

多層ネットの解濃度とハードウェア化における スケーリング法に関する考察

五反田 博[†] 白土 浩[†]
井上 勝裕^{††} 熊丸 耕介^{††}

本論文では、入力パターンがアフィン変換の関係にある学習課題に対して、同一構造の多層ネットは、シグモイド関数の極性や値域幅の大きさによらず等しい濃度の解を持つことを明らかにする。この結果は、多層ネットをアナログハードウェア化する際のスケーリング問題に適用できる。入力パターンやシグモイド関数の値域が k 倍に拡大された場合、バイアスはそのままして荷重を $1/k$ 倍すれば、分離能力や汎化能力は保存される。また、荷重やバイアスの初期値を上記のように設定した場合、バイアス更新と荷重更新に対する学習係数をそれぞれ $1/k^2$ 倍と $1/k^4$ 倍にすれば、誤差逆伝搬 (BP) 学習の収束挙動は元のものと等価になる。さらに、通常のように、荷重とバイアスの初期値を同一分布の一樣乱数で設定し、双方の更新に対する学習係数を等しく与えて BP 学習させた場合、良好な収束を与える初期値は k の増大とともに小さくなることをシミュレーションにより示す。

Studies on Cardinality of Solutions for Multilayer Nets and a Scaling Method in Hardware Implementations

HIROMU GOTANDA,[†] HIROSHI SHIRATSUCHI,[†] KATSUHIRO INOUE^{††}
and KOUSUKE KUMAMARU^{††}

This paper elucidates that multilayer nets of equal structure allow the same cardinality of admissible solutions for learning tasks of which input patterns are related by Affine transform, even if their sigmoid functions are different in polarity and range. This result can be applied to a scaling problem arising on building the nets in analog hardware. In the case of the input patterns and the sigmoid functions multiplied by a scaling factor k , the separation and generalization abilities can be preserved if the weight values are set to $1/k$ times the original ones with keeping the bias values intact. With such initial weights and biases as in the above, the converging behavior of the back propagation (BP) learning in the scaled environment becomes equivalent to that in the original environment, provided that the learning coefficients for bias and weight update are multiplied by $1/k^2$ and $1/k^4$, respectively. Under the BP learning in an ordinary way with initializing both weights and biases by uniform random numbers of identical distribution and adopting the same learning coefficient for both weight and bias update, it is shown by simulations that the initial values resulting in good convergence decrease with increasing k .

1. ま え が き

多層ネットを用いた誤差逆伝搬 (BP) 学習は、パターン認識や制御等の多くの分野で有用性が認められているが、ローカルミニマムへの落ち込みや学習速度の遅さ等の収束問題がある。そのため、ネット構造 (層数やユニット数) や初期値等を適切に定める試み

が多数検討されているが、BP 学習の適用については、依然、試行錯誤の性格を残している。

その中で、ユニット極性の選択については、シグモイド関数が 0 から 1 の値をとる (片極型) ユニットよりも、 -0.5 から 0.5 のように正負の値をとる (両極型) ユニットで構成される (両極型) ネットの方が良好な収束を示すことが報告されている^{1)~4)}。これに関して、著者らは、賣らによる初期値に関する検討結果⁵⁾と Watanabe らによる解に関する検討結果²⁾をもとに、両極型ネットによる収束の片極型ネットに対する優位性を裏付けて、極性による収束の差異はネット規模の増大とともに顕著に開くことを指摘した⁶⁾。ま

[†] 近畿大学九州工学部

Faculty of Engineering, Kinki University in Kyushu

^{††} 九州工業大学情報工学部

Faculty of Computer Science and Systems Engineering,
Kyushu Institute of Technology

た、多層ネットが訓練データの規定する入出力関係を満たす解（荷重とバイアスの組）を持つための必要条件を法ベクトルの集合で定式化し、この必要条件がユニット極性により大きく異なることを明らかにした⁷⁾。

両極型ユニットとしては、一般に、出力範囲が -1 から 1 のものが使用されている。このときの入力空間は $(-1, 1)^m$ と一辺の長さが 2 の超立方体となって、その体積は 2^m と入力次元 m とともに増大する。一方、出力範囲が -0.5 から 0.5 の場合、入力空間は $(-0.5, 0.5)^m$ と単位超立方体となって、その体積は m の大きさに関係なく 1 となる。必要条件を満たす法ベクトルの集合は入力空間を取り囲むように存在するから、入力空間が拡大すれば、探索範囲が広がることになる。しかし、入力空間の大きさの相違が多層ネットの許容解（すべての入力パターンに対してネット出力と教師との誤差がある許容範囲に入るときの荷重とバイアスの組で、実際に BP 学習させて得られる収束解はこれに含まれる）の個数（濃度）や初期値の設定にどのように影響するかは定かでない。

また、ニューラルネットはハードウェア化されてその真価が発揮されることから、VLSI 技術や光技術に基づいて、荷重やシグモイド関数、積和計算を実装したハードウェアが多数試作されている^{8)~10)}。この場合、シグモイド関数は、実装方式に依存して、固有の極性や値域幅をとる。そして、入力パターンは電気ないし光信号に変換されて提示されることになるが、許容される信号の大きさは回路的に制限される。同様な制限が荷重やバイアスの大きさについても存在する。したがって、計算機上で模擬的に構築された学習済みの多層ネットをハードウェア化する場合、実装されるシグモイド関数の特性との整合性を考慮しながら、入力パターン、荷重、バイアスをスケールリングする必要がある。さらに、学習可能なハードウェアの場合¹¹⁾、これまで計算機シミュレーションを通して蓄積されてきた初期値や学習係数に関する経験値や知識をそこでの学習に活用することが重要となる。しかし、入力空間が拡大縮小されたときのスケールリング法や BP 学習の収束挙動に関する議論は見あたらない。

