

錯視図形と相転移

東京理科大学理工学部情報科学科 大矢雅則 (Masanori Ohya)

要約 本論文では, Ising モデルの考えを基に錯視図形の認識過程をモデル化し, 以下の解析を行う. 第一に, 物理学などにおいて論じられる相転移が, 二義図形の錯視現象に現れることを, Shannon のエントロピーを計算し, その導関数の不連続点などによって調べる. 第二に, 認識が個人の経験, 心理状態などによって異なってくるという事を, 認識結果の時間的変化を追跡することによって確認する. すなわち, 認識が起こる直前の素子の状態を初期状態とし, その影響が認識結果にどのように現れるかの実験を上記のモデルによって理論的に説明する. そのため, ある二義図形における認識実現機械を構成するために, 逆伝搬法というニューラルネットワークの学習法を用いることにする.

1. まえがき

認識とは, 抽象的に言えば存在物とその在り方とを関係づけることである. 我々は対象とする存在物を見たり触ったりすることによってその様々な性質を刺激として受け取り, それによって対象物の在り方を認識する. ここでは, 認識論の数理的モデルとして, 錯視現象の解析を通して構成する. 今までの錯視の研究は, 心理学及び医学の分野でのものが主であり, その数理的な研究はそれほど行われていない. 錯視現象を, 数理とくに物理や情報理論との関わりの下で解析することができれば, 認識過程一般の数理的取り扱いが可能になるかもしれない.

以下では認識過程を数理的に追っていくために次のような 3 段階に分けて考えることにする [6].

1. 対象物から刺激を受ける

認識を生じる脳は, ニューロン (以下素子と呼ぶことにする) が多数結合して作られていると考えられる. 我々は感覚器官によって対象物を感じ取るが, それは各々の素子が刺激を受け, その一部は励起したものと考えられている. 今, 認識内容が 2 つのカテゴリ $\mathcal{C}_1, \mathcal{C}_{-1}$ に属する対象のみ取り扱い, i 番目の素子の状態 $\sigma_i \in \{-1, +1\}$ は, $\mathcal{C}_1$ に対応する刺激を受けると $\sigma_i = 1$, $\mathcal{C}_{-1}$ に対応する刺激を受けると $\sigma_i = -1$ の値をとるものとする.

2. 感覚を生じる.

ここでは, 刺激によって作られた n 個の素子の状態 $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$ が

ら得られる閉区間 $[-1, 1]$ 内の実数値 λ を、感覚と呼ぶことにする。 σ 内に $\mathcal{C}_1$ に対応する刺激を受けた素子が多ければ λ の値は $+1$ に近づき、 $\mathcal{C}_{-1}$ に対応する刺激を受けた素子が多ければ -1 に近づくよう、 σ と λ の関係が決められる。

3. 認識を生じる.

対象があらかじめ設定されたカテゴリー $\mathcal{C}_1, \mathcal{C}_{-1}$ のいずれに属するかを感覚 λ によって判別するのが、認識で μ である。二値認識を扱う限り、この μ は -1 か $+1$ の値を取るものと考えられる。以下では $\lambda \in [-1, +1]$ からカテゴリー $\mathcal{C}_1$ に属すると判断すれば $\mu = 1$, $\mathcal{C}_{-1}$ に属すると判断すれば $\mu = -1$ をとるものとする。

2. Ising モデルによる認識のモデル化

認識過程をモデル化するために、磁性体の議論に用いられる Ising モデルの考えを適用することにする [6]。Ising モデルとは、格子点上のスピンが 2 つの状態 $\sigma_i = 1$ (i 番目のスピンは上向き), $\sigma_i = -1$ (i 番目のスピンは下向き) のいずれかをとり、全体としてモデルの性質をつくり出すものである (図 1)。以下では n 個の素子群から成るモデルを系と呼ぶ。

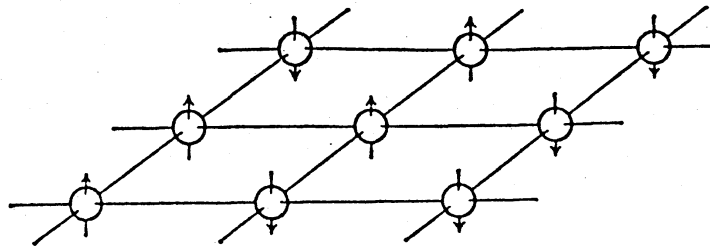


図 1 Ising モデル

一般に、系のエネルギーは

$$E(\sigma_1, \dots, \sigma_n) = - \sum_{i=1}^n \sum_{j=1}^n J_{ij} \sigma_i \sigma_j + \sum_{i=1}^n K_i \sigma_i \quad (1)$$

で定められる。第 1 項はスピン間の相互作用により生じるエネルギーの総和であり、第 2 項は外界からの系の受ける作用を表している。 J_{ij} は相互作用の強さを表す係数で、一般的には i, j スピン間の距離のみに依存すると考えられる。そし

て、その符号は

$$\begin{cases} J_{ij} > 0 \iff \text{強磁性相互作用 (向きをそろえる傾向)} \\ J_{ij} < 0 \iff \text{反強磁性相互作用 (交互に向く傾向)} \end{cases}$$

を意味する。

系の状態の総数は、 n 個の素子 (スピン) の状態によって決定されるので、全部で 2^n 通りになる。今、系が j 番目の状態 $\sigma^{(j)}$ をとる確率とそのエネルギーをそれぞれ p_j, E_j ($j=1, \dots, 2^n$) とおくと、

$$\sum_{j=1}^{2^n} p_j = 1 \quad (2)$$

$$\sum_{j=1}^{2^n} p_j E_j = E \quad (3)$$

が成り立つ。ただし E は平均エネルギーで、一定値にあると仮定する。

系が平衡状態にある確率分布を求めることにする。平衡状態とは、条件 (2), (3) のもとで完全事象系

$$\begin{pmatrix} X \\ P \end{pmatrix} \equiv \begin{pmatrix} \sigma^{(1)} & \sigma^{(2)} & \dots & \sigma^{(2^n)} \\ p_1 & p_2 & \dots & p_{2^n} \end{pmatrix}$$

のエントロピー

$$S(p) \equiv - \sum_{i=1}^{2^n} p_i \log p_i$$

を最大にする確率分布 $p = (p_1, \dots, p_{2^n})$ で記述される状態である。ラグランジェの未定係数 α, β を用いて

$$\frac{\partial}{\partial p_j} \left\{ S(p) + \alpha \left(1 - \sum_{j=1}^{2^n} p_j \right) + \beta \left(E - \sum_{j=1}^{2^n} p_j E_j \right) \right\} = 0$$

を計算すると、

$$p_j = \exp(-1 - \alpha - \beta E_j)$$

となる。式 (2) に代入すると、

$$p_j = \frac{\exp(-\beta E_j)}{\sum_{k=1}^{2^n} \exp(-\beta E_k)} \quad (j=1, \dots, 2^n) \quad (4)$$

