距離型 cos 素子を用いたネットワークモデル

非会員 大 池 浩 一 (千 葉 大)

正 員 小 坪 成 一 (千 葉 大)

正 員 平 田 廣 則 (千 葉 大)

A Network Model using Distance based Cosine Elements

Koichi Oike, Non-member, Seiichi Koakutsu, Member, Hironori Hirata, Member (Chiba University)

In this paper, we propose a new network element named a “distance-based cosine element” for neural networks. We also derive a learning algorithm based on the back-propagation algorithms for multilayer networks. The distance-based cosine element inputs a squared distance between an input pattern vector and its weight vector, and uses an affine transformation of cosine function as its output function. The proposed distance-based cosine network is able to improve its learning speed as well as convergence rate because its output function does not have any saturated regions which cause slow learning speed of the back-propagation learning using sigmoid elements.

We demonstrate the advantages of the proposed network by solving N -bits parity problems and Fisher's Iris classification problem. Experimental results indicate that our distance-based cosine network consistently obtains better results than the conventional sigmoid network in terms of both the learning speed and the convergence rate.

キーワード：階層型ニューラルネットワーク，パリティ問題，距離型 cos 素子

1. ま え が き

現在，多くの分野に応力されている階層型ニューラルネットワークでは，基本素子（ユニット）として積和型シグモイド素子が用いられ，学習則としては D. E. Rumelhart らが考案したバックプロパゲーション（BP）法⁽¹⁾を採用している。

こうした従来のニューラルネットワークの欠点としては，学習に時間がかかることや，ローカルミニマムによる学習の停止などがあげられる。学習時間はネットワークの能力と直接結びつくものではないが，大規模な問題に適用する場合には大きな問題となる。

これらの問題を改善するための方法として，BP 法のアルゴリズムの改良が数多く検討されている。たとえば，誤差評価関数に Kullback Divergence を用いる方法⁽²⁾やユニットの出力値に応じて学習係数を制御する DCLA 法⁽³⁾，最急降下法の代わりに共役勾配法を用いる Quick Prop⁽⁴⁾，共役勾配法とランダム探索法を組み合わせたハイブリッドアルゴリズム⁽⁵⁾などがあり，学習時間の短縮やローカルミニマムからの脱出に効果のあることが報告されている。

一方，ニューラルネットワークに使われている素子自体に着目し，入出力関数を変更することによって積和型シグモイド素子とは異なる特性を持った素子をデザインし，学習時間の短縮や汎化能力の向上を図る試みも行われている。出力関数を変更するものとして，複数のシグモイド関数の合成による非単調な関数を用いた例⁽⁶⁾⁽⁷⁾や，正弦波関数を用いた例⁽⁸⁾があり，パリティ問題や二重らせん問題に対してシグモイド関数よりも学習能力が優れていると報告されている。また，入出力関数ともに変更するものとしては，入力と重みの差の二乗和を入力とし，単調減少の exp 関数を出力関数とする学習距離素子によるネットワーク⁽⁹⁾⁽¹⁰⁾があり，学習能力の向上や大規模化への有効性が報告されている。

本論文で提案する距離型 cos 素子も，こうした非シグモイド素子の一種で，入力と重みの差の二乗和を入力，非単調な周期関数である cos 関数を出力関数とするものである。正規化されていない入力パターンに対応するため，以前著者が提案したモデル⁽¹¹⁾に伸縮度と呼ばれるパラメータを導入している。その特徴として，出力関数に周期関数を使うことにより飽和領域にとらわれることなく高速かつ収束率の高い学習が可能であることがあげられる。

以下第2章では距離型 cos 素子のモデルを提示し、これを中間層および出力層のユニットとする距離型 cos ネットワークを構成して、BP 法に基づいた学習アルゴリズムについて述べる。続いて、第3章では距離型 cos 素子の特徴とその理由について定性的に述べた後、第4章、第5章では N ビットパリティ問題およびフィッシャーのアヤメ分類問題⁽¹²⁾の学習を行い、学習速度や収束特性、汎化能力についてシグモイド素子および非シグモイド素子によるネットワークとの比較検討を行う。最後に、第6章でまとめと今後の課題を述べる。

2. 距離型 cos 素子モデル

〈2・1〉 入出力関数 階層型ニューラルネットワーク(図1)で一般に使用されるシグモイド素子では、入力信号(結合を伝搬してくる他の素子からの出力信号)と、各結合に割り当てられた重みの積を、すべての入力結合について総和したものを入力とし、出力関数としては単調増加関数のシグモイド関数を用いる。

これに対し、本論文で提案する距離型 cos 素子では、入力信号と重みの差の二乗を総和したものを入力とする。出力関数としては周期関数の cos 関数を用いる。なお、sin 関数ではなく cos 関数を採用するのは、入力が0の場合、つまり入力信号と重みが一致した場合に最大値1を出力するためである。

いま、入力パターン p がネットワークに提示され、 n 個のユニット $u_i (i=1 \cdots n)$ からの出力信号 o_{pi} が、それぞれ重み w_{ji} を持つ結合を通じてユニット u_j に伝えられるとする。

シグモイド素子では、内部状態 net_{pj} は以下の式で与えられる。

$$net_{pj} = \sum_{i=1}^n w_{ji} o_{pi} - \theta_j \cdots \cdots \cdots (1)$$

ここで、 θ_j はユニットのしきい値である。

この net_{pj} を入力とする出力関数 f としては、以下のシグモイド関数を用いる。

$$o_{pj} = f(net_{pj}) = \frac{1}{1 + \exp(-net_{pj})} \cdots \cdots \cdots (2)$$

図1 階層ニューラルネットワーク
Fig. 1. Multi layered neural network.