以上の観点から、本論文では、まず、入力パターンがアフィン変換の関係にある学習課題に対して、層数や対応する層のユニット数が同じ多層ネットは、シグモイド関数の極性や値域幅の大きさによらず、等しい濃度の許容解を持つことを示す。次に、この結果を利用して、両極型ネットを対象に、入力パターンやシグモイド関数の値域が k 倍された場合、つまり入力空間が k 倍に拡大された場合、バイアスを同じにして、荷

重を $1/k$ 倍すれば、分離能力や汎化能力を等価に保存できることを指摘する。さらに、荷重やバイアスの初期値を上記のように設定した場合、バイアス更新に対する学習係数を $1/k^2$ 倍、また荷重更新に対する学習係数を $1/k^4$ 倍にすれば、収束挙動は等価になることを述べる。最後に、通常のように、荷重とバイアスの初期値を同一分布幅の一樣乱数で設定し、双方の更新に対する学習係数を等しく与えて学習させた場合、良好な収束を与える初期値は入力空間の拡大とともに小さくなることをシミュレーションにより示す。

2. 準備

2.1 片極型と両極型ネット

バイアスが b 、下位層ユニットとの結合荷重が w_i ($i = 1, 2, \dots, m$)、入出力特性が

$$u = \sum_{i=1}^m w_i x_i + b \quad (1)$$

$$y = f(u) \quad (2)$$

のように与えられるユニットで構成される多層ネットを考える。ここに、 x_i は下位層ユニットからの入力である。そして、シグモイド関数 $f(\cdot)$ が

$$f_U(u) = \frac{1}{1 + e^{-u}} \quad (3)$$

$$f_B(u) = \frac{1}{1 + e^{-u}} - \frac{1}{2} = \frac{1}{2} \tanh\left(\frac{u}{2}\right) \quad (4)$$

で定義されるユニットをそれぞれ片極型ユニットと両極型ユニットと呼ぶことにする^{3), 6)}。そして、多層ネットが前者からなる場合を片極型ネット、後者からなる場合を両極型ネットと呼び、両者は、同一構造、つまり層数や対応する層のユニット数は同じと仮定する。また、以降では、片極型と両極型ネットに付随する変数や集合をそれぞれ添字の U と B で区別する。

多層ネットの場合、ある層の出力はそのまま次の層への入力となるから、入力空間 $\{\mathbf{x} = [x_1, x_2, \dots, x_m]^T\}$ は、片極型ネットのとき $X_U = (0, 1)^m$ 、両極型ネットのとき $X_B = (-0.5, 0.5)^m$ となる。ここに、 T はベクトルの転置を表す。買らは、片極型ネットを対象に、入力空間 X_U を活性領域 $\{\mathbf{x} \mid \rho < f_U(u_j) < 1 - \rho\}$ と飽和領域 $\{\mathbf{x} \mid f_U(u) \leq \rho, f_U(u) \geq 1 - \rho\}$ に分け、活性領域の幅が、

$$g = \frac{2 \ln\{(1 - \rho)/\rho\}}{\|\mathbf{w}\|} \quad (5)$$

となることや、ユークリッド空間 R^m の原点から分離超平面 $H = \{\mathbf{x} \mid \mathbf{w}^T \mathbf{x} + b = 0\}$ までの距離が $|b|/\|\mathbf{w}\|$ となることを明らかにしている⁵⁾。ここに、

$\mathbf{w} = [w_1, w_2, \dots, w_m]^T$ で, $\|\cdot\|$ はユークリッドノルム, $|\cdot|$ は絶対値を表し, ρ は活性領域を規定する 0.1 程度の小さな正数である. 一方, 両極型ネットの場合, $\mathbf{w}$ や b が片極型ネットのものと同しければ, 分離超平面や活性領域 ($\{\mathbf{x} \mid -0.5 + \rho < f_B(\mathbf{u}) < 0.5 - \rho\}$) は, ユークリッド空間 R^m で片極型ネットのときと同じ領域を占める. しかし, 入力空間 X_B と X_U の R^m での位置づけが異なるため, 双方のネットでは荷重やバイアスを等しく与えたとしても, 分離超平面や活性領域による入力空間の分割は違ったものとなる^{3), 6)}.

2.2 入力空間の幾何学的に等価な分割

ユークリッド空間 R^m で, 原点から伸びて分離超平面 H と直交するベクトルは, 買らの結果⁵⁾より,

$$\mathbf{h} = -\frac{\mathbf{b}}{\|\mathbf{w}\|^2} \mathbf{w} \quad (6)$$

と記述されることが導かれる³⁾. 以下では, これを法ベクトルと呼んで, 両極型ユニットと片極型ユニットの分離超平面 H_B と H_U に対する法ベクトルをそれぞれ $\mathbf{h}_B$ と $\mathbf{h}_U$ と記すことにする. このとき, H_B による入力空間 X_B の分割と, H_U による X_U の分割が幾何学的に等価ならば,

$$\mathbf{h}_U = \left(1 + \frac{\mathbf{1}^T \mathbf{h}_B}{2\|\mathbf{h}_B\|^2}\right) \mathbf{h}_B \quad (7)$$

なる関係が得られる^{7), 12)}. ここに, $\mathbf{1}$ は $[1, 1, \dots, 1]^T$ のように要素がすべて 1 の m 次元ベクトルである. さらに, 式 (7) と活性領域幅の定義 (5) から, 両極型ユニットの荷重とバイアスをそれぞれ $\mathbf{w}_B (= [w_{B1}, w_{B2}, \dots, w_{Bm}]^T)$ および b_B と表記するとき, 片極型ユニットの荷重とバイアスをそれぞれ $\mathbf{w}_U = \mathbf{w}_B$ および $b_U = b_B - \mathbf{1}^T \mathbf{w}_B / 2$ と与えれば, 分離超平面のみならず活性領域から見た入力空間の分割も幾何学的に等価となる¹²⁾.