が得られる。ただし β は条件 (3) によって決められるものである。

今求めた確率分布がエントロピー $S(p)$ の値を最大にすることを示そう。上の分布を $p^* = (p_1^*, \dots, p_{2^n}^*)$ とおき、条件 (2), (3) を満たす任意の確率分布 p に対するエントロピーの差を計算すると、

$$S(p^*) - S(p) = S(p | p^*)$$

となるが、相互エントロピー $S(p | p^*)$ は常に非負なので [4]、任意の p に対して $S(p^*) \geq S(p)$ が成り立つことがわかる。この確率分布 p^* はしばしば カノニカル分布 と呼ばれる。以上の Ising モデルとエントロピーを認識の問題に適用し、認識過程をモデル化してみることにする。

素子は一様に並んでいて、それぞれ 2 つの状態 $+1, -1$ のいずれかをとるものとする。素子の間には相互作用がはたらくが、その大きさは素子間の距離のみに依存し、距離が少し離れると極端に小さくなると仮定する。つまり隣接する素子どうしにのみ同じ大きさではたらくと考えられ、系のエネルギー (1) は

$$E(\sigma_1, \dots, \sigma_n) = -J \sum_{\langle i, j \rangle} \sigma_i \sigma_j \quad (5)$$

と書き換えられる。ただし $\langle i, j \rangle$ は隣接するスピン対を表す。系の各状態 $\sigma^{(k)}$ ($k=1, \dots, 2^n$) は先程のカノニカル分布に従って見いだされるものとする。なお、式 (4) におけるパラメータ β は、認識においては図形から与えられる作用（例えばシュレーダーの階段図形においては 左右の壁の面積比など）に相当する。

更に、各素子の他の素子からの影響は、同じ平均値 $\bar{\sigma}$ で与えられると仮定する（平均場の近似）と、系のエネルギー (5) は

$$E(\sigma_1, \dots, \sigma_n) = -\frac{J}{2} z \sigma \sum_{i=1}^n \sigma_i$$

となる。z は各スピンの隣接するスピン数である。

この E を用いて、(4) の確率分布

$$p_j = \frac{\exp(-\beta E_j)}{\sum_{k=1}^{2^n} \exp(-\beta E_k)} \quad (j=1, \dots, 2^n)$$

から素子の平均 $\bar{\sigma}$ を求めよう。

$$\begin{aligned} \text{式 (4) の分母} &= \sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_n=\pm 1} \exp \left[\alpha \sum_{i=1}^n \sigma_i \right] \\ &= \sum_{k=0}^n {}_n C_k \exp \{ \alpha (2k-n) \} \\ &= \exp(-\alpha n) \{1 + \exp(2\alpha)\}^n \\ &= \{ \exp(\alpha) + \exp(-\alpha) \}^n, \end{aligned}$$

$$\begin{aligned} \text{式 (4) の分子} &= \exp \{ -\beta E(\sigma_1^{(j)}, \dots, \sigma_n^{(j)}) \} \\ &= \exp \left(\alpha \sum_{i=1}^n \sigma_i^{(j)} \right). \end{aligned}$$

ただし、 $\alpha = \frac{\beta J z \bar{\sigma}}{2}$ である。これより素子の平均 $\bar{\sigma} = \bar{\sigma}_k (k=1, \dots, n)$ は

$$\begin{aligned} \bar{\sigma} &= \sum_{j=1}^{2^n} \sigma_k^{(j)} p_j \\ &= \sum_{j=1}^{2^n} \sigma_k^{(j)} \frac{\exp \left(\alpha \sum_{i=1}^n \sigma_i^{(j)} \right)}{\{ \exp(\alpha) + \exp(-\alpha) \}^n} \end{aligned}$$

と表される。この式の分子は

$$\begin{aligned}
\sum_{j=1}^{2^n} \{ \sigma_k^{(j)} \exp(\alpha \sum_{i=1}^n \sigma_i^{(j)}) \} &= \sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_n=\pm 1} \{ \sigma_k^{(j)} \exp(\alpha \sum_{i=1}^n \sigma_i) \} \\
&= \sum_{\sigma_i=\pm 1} \sigma_i \exp \alpha \sigma_i \sum_{k=0}^{n-1} C_k \exp \{ \alpha (2k - (n-1)) \} \\
&= \sum_{\sigma_i=\pm 1} \sigma_i \exp \alpha \sigma_i \{ \exp \alpha + \exp(-\alpha) \}^{n-1} \\
&= \{ \exp \alpha - \exp(-\alpha) \} \{ \exp \alpha + \exp(-\alpha) \}^{n-1}
\end{aligned}$$

と計算されるので、 $\bar{\sigma}$ は

$$\begin{aligned}
\bar{\sigma} &= \frac{(\exp \alpha - \exp(-\alpha)) (\exp \alpha + \exp(-\alpha))^{n-1}}{(\exp \alpha + \exp(-\alpha))^n} \\
&= \tanh \alpha
\end{aligned} \tag{6}$$

と求められる。

次に感覚を表す変数 λ について考える。前にも記したように、 λ は系（素子全体）の状態 $(\sigma_1, \dots, \sigma_n)$ を巨視的にとらえた時に、2通りの刺激をそれぞれの程度受け取ったかを表す値である。この λ を、ここでは最も簡単な

$$\lambda \equiv \frac{1}{n} \sum_{i=1}^n \sigma_i$$

の形で定めることにする。

次に認識を表す変数 $\mu \in \{-1, +1\}$ について考える。これは感覚 λ に依存して決まるものなので、条件付確率

$$p(\mu | \lambda) \equiv \frac{1 + \mu \lambda}{2} \tag{7}$$

によって与えることにする。 λ は $(\sigma_1, \dots, \sigma_n)$ によって一意に決まるので、 μ と $(\sigma_1, \dots, \sigma_n)$ の同時確率分布は

$$p(\mu, \sigma_1, \dots, \sigma_n) = p(\mu | \lambda) p(\sigma_1, \dots, \sigma_n)$$

と考えられ、これから μ の周辺分布は

$$\begin{aligned} p(\mu) &= \sum_{\sigma_1=\pm 1} \cdots \sum_{\sigma_n=\pm 1} p(\mu, \sigma_1, \dots, \sigma_n) \\ &= \frac{1 + \mu \bar{\sigma}}{2} \end{aligned} \quad (8)$$

と計算される。よって μ の平均値は

$$\begin{aligned} \bar{\mu} &= 1 \cdot \frac{1 + \bar{\sigma}}{2} + (-1) \cdot \frac{1 - \bar{\sigma}}{2} \\ &= \bar{\sigma} \end{aligned}$$

となり、(6) より認識 $\bar{\mu}$ は

$$\bar{\mu} = \tanh \left(\frac{\beta J_Z \bar{\mu}}{2} \right) \quad (9)$$

と求められる。

3. 錯視図形の認識と相転移

相転移というと、物理学で取り扱う自然現象であると思っている人が多いであろう。しかし、人間の認識、特に錯視現象においては相転移という概念を用いることによってうまく説明できる場合がある。

