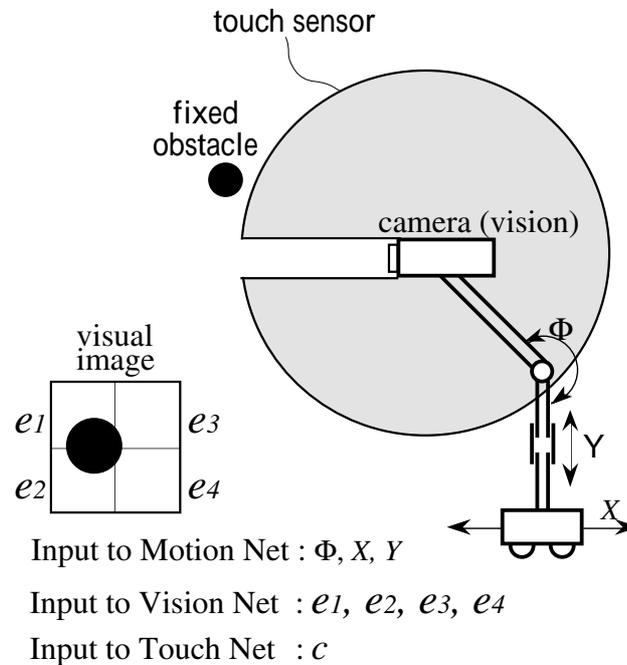


図 3.25 接触センサ、視覚センサを有するロボット

### 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すると、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関連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、それらの情報の関連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、本論文の関連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

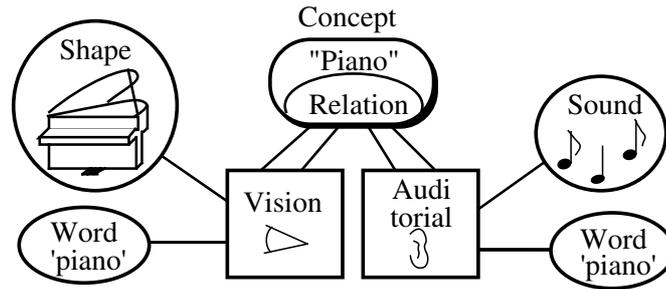


図 3.26 概念形成のモデル

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、さらに相關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られる。

### 3.7 まとめ

異種情報源の情報から相關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關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、学習後の相關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、相關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、抽出する相關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、入力情報と相關情報の上に線形性が強い場合、きれいに直交化された複数の相關情報が抽出できることがわかった。また、相關情報の分布領域が  $n$  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關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。