3. アフィン変換パターンに対する許容解の濃度

入力パターンと教師がそれぞれ $\mathbf{p}$ と $\mathbf{d}$ で与えられる訓練データの集合を $\{(\mathbf{p}, \mathbf{d})\}$ のように表す. また, その課題に対して, シグモイド関数が $f(u)$ で与えられる多層ネットによる許容解の集合を $\Psi(\mathbf{p}, \mathbf{d}, f)$ と表記して解集合と呼ぶことにする. このとき, 両極型および片極型ネットの解集合はそれぞれ $\Psi_B(\mathbf{p}, \mathbf{d}_B, f_B)$ および $\Psi_U(\mathbf{q}, \mathbf{d}_U, f_U)$ のように表される. ここに, 教師 $\mathbf{d}_B$ の要素は -0.5 から 0.5 , $\mathbf{d}_U$ の要素は 0 から 1 の範囲にあるものとする. この場合, すでに報告したように, 入力パターンと教師のそれぞれが, $\mathbf{q} = \mathbf{p} + \mathbf{1}/2$ および $\mathbf{d}_U = \mathbf{d}_B + \mathbf{1}/2$ と平行移動の関係にあれば,

2 つの解集合は $\Psi_B(\mathbf{p}, \mathbf{d}_B, f_B) \iff \Psi_U(\mathbf{q}, \mathbf{d}_U, f_U)$, つまり 1 対 1 に対応して, 2 つのネットは同じ濃度の解集合を持つ¹²⁾.

上記は, 訓練データが平行移動の関係にある課題は本質的に等価であり, この等価な学習課題に対して, 同一構造の多層ネットは, シグモイド関数の値域幅が同じなら極性によらず, 同じ濃度の解集合を持つことを意味している. これからの発展で, 次のことが類推される. すなわち, 入力パターンが $\{\mathbf{p}\}$ で与えられる課題と, 入力パターンが $\{\mathbf{p}\}$ をアフィン変換して

$$\mathbf{q} = A\mathbf{p} + \mathbf{a} \quad (8)$$

と与えられる課題は等価であり, これらの課題に対して, 同一構造の多層ネットは, シグモイド関数の極性や値域幅によらず, 同じ濃度の解集合を持つと考えられる. ここに, A は正則な変換行列, $\mathbf{a}$ は定数ベクトルである. 以下では, このことを証明する.

3.1 ユニット極性が異なる場合

訓練データ $\{(\mathbf{p}, \mathbf{d}_B)\}$ に対する両極型ネットの解集合 $\Psi_B(\mathbf{p}, \mathbf{d}_B, f_B)$ と, 入力パターンが式 (8) で与えられる $\{(\mathbf{q}, \mathbf{d}_U)\}$ に対する片極型ネットの解集合 $\Psi_U(\mathbf{q}, \mathbf{d}_U, f_U)$ を比較する. この場合, 教師が $\mathbf{d}_U = \mathbf{d}_B + \mathbf{1}/2$ なる関係にあれば, $\Psi_B(\mathbf{p}, \mathbf{d}_B, f_B) \iff \Psi_U(\mathbf{q}, \mathbf{d}_U, f_U)$ となって, 双方のネットによる解集合は同じ濃度になることが以下のように示される. 図 1 は, 議論の流れを明らかにするため, 3 層の場合を例に, 前向き信号が等価な両極型ネットと片極型ネットの関係を概念的に示したものである.

まず, 双方で第 1 中間層ユニットの荷重とバイアスが

$$\mathbf{w}_U = (A^T)^{-1} \mathbf{w}_B \quad (9)$$

$$b_U = b_B - \mathbf{w}_B^T A^{-1} \mathbf{a} \quad (10)$$

の関係にあれば, 総入力 u_U と u_B は $\mathbf{w}_U^T \mathbf{q} + b_U = \mathbf{w}_B^T \mathbf{p} + b_B$ のように等しくなることが導かれる. そして, この場合, $\mathbf{p}$ と $\mathbf{q}$ に対する第 1 中間層ユニットの出力をそれぞれ y_B と y_U とすると, シグモイド関数の定義 (3), (4) から,

$$y_U = y_B + 1/2 \quad (11)$$

となって, 片極型ユニットの出力は, 両極型に比べて $1/2$ 大きくなることが分かる.

次に, 第 2 中間層の場合, 式 (11) より, y_B を要素とするベクトル $\mathbf{y}_B$ を両極型ユニットへの入力とすると, これを $1/2$ 平行移動した $\mathbf{y}_U = \mathbf{y}_B + \mathbf{1}/2$ が片極型ユニットへ入力されることになる. この場合, 式 (8) の変換行列と定数ベクトルがそれぞれ $A = I$ (単位行列) と $\mathbf{a} = \mathbf{1}/2$ となることを考慮して, 荷重とバイアスを式 (9), (10) より $\mathbf{w}_U = \mathbf{w}_B, b_U = b_B - \mathbf{1}^T \mathbf{w}_B / 2$