まず相転移について物理学を越えた一般的な定義を与えておこう[4]。様々な自然科学の研究においては、その対象としている系が指し示す“状態”(これを ϕ と書く)の変化の様子を見ることが重要である。このとき、系特有のあるパラメータ β の変化に伴ってこの状態が巨視的に異なるものに変化することを相転移と考えている。例えば H_2O は、温度というパラメータの変化によって異なる 3 つの状態、氷、水、蒸気(およびそれらの混合)に変化する。相転移とは、状態があるパラメータ β を境にして 2 つ(以上)の巨視的に異なる状態の間を移り変わることを言うのである。では、巨視的な量とはどのようなものであり、それはどのように計算しうるものかという、これに関する一般論はなく、対象に

応じて個々に決められているのが現状である。例えばエントロピー S 等のいわゆる熱力学的な諸量を使うと

$$\left\{ \begin{array}{l} S \text{ が } \beta \text{ に関して不連続} \iff \text{第 1 種相転移} \\ \beta \frac{\partial S}{\partial \beta} \text{ が } \beta \text{ に関して不連続} \iff \text{第 2 種相転移} \end{array} \right.$$

のように表される。

さて、錯視現象における認識を相転移という観点から考えることにしよう。対象は二義図形のみとし、パラメータ θ を適当に決めた時にその変化につれて図形の認識にどのような変化が現われるかを調べることにする。 n 個の素子に伝えられる図形からの刺激は θ に依存するものであるから、 $\beta(\theta)$ とおく。素子の応答に応じて感覚を生じ、このとき認識 $\mu(\theta)$ を生じる条件付確率は、(7) より

$$p(\mu(\theta) | \lambda) = \frac{1 + \mu(\theta)\lambda}{2},$$

そして認識 $\mu(\theta)$ の平均は、 $J(\theta) = \beta(\theta)J_Z/2$ とおくと (9) より

$$\bar{\mu}(\theta) = \tanh(J(\theta)\bar{\mu}(\theta)) \quad (10)$$

で与えられていた。

まず認識の巨視的な量のひとつである $\bar{\mu}(\theta)$ の、 $J(\theta)$ に対する変化率を調べる。 $\bar{\mu}(\theta)$ の微分

$$\frac{d\bar{\mu}(\theta)}{dJ(\theta)} = \{1 - \tanh^2(J(\theta)\bar{\mu}(\theta))\} \left\{ \bar{\mu}(\theta) + \frac{d\bar{\mu}(\theta)}{dJ(\theta)} \right\}$$

より、

$$\frac{d\bar{\mu}(\theta)}{dJ(\theta)} = \frac{2\bar{\mu}(\theta)^2(1-\bar{\mu}(\theta)^2)}{2\bar{\mu}(\theta) - (1-\bar{\mu}(\theta)^2)\log\frac{1+\bar{\mu}(\theta)}{1-\bar{\mu}(\theta)}} \quad (11)$$

が得られる. 式 (11) の分子を f , 分母を g とおくと

$$\lim_{\bar{\mu} \rightarrow \pm 0} f = \lim_{\bar{\mu} \rightarrow \pm 0} g = 0$$

なので, l'Hospital の定理を用いて極限が

$$\begin{aligned} \lim_{\bar{\mu} \rightarrow \pm 0} \frac{d\bar{\mu}(\theta)}{dJ(\theta)} &= \lim_{\bar{\mu} \rightarrow \pm 0} \frac{df/d\bar{\mu}}{dg/d\bar{\mu}} \\ &= \lim_{\bar{\mu} \rightarrow \pm 0} \frac{2(1-4\bar{\mu}^2)}{\log \frac{1+\bar{\mu}}{1-\bar{\mu}}} \\ &= \pm \infty \end{aligned}$$

と求められる. $\bar{\mu}(\theta)=0$ において $J(\theta)$ に対する $\bar{\mu}(\theta)$ の変化率に不連続 (発散) が起こるということは, $\bar{\mu}(\theta)$ がその近傍で $J(\theta)$ に対して急激な変化をする, すなわち認識の相転移が起こったと考えることができる. $\bar{\mu}(\theta)=0$ とは, $p(\mu(\theta)=1)=p(\mu(\theta)=-1)=1/2$ なる場合であるから, この時の認識の状態 ϕ は,

$$\phi(\mu) = \begin{cases} \mu(\theta)=1 \text{ と認識しやすい} & \bar{\mu} > 0 \\ \mu(\theta)=-1 \text{ と認識しやすい} & \bar{\mu} < 0 \end{cases}$$

といえる. これはごく自然なことである.

次に, 各々の素子以外による作用の総和 $J(\theta)\bar{\sigma}=J(\theta)\bar{\mu}(\theta)$ に対するエントロピーの変化率を考える. 完全事象系 $(\mu(\theta), \{p(\mu(\theta))\})$ のエントロピー $S(p)$ は, (8) より,

$$S(p) = -\frac{1+\bar{\mu}}{2} \log \frac{1+\bar{\mu}}{2} - \frac{1-\bar{\mu}}{2} \log \frac{1-\bar{\mu}}{2}$$

となる。 $J(\theta) \bar{\mu}(\theta)$ に対する $S(p)$ の変化率を、次の合成関数の微分により求める：

$$\frac{dS(p)}{dJ(\theta) \bar{\mu}(\theta)} = \frac{dS(p)}{d\bar{\mu}(\theta)} \cdot \frac{d\bar{\mu}(\theta)}{dJ(\theta) \bar{\mu}(\theta)}$$

右辺の 2 つめの微分は、(10) を変形した

$$J(\theta) \bar{\mu}(\theta) = \tanh^{-1} \bar{\mu}(\theta) = \frac{1}{2} \log \frac{1+\bar{\mu}(\theta)}{1-\bar{\mu}(\theta)}$$

から得られる。すなわち

$$\frac{d\bar{\mu}(\theta)}{dJ(\theta) \bar{\mu}(\theta)} = 1 - \bar{\mu}(\theta)^2$$

よって

$$\frac{dS(p)}{dJ(\theta) \bar{\mu}(\theta)} = -\frac{1}{2} \log \frac{1+\bar{\mu}(\theta)}{1-\bar{\mu}(\theta)} (1 - \bar{\mu}(\theta)^2)$$

となるが、さらに 2 回微分は、

$$\frac{d^2 S(p)}{d(J(\theta) \bar{\mu}(\theta))^2} = \left\{ \bar{\mu}(\theta) \log \frac{1+\bar{\mu}(\theta)}{1-\bar{\mu}(\theta)} - 1 \right\} (1 - \bar{\mu}(\theta)^2)$$

と計算される。右辺=0 の解を求めると、

$$\bar{\mu}(\theta) = \pm 1, \pm 0.648$$

を得るが、このうち $\bar{\mu}(\theta) = \pm 0.648$ は $S(p)$ の変曲点なので、 $S(p)$ の変化率が一番大きいといえる。よって被験者の回答の予測に関する不確定さ、すなわち判断の困難さも最も大きく、この値を境にして判断の困難さは大きな差を生じる