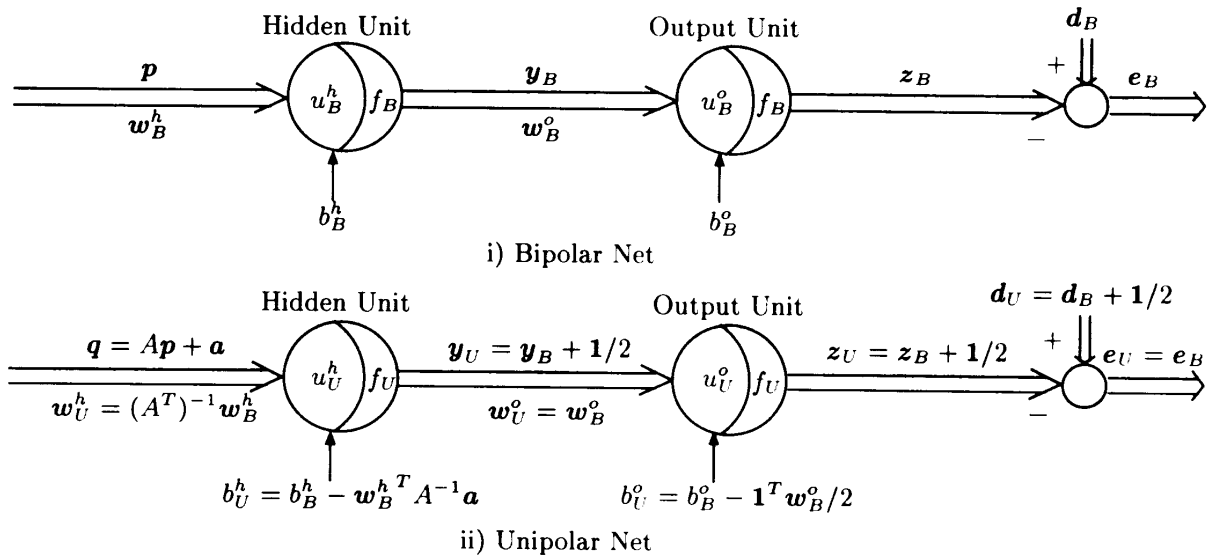


図1 前向き信号の等価な両極型と片極型ネットの概要図 (w_B^h や w_B^o 等の表記における上付添字 h と o はそれぞれ中間および出力層ユニットを指している.)

Fig.1 A schematic diagram of bipolar and unipolar nets equivalent in feedforward signals (In the notations such as w_B^h and w_B^o , the superscript h and o denote the hidden and output units, respectively).

と与えれば、第2中間層ユニットへの総入力も双方で等しいことが導かれる。同様のことが上位層ユニットについても繰り返される。したがって、両極型ネットからの出力を z_B とすると、片極型ネットからの出力は $z_U = z_B + 1/2$ となる。

以上のことは、双方のネットにおける前向き信号の伝達特性が等価であることを示唆している。同時に、教師が $d_U = d_B + 1/2$ なる関係にある場合、 $\{(p, d_B)\}$ に対する両極型ネットの誤差特性と、 $\{(q, d_U)\}$ に対する片極型ネットの誤差特性は一致することを示している。したがって、双方のネットによる解集合は、 $\Psi_B(p, d_B, f_B) \iff \Psi_U(q, d_U, f_U)$ のように1対1に対応して、同じ濃度になることが分かる。

また、式(11)の場合、式(8)の関係にある p' と q' に対するユニット出力をそれぞれ y'_B と y'_U とすると、

$$y'_U - y_U = y'_B - y_B \quad (12)$$

となって、ユニット出力から見た入力パターンに対する分離能力は双方で等しくなる。さらに、 p' と q' に対するネット出力をそれぞれ z'_B と z'_U とすると、 $z'_U - z_U = z'_B - z_B$ となって、ネット出力から見た分離能力も等しくなることが分かる。言い換えると、2つのパターン集合 $\{p\}$ と $\{q\}$ がアフィン変換の関係にあり、変換行列 A と定数ベクトル a が既知ならば、荷重とバイアスを式(9)、(10)のように与えることにより、 $\{p\}$ に対する分離能力と $\{q\}$ に対する分離能力を等価にする多層ネットが構築できる。ただし、

この場合、双方のネットで、法ベクトルや活性領域幅は異なり、 R^m における分離超平面や活性領域の配置も異なることに注意されたい。また、以下の場合についても、比較対象のネット間で、前向き信号の伝達特性や入力パターンに対する分離能力の等価性が上述と同様に導かれることに注意されたい。

3.2 ユニット極性が同一の場合

シグモイド関数が同じく式(4)で与えられる2つの両極型ネットの解集合 $\Psi_B(p, d_B, f_B)$ と $\Psi_B(r, d_B, f_B)$ を比較する。ここに、 r は p と式(8)のようにアフィン変換の関係にあるものとする。この場合、第1中間層の荷重とバイアスを式(9)、(10)のように与えれば、双方の総入力は同じになって、ユニット出力も同じになる。したがって、第2中間層以降の荷重とバイアスが双方で等しければ、前向き信号の伝達特性は一致する。そして、教師も等しいから、双方の誤差特性は一致して、解集合 $\Psi_B(p, d_B, f_B)$ と $\Psi_B(r, d_B, f_B)$ は1対1に対応することになる。また片極型ネットについても、 $\Psi_U(q, d_U, f_U)$ と $\Psi_U(s, d_U, f_U)$ の1対1対応が同様に導かれる。ここに、 s は q とアフィン変換の関係にあるものとする。

3.3 ユニットの出力幅が異なる場合

シグモイド関数が k 倍されて $kf_B(\cdot)$ と与えられる両極型ネットの荷重とバイアスをそれぞれ w_B^* と b_B^* で表す。そして、このネットに入力パターンが同じで教師のみが k 倍された訓練データ $\{(p, kd_B)\}$ を提示し

たときの解集合 $\Psi_B(\mathbf{p}, k\mathbf{d}_B, k\mathbf{f}_B)$ と, $\Psi_B(\mathbf{p}, \mathbf{d}_B, \mathbf{f}_B)$ を比較する. この場合, 式 (8) の A と $\mathbf{a}$ はそれぞれ単位行列と零ベクトルになるから, 式 (9), (10) より, 第1中間層ユニットが $\mathbf{w}_B^* = \mathbf{w}_B$, $b_B^* = b_B$ なる関係であれば, その出力は $\mathbf{y}_B^* = k\mathbf{y}_B$ となる. 第2中間層については, $A = kI$, $\mathbf{a} = \mathbf{0}$ を式 (9), (10) に代入して, $\mathbf{w}_B^* = \mathbf{w}_B/k$, $b_B^* = b_B$ となることから, 総入力は双方で等しく, 出力は k 倍されることが分かる. 以下, 同様な手順により, ネット出力が $\mathbf{z}_B^* = k\mathbf{z}_B$ と k 倍されることが導かれる. したがって, 誤差が k 倍されるという違いを許せば, $\Psi_B(\mathbf{p}, \mathbf{d}_B, \mathbf{f}_B)$ と $\Psi_B(\mathbf{p}, k\mathbf{d}_B, k\mathbf{f}_B)$ は1対1に対応することになる. 同様に, 片極型ネットについても, $\Psi_U(\mathbf{q}, \mathbf{d}_U, \mathbf{f}_U)$ と $\Psi_U(\mathbf{q}, k\mathbf{d}_U, k\mathbf{f}_U)$ の1対1対応が導かれる.

さらに, 上述と3.2節の結果を両極型ネットの解集合 $\Psi_B(\mathbf{p}, \mathbf{d}_B, \mathbf{f}_B)$ と $\Psi_B(\mathbf{r}, k\mathbf{d}_B, k\mathbf{f}_B)$ に適用すれば, これらの要素も1対1に対応することが導かれる. また, 片極型ネットについても, これと同様に, 解集合 $\Psi_U(\mathbf{q}, \mathbf{d}_U, \mathbf{f}_U)$ と $\Psi_U(\mathbf{s}, k\mathbf{d}_U, k\mathbf{f}_U)$ の要素も1対1に対応することが導かれる. したがって, 同一極性のネットは, 構造が同じなら, アフィン変換された入力パターンに対し, シグモイド関数の値域幅によらず, 同じ濃度の解集合を持つことが分かる.

3.4 ユニットの極性と出力幅が異なる場合

シグモイド関数の極性や値域幅それに訓練データの観点から, 以上の対応関係をまとめると図2のようになる. (図中の“ $\iff$ ”は1対1対応を示し, 付記の数字はそれを証明した節を表す.) そして, 図の対応関係から, たとえば $\Psi_B(\mathbf{p}, \mathbf{d}_B, \mathbf{f}_B) \iff \Psi_U(\mathbf{s}, k\mathbf{d}_U, k\mathbf{f}_U)$ が導かれて, ユニットの極性と出力幅がともに異なる場合でも解集合の濃度は等しくなることが分かる.

以上のことから, 多層ネットは, 同一構造, すなわち各層でのユニット個数が同じなら, シグモイド関数の極性や値域幅に関係なく, 等価な学習課題に対して同じ濃度の解集合を持つことが結論づけられる.

4. スケーリングに関する考察

計算機上で学習済みの多層ネットをアナログ回路でハードウェア化する場合, 実現可能なシグモイド特性に準拠して, 入力パターンおよび荷重やバイアスをスケーリングする必要がある. このスケーリングはすべての層の入力空間にわたるから, 各層の荷重やバイアスの値はそれに応じて変換されなければならない. ここでは, まず, このスケーリング問題に前述の結果を適用して, 両極型ネットを対象に, 入力パターンやシグモイド関数の値域幅が拡大縮小された場合でも, パ

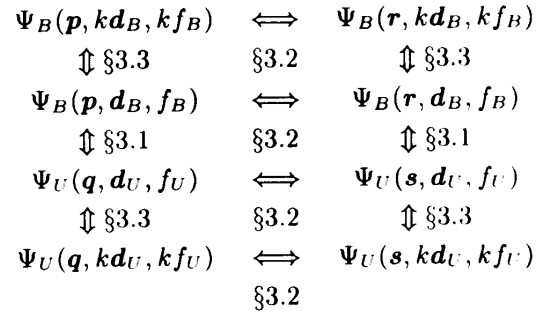


図2 解集合間の1対1対応

Fig. 2 One-to-one correspondence between solution sets.

ターンの分離能力や汎化能力を等価に保存するネットが構築できることを述べる.

また, BP学習の収束能力は, 学習係数や慣性係数の値に影響されるが, 適切な値は学習課題に依存するため, その設定は経験に委ねざるをえない現状にある. そのため, 良好な学習過程をハードウェアで実現するには, これまでソフトウェア的に蓄積されてきた学習係数や慣性係数に関する経験値をいかに活用するかが重要となる. したがって, 次に, 入力パターンやシグモイド特性が拡大縮小されたときのBP学習について, 学習係数や慣性係数をどのように与えれば, 収束挙動が等価になるかを議論する.

以上の2点を議論するにあたり, 定義や仮定を次のように設ける. すなわち, 計算機上で, 式 (4) の $f_B(\cdot)$ をシグモイド関数として, 訓練データ $\{(\mathbf{p}, \mathbf{d}_B)\}$ によりすでにBP学習済みの両極型ネットを $\mathcal{N}$, またこれをハードウェアとして実現すべきネットを $\mathcal{N}^*$ と表す. さらに, $\mathcal{N}^*$ は $\mathcal{N}$ と同一構造として, $\mathcal{N}^*$ のシグモイド特性と訓練データをそれぞれ

$$f_B^*(u^*) = kf_B(u) \quad (13)$$

$$\{(\mathbf{p}^*, \mathbf{d}_B^*)\} = \{(k'\mathbf{p}, k\mathbf{d}_B)\} \quad (14)$$

のように定義する. この場合, k' と k は同じとしても議論の本質は失われない. したがって, 以下では, 便宜のため, $k' = k$ と仮定し, k を拡大率と呼ぶことにする. このとき, $\mathcal{N}^*$ の入力空間は, $\mathcal{N}$ に比べて, すべての層で k 倍に拡大されることになる.