と考えられる。 $\bar{\mu} = \pm 0.648$ において相転移が起こっていると受け取れるわけである。認識の状態 ϕ は、

$$\phi(\bar{\mu}) = \begin{cases} \mu = 1 \text{ なる認識を生じやすい} & \bar{\mu} > 0.648 \\ \text{どちらも断定しにくい} & |\bar{\mu}| < 0.648 \\ \mu = -1 \text{ なる認識を生じやすい} & \bar{\mu} < -0.648 \end{cases}$$

とみなせる。

そこで、以上の考察をシュレーダーの階段図形（図 2）の実際の認識に適用してみることにしよう。

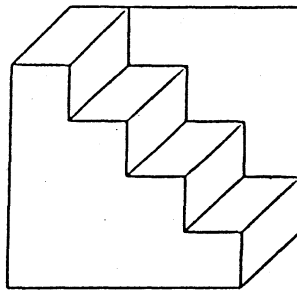


図 2 シュレーダーの階段図形

この図形は左下の壁の方が右上の壁よりも手前に見えてくれれば階段を見おろしたように見え、逆であれば天井に張りついた階段を見上げるように見えるという二義図形である。図 2 の場合、後者に見えるのは比較的 unnatural なので、それを解消するために 45 度傾けたものを実験に用いることにする（図 3）。

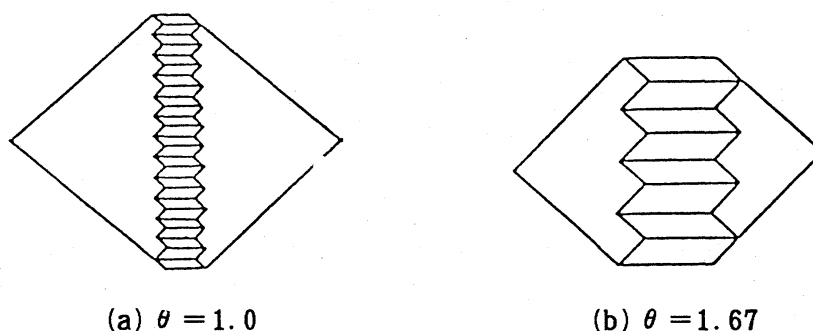


図 3 45 度傾けたシュレーダーの階段図形

なお、実験は左右の壁の面積比 S_L/S_R をパラメータ θ とし、被験者には θ の値を変化させた図を見てもらい、どちらの壁が手前に見えるか答えてもらう。図 3 を見ると、 θ の値が大きい図形ほど左の壁が手前に見える確率が高くなり、ちょうど左右同面積の $\theta = 1.0$ 辺りで答えが半分ずつに別れると予想される。表 1 は面積比 0.6~1.67 の 9 図形を使用して被験者 89 名に対する実験を行った結果である。 $p_i (\mu = 1)$ は、左の壁が手前に見えるという事象 $\mu = 1$ を得る確率である。

θ_i	0.6	0.71	0.8	0.9	1.0	1.11	1.25	1.4	1.67
人数	12	18	26	36	43	61	71	73	74
$p_i (\mu = 1)$	0.13	0.20	0.29	0.40	0.48	0.69	0.80	0.82	0.83

表 1 各 θ_i において左の壁が手前に見えると解答した人数及び確率

ここで $\eta \equiv \log \theta$ とパラメータを置き換え、実験結果をもとに $p(\mu(\eta) = +1)$ を関数近似する。ここでは

$$p(\mu(\eta) = +1) = \frac{\exp(a\eta + b)}{1 + \exp(a\eta + b)}$$

の形の S 字形関数を用い、最小二乗法によって定数 a, b の値を決める。すなわち、 $\eta_i = \log \theta_i$ ($i = 1, \dots, 9$) とおき、各

$$p_i = \frac{\theta_i \text{ の図形を見て左が手前に見えた人数}}{\text{被験者数 (= 89 名)}}$$

の値に対して

$$\sum_{i=1}^9 \left\{ p_i - \frac{\exp(a\eta + b)}{1 + \exp(a\eta + b)} \right\}^2$$

を最小にするような a, b を求めるのである。結果は $a=4.271, b=0.110$ となる。この時 $\bar{\mu}(\eta)$ は,

$$\bar{\mu}(\eta) = \frac{\exp(4.271\eta + 0.11) - 1}{\exp(4.271\eta + 0.11) + 1}$$

で近似されるので、先ほど論じた相転移点 $\bar{\mu}=0$ は $\eta = -0.026$, すなわち $\theta = 0.974$ の点とわかる。状態 ϕ は

$$\phi(\theta) = \begin{cases} \text{左が手前に見えやすい} & \theta < 0.974 \\ \text{右が手前に見えやすい} & \theta > 0.974 \end{cases}$$

と表される。

次に、作用の総和に対するエントロピーの変曲点から求めた相転移点 $\bar{\mu} = \pm 0.648$ を考える。 $\bar{\mu} = 0.648$ の場合には $\theta = 1.399$ ($\eta = 0.336$), $\bar{\mu} = -0.648$ の場合には $\theta = 0.679$ ($\eta = -0.387$) なので、状態は

$$\phi(\theta) = \begin{cases} \text{左が手前に見えやすい} & \theta < 0.679 \\ \text{どちらが手前か迷う} & 0.679 < \theta < 1.399 \\ \text{右が手前に見えやすい} & \theta > 1.399 \end{cases}$$

と表される。

また、この近似関数を用いてエントロピー $S(p)$ の面積比 θ に対する変化率を考える。確率分布 p を $p \equiv (p_1, p_2)$, $p_1 \equiv \text{Prob}(\mu = +1)$, $p_2 \equiv \text{Prob}(\mu = -1)$ とおく。

$$\begin{aligned} S(p) &= -p(\mu=1) \log p(\mu=1) - (1-p(\mu=1)) \log (1-p(\mu=1)) \\ &= \frac{\eta' \exp \eta'}{1 + \exp \eta'} + \log (1 + \exp \eta') \end{aligned}$$

$$\frac{dS(p)}{d\theta} = -\eta' \cdot \frac{a \exp \eta'}{(1 + \exp \eta')^2} \cdot \frac{1}{\exp \eta'}$$

$$\begin{aligned}\frac{d^2 S(P)}{d\theta^2} &= \frac{d}{d\eta} \left\{ -\frac{a\eta' \exp \eta'}{\exp \eta (1 + \exp \eta')^2} \right\} \cdot \frac{d\eta}{d\theta} \\ &= -\frac{a \exp \eta' \{ \eta' (a-1 - (1+a) \exp \eta') + a(1 + \exp \eta') \}}{\exp 2\eta (1 + \exp \eta')^3}\end{aligned}$$

ただし, $\eta' \equiv a\eta + b$ とおく. $a=4.271$, $b=0.110$ を代入すると, S の変曲点が $\theta = 0.6185, 1.309$ と得られる. よって $\theta = 0.6185, 1.309$ において相転移が起こり, 状態 ϕ は

$$\phi(\theta) = \begin{cases} \text{左の壁が手前に見えやすい} & \theta < 0.6185 \\ \text{どちらが手前に見えるか迷う} & 0.6185 < \theta < 1.309 \\ \text{右の壁が手前に見えやすい} & \theta > 1.309 \end{cases}$$

と表されるといえる.