4.1 分離および汎化能力の等価なネット

入力パターン $\mathbf{p}^*$ と $\mathbf{p}$ は式 (8) において $A = kI$, $\mathbf{a} = \mathbf{0}$ の関係にある. したがって, $\mathcal{N}$ の荷重とバイアスをそれぞれ $\mathbf{w}$ と b とするとき, 式 (9), (10) に準じて, $\mathcal{N}^*$ の荷重 $\mathbf{w}^*$ とバイアス b^* をそれぞれ

$$\mathbf{w}^* = \mathbf{w}/k \quad (15)$$

$$b^* = b \quad (16)$$

と与えれば,

$$\mathbf{y}^* = k\mathbf{y} \quad (17)$$

となることが導かれる。ここに、 $\mathbf{y}$ と $\mathbf{y}^*$ はそれぞれ $\mathcal{N}$ と $\mathcal{N}^*$ の対応する層のユニット出力である。したがって、 $\mathcal{N}^*$ の各層に対する入力、 $\mathcal{N}$ の対応する層の入力に比べて、 k 倍されることになる。これは、2つのネットの前向き信号伝達特性が等価であることを意味している。

一方、荷重とバイアスが式 (15), (16) で与えられるとき、式 (5), (6) から、法ベクトルや活性領域幅も

$$\mathbf{h}^* = k\mathbf{h} \quad (18)$$

$$g^* = kg \quad (19)$$

と k 倍されることが分かる。このことは、双方の入力空間が法ベクトルや活性領域により幾何学的に相似な形で分割されることを意味する、と同時に入力パターンに対する分離能力が等価となることを意味している。

そして、 $\mathcal{N}$ と $\mathcal{N}^*$ において、ネット出力と教師の誤差をそれぞれ e および e^* と記すとき、

$$e^* = ke \quad (20)$$

となって、両者による誤差特性は等価になることが導かれる。このことは、入力パターンが k 倍された場合、教師も k 倍して、荷重とバイアスを式 (15), (16) のように与えれば、 $\mathcal{N}^*$ と $\mathcal{N}$ の汎化能力が等価になることを示唆している。

4.2 等価な収束挙動を与える学習係数

入力空間が式 (13), (14) のように拡大されるとき、荷重やバイアスの初期値を式 (15), (16) の関係で与え、訓練データを $\mathcal{N}$ と $\mathcal{N}^*$ に同一順序で提示して学習させる場合を考える。そして、 $\mathcal{N}^*$ の学習係数や慣性係数を $\mathcal{N}$ のものとどのような関係で与えれば、両者の収束挙動が等価になるかを考察する。

BP 学習において、 $\mathcal{N}$ の出力層ユニットの荷重 $\mathbf{w}$ とバイアス b は、逆伝搬誤差 $\delta = f'_B(\cdot)e$ をもとに、

$$\Delta\mathbf{w} = \eta\delta\mathbf{x} + \alpha\Delta\mathbf{w}(\text{old}) \quad (21)$$

$$\Delta b = \mu\delta + \alpha\Delta b(\text{old}) \quad (22)$$

と更新される。ここに、 $f'_B(\cdot)$ は $f_B(\cdot)$ の微分、 η と μ はそれぞれ荷重とバイアス更新の学習係数、 α は慣性係数、 $\mathbf{x}$ は出力層ユニットへの入力ベクトルである。また、 $\mathcal{N}^*$ の出力層ユニットの荷重 $\mathbf{w}^*$ とバイアス b^* は、逆伝搬誤差 $\delta^* = f'^*_B(\cdot)e^*$ をもとに、

$$\Delta\mathbf{w}^* = \eta^*\delta^*\mathbf{x}^* + \alpha^*\Delta\mathbf{w}^*(\text{old}) \quad (23)$$

$$\Delta b^* = \mu^*\delta^* + \alpha^*\Delta b^*(\text{old}) \quad (24)$$

と更新される。ここに、 η^* 、 μ^* 、 α^* 、 $\mathbf{x}^*$ は $\mathcal{N}^*$ に対して上記と同様に定義される量である。

荷重とバイアスの初期値をそれぞれ式 (15), (16) の関係で与えた場合、前節で述べたように、 $\mathcal{N}$ と $\mathcal{N}^*$ の前向き信号伝達は式 (17) のように等価となり、誤

差特性も式 (20) のように等価になる。したがって、式 (20) と式 (13) から、 $\mathcal{N}$ と $\mathcal{N}^*$ の逆伝搬誤差について、

$$\delta^* = k^2\delta \quad (25)$$

なる関係が得られる。この関係と式 (17) を式 (21) から式 (24) に適用すると、最初の提示パターンに対して、 $\Delta\mathbf{w}(\text{old})$ 、 $\Delta b(\text{old})$ 、 $\Delta\mathbf{w}^*(\text{old})$ 、 $\Delta b^*(\text{old})$ はゼロとなることから、次の関係が導かれる。すなわち、荷重更新とバイアス更新の学習係数をそれぞれ

$$\eta^* = \eta/k^4 \quad (26)$$

$$\mu^* = \mu/k^2 \quad (27)$$

と設定すれば、荷重とバイアスの変化分について、

$$\Delta\mathbf{w}^* = \Delta\mathbf{w}/k \quad (28)$$

$$\Delta b^* = \Delta b \quad (29)$$

なる関係が得られる。式 (28), (29) は、 $\mathcal{N}$ と $\mathcal{N}^*$ の出力ユニットの荷重とバイアスが、更新後も、更新前の式 (15), (16) と同じ関係を保存していることを意味している。同様に、中間層ユニットについても、更新後の荷重やバイアスが式 (15), (16) の関係を保存することが導かれる。したがって、最初の提示パターンによる更新後も、 $\mathcal{N}$ と $\mathcal{N}^*$ の前向き信号伝達特性は等価となり、式 (17) が保証されることになる。