以上のように, 二義錯視図形の認識は, Ising モデルをベースとした認識過程のモデル化により非常にうまく説明できることがわかる. このことは錯視図形の認識に相転移が深くかかわっていることを示していると思われる.

4. 認識における記憶学習

前節では, ひとつのパラメータの図形に対する認識のみを考え, 刺激を受ける前の素子の初期状態を考慮に入れていない. 我々が認識を行なう際には, 無意識のうちに何かしらの潜在意識がはたらくものである. 例えば図 4 を見せられどちらが長いかと尋ねられたら, (b) かあるいは同じと答える人がほとんどであろう. しかし実際は, 同じと答えた人の中にも (b) が長く見えた人がいるはずである. 経験的に (b) の方が長く見えることを知っているせいで, "見た目より (b) は短いはずだ" と無意識のうちに脳にはたらきかけてしまい, 注意深く同じ長さであることを見抜いたと考えられる. また, 二義図形の認識においても個人の傾向, つまり判断のくせがあったりする. というのは, 例えば図 5 において初めに少女に見えた人は何度見ても少女に見えてしまうし, 逆にはじめに老婆に見えた人は少女に見ようとしてもやはり老婆に見てしまいがちだったりする [5]. これらのような現象は, 経験によって脳がその対象物についての情報を身につけ, 次回の判断に強く影響したために生じたことと思われる. このように考えると認識過程とは, 以前の経験をもとに連続的に発展するダイナミクスであると考えられる. あ

らかじめ情報を蓄積して次の認識を行なう（刺激を受ける）ということは，初期状態をもった脳が，刺激によって次の時刻の状態へ変化発展することというわけである．以下ではこの初期状態による効果を 記憶学習効果 と呼ぶことにする．

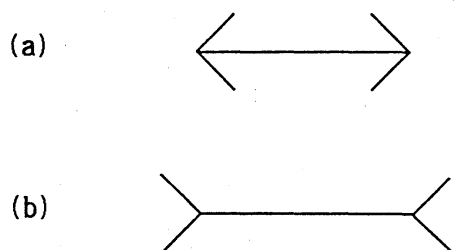


図 4



図 5

この節では，前節と同様にシュレーダーの階段図形（図 3）における記憶学習の効果について調べてみることにする．まず，認識過程に時間を導入するためにここで用いる，逆伝搬法というニューラルネットの学習法を復習する．

4.1 逆伝搬法

学習とは，システムが環境からの入力に応じて自身の構造を作り替えていく（自己組織化）ことである．例えば認識，記憶，概念形成など，脳のもつすぐれた機能は環境との相互作用を通してニューラルネットワークが行なう自己組織化過程と考えられる．生物神経系においては，ニューロン間のシナプス結合の強さを変化させることによって学習を行なっていると考えられている．従ってニューラルネットにおいては，学習といえば通常シナプス結合係数をなんらかのアルゴリズムによって変化させることを指す．

学習には大きく 2 通りのものがある．1 つはある入力に対してシステムが出力すべき値を外部からコントロールされている学習で，これは“教師あり学習”と呼ばれる．与えられる望ましい出力を“教師信号”という．これに対して“教師なし学習”とは，システムが，例えば入力パターンの出現頻度など環境の性質を自分で学習し，自己の構造に反映させる学習である．

ここでは認識への応用に用いるための学習法則として逆伝搬法（back propagation）を取り上げることにする．この方法は，上の 2 通りの学習則のうち教師あり学習のひとつである．まず，この逆伝搬法の基本となっているパーセプトロンについて説明することにしよう．

パーセプトロンは，1961年 Rosenblatt によって提案された有名な学習機械モデルである[7]．中心となるのは図 6 のような McCulloch-Pitts 形の素子に教師信号を与えた素子である．これを基本パーセプトロンと呼ぶことにしよう．

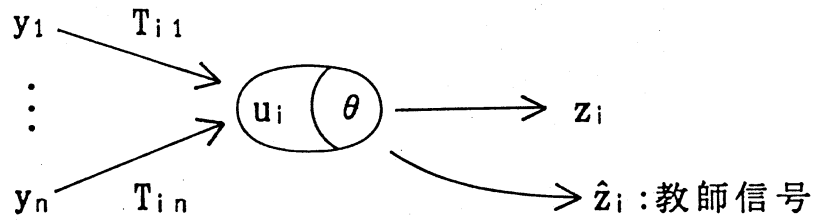


図 6 基本パーセプトロン

ここで θ はしきい値, u_i は入力 of 総和を表す. この素子の動作は, 式で表すと

$$u_i = \sum_{j=1}^n T_{ij} y_j - \theta \quad (12)$$

$$z_i = 1(u_i) = \begin{cases} 0 & u_i \leq 0 \\ 1 & u_i > 0 \end{cases} \quad (13)$$

となる.

n 個の要素をもつパターンベクトル $(y_1, \dots, y_n)$ の集合が与えられた時, ベクトルの各成分を 1 つ 1 つ座標軸に割り当てた n 次元のパターン空間ができあがる. 同じカテゴリーに属するパターンは, 通常数学的にも似たベクトルであると考えられるので, これらの点はパターン空間内にあるカテゴリーを作っていると考えられる. したがって, パターン認識とはパターン空間をカテゴリーごとに分割することに等しい. ひとつのパターン認識問題が与えられた時, 基本パーセプトロンは外部から受けた教師信号によって結合係数 T_{ij} を変化させ, 反復的に学習する. いま, $n+1$ 次元の係数ベクトル $T_i \equiv (T_{i1}, \dots, T_{in}, \theta)$ とパターンベクトル $Y \equiv (y_1, \dots, y_n, -1)$ を用いると, 式 (12), (13) は

$$z_i = \begin{cases} 0 & T_i \cdot Y \leq 0 \\ 1 & T_i \cdot Y > 0 \end{cases} \quad (14)$$

と書き換えられる. 最も一般的なパーセプトロンの学習則は, $\hat{z}_i$ を教師信号とするとすると, 入力パターンが与えられるたびに T_i を

$$\partial T_i = -\epsilon (z_i - \hat{z}_i) Y$$

だけ変化させるというものである。ただし ϵ は非常に小さな正定数である。
 z_i が誤った ($\hat{z}_i$ と異なる) 出力を出した時のみ T_i が変化する (改善される) ことがわかる。この学習則によって有限回学習を繰り返せば、パーセプトロンは正しい識別をするようになることが証明されている (パーセプトロンの収束定理 [7])。

式 (12) の u_i に 0 を代入したものは、パターン空間上の超平面を表す式となる。すると (13) の意味は、入力されたパターンがこの超平面のどちら側にあるかを判別することになる。このように超平面によってパターンを 2 つのカテゴリに分離できるような問題を線形分離可能な問題という。基本パーセプトロンは任意の線形分離可能な問題を有限回の反復学習で達成できることが示されているわけだが、現実にはそのままで線形分離可能な問題は非常に少ない。そこでなんらかの前処理を施すために層の数を増やしたパーセプトロンなどが考えられた (図 7)。

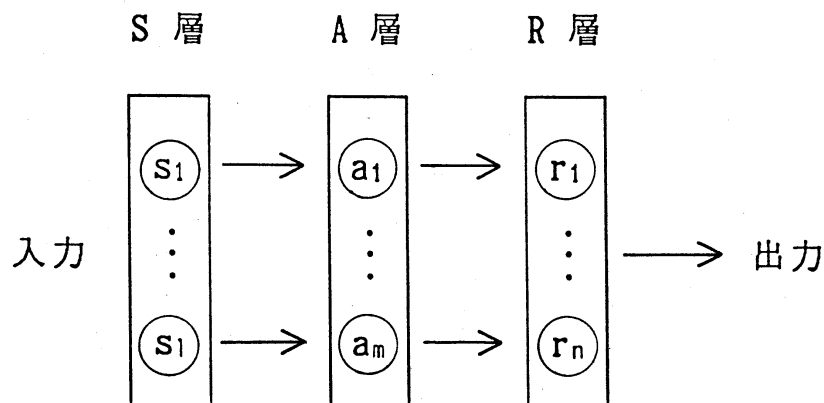


図 7 3 層パーセプトロン

ところが問題は、感覚層 (sensory units : S 層) で受け取った入力をどのように変換して連合層 (association units : A 層) へ送れば反応層 (response units : R 層) で線形分離処理を行えるであろうか、ということになる。A 層の素子数が十分多ければこの問題は解決しうることが [1] において示されているが、素子数や層数を極端に増やすことは現実的には計算量が莫大なものとなり難しい。そこで考え出されたのが逆伝搬法である。

逆伝搬法とは、ある程度現実的な数の中間層素子に積極的に学習をさせ、各問題に必要な回路を中間層に形成させる方法である。

まず準備として、上で述べたパーセプトロンの学習則を一般化しておこう。i

番目の素子は、他の素子の出力 y_j を結合係数 T_{ij} を乗じて受け、それらの入力
の総和

$$u_i = \sum_j T_{ij} y_j \quad (15)$$

に単調増加関数 f を施した

$$y_i = f(u_i) \quad (16)$$

を出力することにする。ここでは、以下で解析するために微分可能な関数

$$f(u_i) \equiv \tanh(u_i)$$

を用いることにする。学習の評価としては、 $\hat{y}_i(x)$ を入力パターン x に対して素
子 i が出力すべき値（教師信号）、 $y_i(x)$ をその時の実際の出力として、誤差関
数

$$E(x) \equiv \frac{1}{2} \sum_x \sum_i (y_i(x) - \hat{y}_i(x))^2 \quad (17)$$

を考える。 $y_i(x)$ がその時点での T_{ij} に依存しているため、 E は間接的に T_{ij}
の関数であるといえる。この関数値を最小にするような T_{ij} を求めていく方法を
最小 2 乗平均誤差法という。

各係数を学習させる規則としては、 T_{ij} を

$$\delta T_{ij} = -\epsilon \frac{\partial E}{\partial T_{ij}} \quad (18)$$

に従って変化させればよい。これは、各 T_{ij} を軸、 E を高さと考えた誤差曲面上
を、最も急な傾斜方向に進んでいくことに相当する（このような学習法を最急降
下法という）。

合成関数の微分により、

$$\frac{\partial E}{\partial T_{ij}} = \sum_x \frac{\partial E}{\partial y_i(x)} \cdot \frac{dy_i(x)}{du_i(x)} \cdot \frac{\partial u_i(x)}{\partial T_{ij}}$$

である。これに式 (15), (16) を考慮すると, (18) は

$$\delta T_{ij} = -\epsilon \sum_x \frac{\partial E}{\partial y_i(x)} f'(u_i) y_j(x) \quad (19)$$

となる。よって中間層が学習しない時には式 (17) から

$$\delta T_{ij} = -\epsilon \sum_x \{y_i(x) - \hat{y}_i(x)\} (1 - y_i^2(x)) y_j(x)$$

という学習則が得られる。

さて, いくつかの中間層をもつ場合を考えてみよう。図 8 のような, どの素子も 1 つ前の層からのみ入力を受け, 次の層へ出力を送るネットワークについて上と同様に学習則を導こうとすると, (19) の $\partial E / \partial y_i(x)$ の値がすぐには求められない。この値は, 次の層の素子 k を用いて

$$\frac{\partial E}{\partial y_i} = \sum_k \frac{\partial E}{\partial y_k} \cdot \frac{dy_k}{du_k} \cdot \frac{\partial u_k}{\partial y_i}$$

と展開でき,

$$\frac{\partial E}{\partial y_i} = \sum_k \frac{\partial E}{\partial y_k} f'(u_k) T_{ki}$$

となる (x は省略してある)。この式を見るとわかるように, 微分値は出力層より逆向きに順々に計算される, つまり出力の誤差は前の層へ前の層へと伝えられていく。この考え方が, "逆伝搬" と呼ばれるゆえんである。

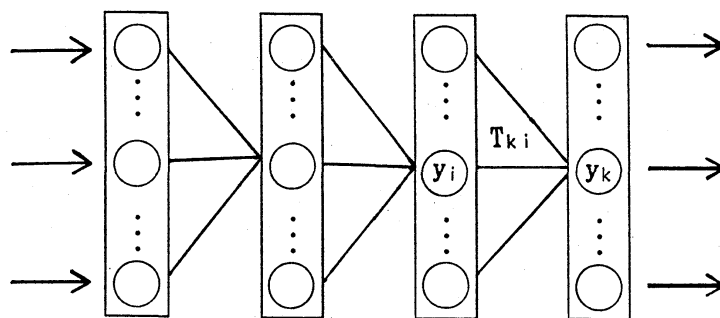


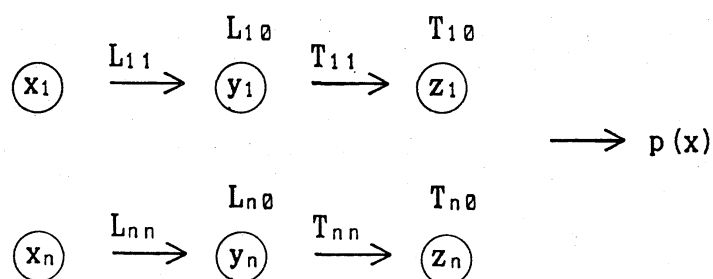
図 8 多層パーセプトロン

バックプロパゲーションは、任意の初期値に対して誤差が極小（最小ではない）となるということが保証されているが、問題点もいくつかある。学習速度を調整する ϵ の値が大きすぎると学習は振動してしまい、小さすぎると収束が非常に遅くなるので定めづらい。また中間層の素子数も、問題によって必要とする数が異なり、実験的に決めるしかない状態である。