2 番目の提示パターンについては、 $\Delta\mathbf{w}^*(\text{old}) = \Delta\mathbf{w}(\text{old})/k$ 、 $\Delta b^*(\text{old}) = \Delta b(\text{old})$ なる関係のもとで、式 (21) から式 (24) のように更新される。この場合、慣性係数を $\mathcal{N}$ と $\mathcal{N}^*$ の間で

$$\alpha^* = \alpha \quad (30)$$

と等しく設定すれば、出力ユニットの荷重の変化分について、 $\Delta\mathbf{w}^* = \eta^*\delta^*\mathbf{x}^* + \alpha^*\Delta\mathbf{w}^*(\text{old}) = (\eta\delta\mathbf{x} + \alpha\Delta\mathbf{w}(\text{old}))/k = \Delta\mathbf{w}/k$ のように式 (28) が成立し、バイアスの変化分についても式 (29) が成立する。同様に、中間層ユニットの荷重やバイアスの変化分についても式 (28), (29) が成立することが導かれる。

上述の議論を以降の提示パターンについても繰り返すことにより、次の結論に達する。すなわち、 $\mathcal{N}$ と $\mathcal{N}^*$ において、荷重とバイアスの初期値をそれぞれ $\mathbf{w}^* = \mathbf{w}/k$ と $b^* = b$ の関係で与えて、学習係数を $\eta^* = \eta/k^4$ および $\mu^* = \mu/k^2$ 、また慣性係数を $\alpha^* = \alpha$ と設定した場合、2つのネットによる BP 学習の収束挙動は等価になる。

4.3 バイアスの駆動入力を考慮する場合

通常、ユニットへの総入力は式 (1) のように表現される。これはバイアスの駆動入力を 1 なる固定値と見なすことにはかならない。一方、ハードウェア化においては、荷重とバイアスは区別されることなく同一方式で実現されることが多い^{8)~11)}。この場合、式 (1) を $\mathbf{w}^T\mathbf{p} + b = (\mathbf{w}^T/k)(k\mathbf{p}) + (b/k)k$ と変形して分かるよ

うに、バイアスの駆動入力を k なる固定入力として扱う必要がある。その際、分離能力や汎化能力を等価にする荷重やバイアスは、 $w^* = w/k$, $b^* = b/k$ と同じく $1/k$ 倍にスケールすればよいことが導かれる。また、等価な収束挙動については、荷重およびバイアス更新の学習係数を $\eta^* = \eta/k^4$ および $\mu^* = \mu/k^4$ と同じく $1/k^4$ 倍すればよいことが導かれる。

5. スケール法と適切な初期値

前章では、入力空間がスケールされた場合、初期値を式 (15), (16) のように荷重とバイアスで区別して設定し、更新アルゴリズムについても、学習係数を荷重とバイアスで式 (26), (27) のように区別して与えれば、収束挙動は等価になることを述べた。しかし、現状では初期値や学習係数の設定法が確立されていないため、通常、初期値は荷重とバイアスとで区別されることなく同じ分布幅の一樣乱数で与えられ、学習係数についても同じ値が用いられる。

このように通常の形態で BP 学習させる場合、良好な収束を与える初期値の大きさには適切な範囲が存在することが知られている^{5),6)}。しかし、その範囲は学習課題に依存して特定が容易でないため、一般に、初期値としては小さな乱数を採用することが多い。これに関して、入力空間が拡大された場合、前章までの議論から、少なくとも荷重の初期値はさらに小さい方が収束は良好になると考えられるが、詳細は定かでない。したがって、ここでは、通常のように、荷重とバイアスの初期値を同一分布幅 $(-W, W)$ の一樣乱数で設定し、荷重更新とバイアス更新の学習係数を等しく与えて学習させた場合、良好な収束を与える初期値の大きさ W が入力空間のスケールにどのように影響されるかをシミュレーションにより調べる。

シミュレーションは、8 次のパリティ問題を対象に、8 - 30 - 1 の両極型ネット⁷⁾で、学習係数を $\eta = 0.5$ 、慣性係数を $\alpha = 0.9$ として、 $W = 0.3, 0.5, 1.0, 1.5, 2.0, \dots, 5.0$ の場合について行った。その際、教師は $\pm 0.4k$ として与え、収束判定基準は 256 個の入力パターンのすべてに対して誤差の絶対値が $0.1k$ 未満に収まることとした。図 3 は、拡大率 k を 1.0 (○), 3.0 (△), 5.0 (□), 10.0 (×) 倍としたときの収束率 (Ratio: 50 回の試行における収束回数⁸⁾の比) と収束したときの平均学習回数 (Cycle) の結果である。図より、 $k = 3$ のとき $1 \leq W \leq 2$ 、また $k = 5$ のとき $0.5 \leq W \leq 1.5$ なる範囲の初期値に対して良好な収束が得られていることが読みとれる。また、その他の拡大率についても、 k が大きくなるに

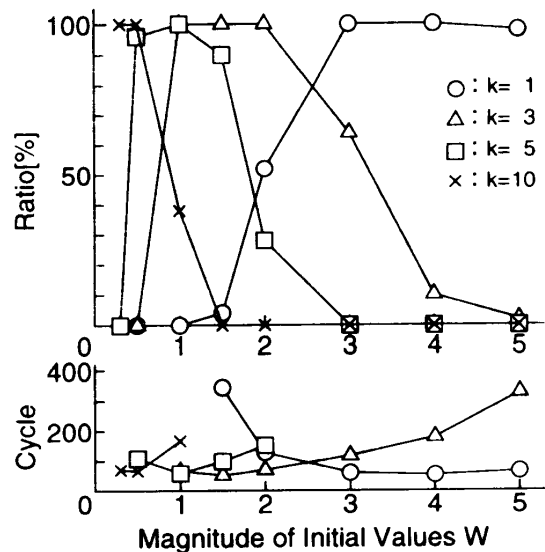


図 3 初期値設定とスケール法から見た典型的な BP 学習の収束
Fig. 3 Convergence of typical BP learnings from initialization and scaling perspectives.