4.2 認識における記憶学習

ここでは、表 1 で与えた実験データをもとにシュレーダーの階段図形の認識機械をつくり、その機械へいくつかの初期状態を入力した場合の認識結果（出力）を比較することにする。認識機械は、表 1 の実験値をもとに教師信号をつくり、上で述べた逆伝搬法によって作成する。

認識機械の作成法を説明しよう。ここでは図 9 のような n 入力 n 出力 2 段パーセプトロンを考えることにする。

図 9 n 入力 n 出力 2 段パーセプトロン

この機械は“パラメータ θ の図形を見る”という行為に相当し、脳のある初期状態を入力すると、 θ の図形を見た認識結果を出力するものとする。同じ機械を通過、つまり同じ図形を見たとしても、脳の初期状態によっては別の判断をす

る可能性があるということがその出力結果に現れると考えられるので、それを比較してみようというわけである。なお、そのような機械を作るということは、ひとつのパラメータ θ に対し、それぞれ結合係数や誤差関数を与える必要があるのである。のだが、簡単のため以下では $T_{ij}(\theta)$, $L_{ij}(\theta)$, $E(\theta)$, $\hat{p}(\theta)$ の θ は省略する。

入力層 x から中間層 y への結合係数を L_{ij} , 中間層 y から出力層 z への結合係数を T_{ij} とし、各素子の出力を次のように定める：

$$\left\{ \begin{array}{l} y_j(x) = \tanh(L_{j0} + L_{j1}x_1 + \cdots + L_{jn}x_n) \\ z_k(x) = \tanh(T_{k0} + T_{k1}y_1(x) + \cdots + T_{kn}y_n(x)) \\ p(x) = \frac{1}{2} \left(1 + \frac{z_1(x) + \cdots + z_n(x)}{n} \right) \end{array} \right.$$

$p(x)$ は "左が手前に見える確率" を意味し、式 (7) に $\mu = +1$ を代入したものになっている。誤差関数は

$$E(x) \equiv \frac{1}{2} \sum_{x \in \hat{X}} (p(x) - \hat{p})^2 \quad (20)$$

で定め、これを最小にするような T_{ij} , L_{ij} を逆伝搬法で学習させ、機械を作る。ここで、 $\hat{p}$ は教師信号、集合 $\hat{X}$ は $\hat{p}$ が与えられるような入力パターンの集合である。

各結合係数は、式 (18) に従って変化するものとし、それぞれ計算してみる。ただし、 T_{i0} , L_{i0} による微分も同時に考えられるように、 $x_0 = y_0(x) = 1$ とおく。

$$\frac{\partial E(x)}{\partial T_{ij}} = \sum_{x \in \hat{X}} (p(x) - \hat{p}) \frac{\partial p(x)}{\partial T_{ij}} \quad (i=1 \sim n, j=0 \sim n)$$

$$\begin{aligned} \frac{\partial p(x)}{\partial T_{ij}} &= \frac{1}{2n} \cdot \frac{\partial}{\partial T_{ij}} \{z_1(x) + \cdots + z_n(x)\} \\ &= \frac{1}{2n} (1 - z_i(x)^2) y_j(x) \end{aligned}$$

よって,

$$\frac{\partial E(x)}{\partial T_{ij}} = \frac{1}{2n} \sum_{x \in \hat{X}} (\hat{p}(x) - p) (1 - z_i(x)^2) y_j(x).$$

$$\begin{aligned} \frac{\partial E(x)}{\partial L_{ij}} &= \sum_{x \in \hat{X}} (\hat{p}(x) - p) \frac{\partial \hat{p}(x)}{\partial L_{ij}} \\ &= \frac{1}{2n} \sum_{x \in \hat{X}} (\hat{p}(x) - p) \frac{\partial}{\partial L_{ij}} \{z_1(x) + \dots + z_n(x)\} \end{aligned}$$

$$(i=0 \sim n, j=1 \sim n)$$

$$\begin{aligned} \frac{\partial z_m(x)}{\partial L_{ij}} &= (1 - z_m(x)^2) T_{mi} \frac{\partial y_i(x)}{\partial L_{ij}} \\ &= (1 - z_m(x)^2) T_{mi} (1 - y_i(x)^2) x_j \end{aligned}$$

よって,

$$\frac{\partial E(x)}{\partial L_{ij}} = \frac{1}{2n} \sum_{x \in \hat{X}} (\hat{p}(x) - p) (1 - y_i(x)^2) x_j \sum_{m=1}^n T_{mi} (1 - z_m(x)^2).$$

従って, T_{ij} , L_{ij} の学習則がそれぞれ

$$\left\{ \begin{array}{l} T_{ij}^{\text{new}} = T_{ij}^{\text{old}} - \frac{\epsilon}{2n} \sum_{x \in \hat{X}} (\hat{p}(x) - p) (1 - z_i(x)^2) y_j(x) \\ L_{ij}^{\text{new}} = L_{ij}^{\text{old}} - \frac{\epsilon}{2n} \sum_{x \in \hat{X}} (\hat{p}(x) - p) (1 - y_i(x)^2) x_j \sum_{m=1}^n T_{mi} (1 - z_m(x)^2) \end{array} \right. \quad (i=1 \sim n, j=0 \sim n)$$

と求められた。ただし $x_0 = y_0(x) = 1$ である。

教師信号としては、式 1 で表 1 のデータを近似した、左の壁が手前に見える

確率

$$\hat{p}(\theta) = \frac{\exp(4.271\eta + 0.11)}{1 + \exp(4.271\eta + 0.11)}$$

を用いる。ただし $\eta \equiv \log \theta$ である。ひとつの機械に対してひとつの $\hat{p}$ (厳密には $\hat{p}(\theta)$) を使うが、この値は入力に依存していない。言い替えれば被験者が初めて図形を見たときに判断した結果の値である。よってここでは誤差関数における集合 $\hat{X}$ は、 2^n 個の要素をもつ入力全ての集合 $X \equiv \{(x_1, \dots, x_n); x_i = \pm 1\}$ の部分集合

$$\hat{X} \equiv \{(x_1, \dots, x_n); x_i = \pm 1, \mid \sum_i x_i \mid \leq f_n\}$$

で定めることにする。条件 $\mid \sum_i x_i \mid \leq f_n$ は、初めて図形をみるということに対応して与えたもので、初期状態がかたよっていない入力パターンを指す。例えば $n=6$, $f_n=2$ なら

$$(+1, +1, +1, +1, +1, -1) \in \hat{X}$$

$$(-1, +1, +1, +1, -1, -1) \in \hat{X}$$

という具合である。

以上のような学習則に従って各パラメータに対する認識機械を作り、その機械を用いて 2 種類の実験を行ってみよう。実験はすべて $n=8$, $f_n=2$, $\epsilon=0.001$ で行う。なお、 $p_{1.0}(x)$ は、状態 x において $\theta=1.0$ の図形を見たときに左の壁が手前に見える確率を表すものとする。

実験 1 A. 直前に極端な図形を見てから、最も判断に迷うパラメータの図形を見たら、その認識結果には直前の影響が残るであろう。例えば明らかに左の壁が手前に見えやすい $\theta=1.67$ の図形 (図 2(b)) を見た後に左右同面積の図形 (図 2(a)) を見たら、左が手前に見えてしまう人が多いのではないかと考えられる。

実験は、左の壁を手前と判断しやすい図形を見た直後の脳の状態を

$$x_{+8} \equiv (+1, +1, \dots, +1)$$

と考へ、 $\theta = 1.0$ の認識機械にこれを初期状態として入力してみることにする。
また、逆にかたよった状態

$$x_{-8} \equiv (-1, -1, \dots, -1)$$

及び、ひとつだけ -1 を含む状態

$$x_{+6} \in X_{+6} \equiv \{(x_1, \dots, x_8) \mid \sum_{i=1}^8 x_i = 6\},$$

ひとつだけ $+1$ を含む状態

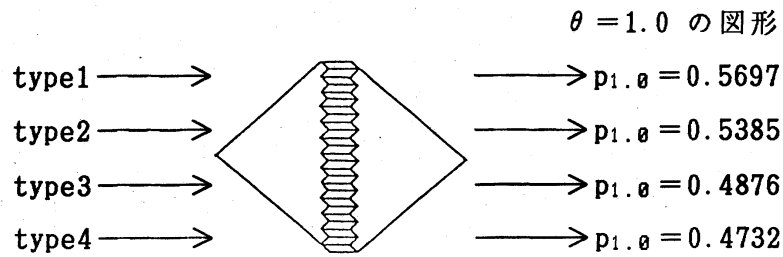
$$x_{-6} \in X_{-6} \equiv \{(x_1, \dots, x_8) \mid \sum_{i=1}^8 x_i = -6\}$$

についても認識結果を求め比較する。

B. また現実的に考えてみると、初期状態によってははじめは記憶学習効果が現れるが、その後何度か $\theta = 1.0$ の図形を見たら次第に迷いが生じてくることが予想される。そこで、認識した(パーセプトロンを通過した)素子の状態を再び入力するという操作を繰り返し、その度の認識結果を調べてみることにする。 x_{+8} と x_{+6} を X^+ 、 x_{-8} と x_{-6} を X^- とグループにして、各グループの要素をそれぞれ 7 回認識機械に通して、平均値を求めることにする。

結果は次のようになった。やはり初期状態の影響は明らかにみられ、記憶学習していることがわかる(A)。また、その後何度か $\theta = 1.0$ の図形を見続けると 2, 3 回でその認識結果は $1/2$ に近づく、つまり次第に考えにかたよりがなくなり迷いが生じて、手前に見える壁が左右同確率になってきたと考えられる。 $\theta = 3.0$ の図形においては、すぐに 1 に近い確率で左が手前に見えるといえる(B)。

A.



B.

回数	$p_{1.0}(x_{+1})$	$p_{1.0}(x_{-1})$
1	0.5419311	0.4859725
2	0.5050419	0.5101546
3	0.4957973	0.4942512
4	0.4940580	0.5008863
5	0.5055548	0.5036111
6	0.5121380	0.5043819
7	0.5065822	0.5134655

繰り返し $\theta = 1.0$ の図形を見た時の $p_{\theta}(x)$ の変化

実験 2 非常にかたよった初期状態 $x_{+8} = (+1, \dots, +1)$, $x_{-8} = (-1, \dots, -1)$ において、それぞれ最も判断しづらい $\theta = 1.0$ の図形を見たら、その認識結果 $p_{1.0}(x(1))$, $p_{1.0}(x(4))$ には差が生じることが実験 1 によって確かめられた。では、それよりも左の壁が手前に見えやすい $\theta = 1.3$ などの図形においては $p_{1.3}(x(1))$, $p_{1.3}(x(4))$ の差はどうなるであろうか。ここでは 2^8 通りのパターン x のうち、認識結果 $p_{\theta}(x)$ が最大値をとるものと最小値をとるものの差を δp_{θ} とし、 $\theta = 1.0, 1.3, 3.0$ の場合についてそれぞれ調べてみた。

θ	1.0	1.3	3.0
max	0.5697	0.8192	0.9869
min	0.4441	0.7213	0.9237
δp_{θ}	0.1255	0.0978	0.0632

θ による認識結果のばらつき方

上の結果から、予想どおり θ が大きい方が、認識結果のばらつきが小さいことがわかった。これは、たとえ初期状態がかたよっていても極端な図形に対しては記憶学習の効果が小さいことを意味する。

以上のように、“パラメータ θ の図形を見る”ことに相当する認識機械を逆伝搬法というニューラルネットの学習法によって構成し、脳（素子）の記憶学習効果を確認できた。実際我々の経験を考えてみても、認識過程は時間的に発展するダイナミクスのひとつであることは明らかであり、その過程を実現することは非常に興味深いといえよう。しかし、人間の記憶学習に作用する要素は無限にある。その中からどれを考慮し認識機械の構成に反映させるか、また、認識する図形のパラメータが複数個存在する場合はどのように合成させるかなど、残された課題は少なくない。

5.まとめ

本論文では、認識過程を数理的に取扱い、二義図形の錯視現象について相転移及び記憶学習という 2 つの面から考察した。

まずイジングモデルによって認識過程をモデル化し、カノニカル分布に従う系の状態を計算した。

物理学における相転移の考え方をシュレーダーの階段図形の認識に適用し、いかなるパラメータにおいて知覚に大きな変化が生じるかを、計算と実験によって調べることができた。また、直前の記憶が認識に影響を及ぼすことをシュミレーションによって確かめるために、逆伝搬法というニューラルネットの学習法を用いて認識機械を構成し、いくつかの脳の初期状態をその機械に入力した結果を考察した。記憶学習効果が個人の以前の経験などによって現れること、そしてそれは判断に最も迷うであろうパラメータの図形において一番顕著にわかることなどが明らかになった。

今後の課題としては、認識過程のモデル化をいかにすればより実際のそれに近

づけることができるか，ということが挙げられる．というのは，2，3のどちらの議論においてもその数理的基本としているのは1のモデルであり，この改善は直接影響を及ぼすからである．

文 献

- [1] 甘利俊一: "神経回路網の数理", 産業図書, 東京, 1978.
- [2] J. J. Hopfield: Newrons with graded response have collective computational properties like those of twostate neurons, Proc. Natl. Acad. Sci, USA 81, 3088-3092, May, 1984.
- [3] 中野馨 編: "ニューロコンピュータの基礎", コロナ社, 東京, 1990.
- [4] 大矢雅則: 錯視図形に見る相転移, 数理科学, NO. 301, 58-61, July, 1988.
- [5] 今井省吾: "錯視図形", サイエンス社, 1984.
- [6] 大矢雅則: "認識過程の数理と最適値問題", 朝倉書店, 準備中.
- [7] F. Rosenblatt: "principles of Neurodynamics", Spartan, 1961.