つれて、良好な収束を与える W は小さくなっていることが分かる。これは、 η や α の値を変えたり、学習課題を変えた場合でも同様であった。以上は、入力空間が拡大したときの適切な初期値に対する設定指針を示唆している。

6. む す び

本論文では、入力パターンがアフィン変換の関係にあり本質的に等価と考えられる学習課題に対して、同一構造の多層ネットは、シグモイド関数の極性や値域幅の大きさによらず、等しい濃度の許容解を持つことを指摘した。また、この結果を利用して、両極型ネットを対象に、入力空間が k 倍にスケールされた場合、バイアスを同じにして、荷重を $1/k$ 倍すれば、分離能力や汎化能力は等価になることを指摘した。さらに、荷重やバイアスを上記のように初期化した場合、バイアス更新に対する学習係数を $1/k^2$ 倍、また荷重更新に対する学習係数を $1/k^4$ 倍にすれば、収束挙動は等価になることを述べた。最後に、通常のように、荷重とバイアスの初期値を同一分布幅の一樣乱数で設定し、双方の更新に対する学習係数を等しく与えて学習させた場合、良好な収束を与える初期値は入力空間の拡大とともに小さくなることをシミュレーションにより示した。そして、これらの結果が多層ネットのハードウェア化に際して有用になることを述べた。

シグモイド特性を CMOS トランジスタ回路で実現した場合、 $(k/2) \tanh(u/2G) + \text{オフセット}$ のように^{10),11)}、値域幅 k だけでなく G の大きさも制約さ

れるうえに、オフセットが生じる。オフセットに関する問題は、ここで議論したユニット極性の違いに起因する問題と本質的に同じである。したがって、オフセットについては、3.1 節と同様な手順をふむことで分離および汎化能力の等価なネットが構築可能と考えられる。一方、 G は物性論的に決まる熱電圧であり⁴⁾、必ずしも式(4)のように $G = 1$ となるわけではない。さらに、集積度が高くなるにつれてデバイス特性のばらつきの問題が顕在化することになる¹¹⁾。したがって、オフセットや G のばらつきを考慮に入れたスケーリング法について今後の検討を予定している。

参 考 文 献

- 1) Gallant, S.I.: *Neural Network Learning and Expert Systems*, MIT Press, Cambridge (1993).
- 2) Watanabe, E. and Shimizu, H.: Algorithm for Pruning Hidden Units in Layered Neural Network for Binary Pattern Classification Problem, *Proc. IJCNN*, pp.327-330 (1993).
- 3) 五反田博, 植田吉祥, 川崎武士, 田中伸幸: 逆伝搬学習におけるシグモイド関数と収束の関係, 信学技報, NC93-115, pp.25-32 (1994).
- 4) Haykin, S.: *Neural Networks-A Comprehensive Foundation*, Macmillan Publishing, New York (1994).
- 5) 賈 棋, 戸田尚宏, 白井支朗: ニューラルネットワークにおける逆伝搬学習アルゴリズムの初期値設定に関する一考察, 信学論, Vol.J73-D-II, No.8, pp.1179-1185 (1990).
- 6) 五反田博, 植田吉祥, 川崎武士: ユニット極性の逆伝搬学習に及ぼす影響, 信学論, Vol.J78-D-II, No.9, pp.1372-1382 (1995).
- 7) 五反田博, 植田吉祥, 白土 浩: ユニット極性から見た収束挙動の等価なネットワーク構成, 信学技報, NC94-96, pp.159-166 (1995).
- 8) Maher, M.A.C., Deweerth, S.P., Mahowald, M.A. and Mead, C.A.: Implementing Neural Architectures Using Analog VLSI Circuits, *IEEE Trans. Circuits and Syst.*, Vol.36, No.5, pp.643-652 (1989).
- 9) 久間和生, 中山 高: ニューロコンピュータ工学, 工業調査会 (1992).
- 10) Lansner, J.A. and Lehmann, T.: An Analog CMOS Chip Set for Neural Networks with Arbitrary Topologies, *IEEE Trans. Neural Networks*, Vol.4, No.3, pp.441-444 (1993).
- 11) Dolenko, B.K. and Card, H.C.: Tolerance to Analog Hardware of On-chip Learning in Back-propagation Networks, *IEEE Trans. Neural Networks*, Vol.6, No.5, pp.1045-1052 (1995).
- 12) Gotanda, H., Ueda, Y. and Shiratsuti, H.: So-

lution Space and BP Learning Behaviour of Multilayer Networks Whose Units Are Different in Polarity, *J. Robotics and Mechatronics*, Vol.7, No.4, pp.336-343 (1995).

(平成 8 年 3 月 14 日受付)

(平成 8 年 5 月 10 日採録)



五反田 博 (正会員)

昭和 56 年九州大学大学院博士課程修了。同年近畿大学九州工学部講師を経て、昭和 59 年同助教授となり現在に至る。工博。昭和 61~62 年カリフォルニア大学バークレー校客員研究員。システム同定、信号処理、神経回路網に関する研究に従事。IEEE、電子情報通信学会、計測自動制御学会、神経回路学会等会員。



白土 浩

平成 8 年近畿大学大学院産業技術研究科電子情報工学専攻博士前期課程修了。現在、九州工業大学大学院情報工学研究科情報科学専攻博士後期課程に在学中。ニューラルネットワークに関する研究に従事。電子情報通信学会会員。



井上 勝裕

昭和 56 年九州大学大学院工学研究科博士課程修了。同年同大学大学院総合理工学研究科助手、昭 58 大分大学工学部講師、助教授を経て、平 2 九州工業大学情報工学部制御システム工学科助教授、現在に至る。システム同定、パターン認識理論とその応用に関する研究に従事。工博。電気学会、計測自動制御学会、日本 ME 学会等会員。



熊丸 耕介

昭和 45 年九州大学大学院博士課程修了。同年同大学工学部電気工学科助手、昭和 47 年同講師、昭和 48 年同助教授、昭和 62 年同情報工学科教授を経て、平成元年九州工業大学情報工学部制御システム工学教授、現在に至る。システム同定、状態推定理論とその応用に関する研究に従事。工博。電気学会、計測自動制御学会、日本 ME 学会等会員。