図2からも明らかなように、この関数は単調増加関数で、出力信号 o_{pj} は $(0, 1)$ の範囲に収まる。

一方、距離型 cos 素子では、内部状態 net_{pj} は以下の式で与えられる。

$$net_{pj} = \frac{\rho_j^2}{n} \sum_{i=1}^n (w_{ji} - o_{pi})^2 \cdots \cdots \cdots (3)$$

この場合、 net_{pj} は必ず非負となる。なお、総和を入力数 n で割っているのは、入力結合数の違いによる出力への影響を押さえるためである。また、 ρ_j は出力関数の適切な領域で学習を行えるように cos 関数の周期を調整する伸縮度で、各ユニットごとに設けられ、重み w_{ji} と同様に学習により値が更新される。

出力関数 f としては cos 関数をアフィン変換した以下の式が用いられる。

$$o_{pj} = f(net_{pj}) = \frac{\cos(\pi net_{pj}) + 1}{2} \cdots \cdots \cdots (4)$$

図3からも明らかなように、この関数は周期関数で、出力信号 o_{pj} は $[0, 1]$ の範囲に収まる。

本論文では、1層以上の中間層を持つ階層型ニューラルネットワークの中間層および出力層のユニットに、(3)、(4)式で表される距離型 cos 素子を使用したものを、距離型 cos ネットワークと呼ぶ。入力層のユニットはネットワークに与えられる入力信号 x_{pi} をそのまま出力する線形素子とする。

$$o_{pi} = net_{pi} = x_{pi} \cdots \cdots \cdots (5)$$

〈2・2〉 学習アルゴリズム 図1のような階層型ニューラルネットワークに対する教師あり学習法としては、D. E. Rumelhart らが考案したバックプロパゲーション (BP) 法⁽¹⁾が広く用いられている。BP 法による学習では、ネットワークの出力信号と教師信号を比較し、その誤

図2 シグモイド素子の出力関数
Fig. 2. Output function of sigmoid element.

図3 距離型 cos 素子の出力関数
Fig. 3. Output function of cosine element.

差に基づいた重みの修正を出力層から入力層に向かって伝搬させる。

本論文では、距離型 cos ネットワークに対して、この BP 法に基づく学習を行うものとする。以下の説明では、簡単のため中間層 1 層の距離型 cos ネットワークを考える。また、各層ユニット数は、入力層が n 、中間層が m 、出力層が l 個であるとする。

あるパターン p の入力信号 $x_{pi}(i=1, \dots, n)$ が入力層の各ユニット u_i に与えられると、(5)式および(3)、(4)式に従って入力層から中間層の各ユニット $u_j(j=1, \dots, m)$ 、 u_j から出力層の各ユニット $u_k(k=1, \dots, l)$ へと信号が伝搬し、最終的な出力信号 o_{pk} が得られる。これを与えられた教師信号 t_{pk} と比較し、二乗誤差 E_p を計算する。

$$E_p = \frac{1}{2} \sum_{k=1}^l (t_{pk} - o_{pk})^2 \quad \dots\dots\dots (6)$$

この E_p から、出力層のユニット u_k の学習信号 δ_{pk} が以下のように定義される。

$$\begin{aligned} \delta_{pk} &\equiv -\frac{\partial E_p}{\partial net_{pk}} \\ &= (t_{pk} - o_{pk})f'(net_{pk}) \quad \dots\dots\dots (7) \end{aligned}$$

ここで、 f' は出力関数 f の導関数で、

$$f'(net_{pk}) = -\frac{\pi}{2} \sin(\pi net_{pk}) \quad \dots\dots\dots (8)$$

これを(7)式に代入すると、

$$\delta_{pk} = -\frac{\pi(t_{pk} - o_{pk})}{2} \sin(\pi net_{pk}) \quad \dots\dots\dots (9)$$

となる。

よって、出力層のユニット u_k と中間層のユニット u_j 間の重みの修正値 Δw_{kj} 、および u_k の伸縮度 ρ_k の修正値 $\Delta \rho_k$ は、一般化モデル則⁽¹⁾により以下の式で与えられる。

$$\Delta w_{kj} = \eta_w \delta_{pk} \frac{2\rho_k^2}{m} (w_{kj} - o_{pj}) \quad \dots\dots\dots (10)$$

$$\Delta \rho_k = \eta_\rho \delta_{pk} \frac{2\rho_k}{m} \sum_{j=1}^m (w_{kj} - o_{pj})^2 \quad \dots\dots\dots (11)$$

同様に、中間層のユニット u_j の学習信号 δ_{pj} 、重みの修正値 Δw_{ji} 、伸縮度の修正値 $\Delta \rho_j$ は、それぞれ以下のよう

$$\delta_{pj} = \frac{\pi}{m} \sin(\pi net_{pj}) \sum_{k=1}^l \delta_{pk} \rho_k^2 (w_{kj} - o_{pj}) \quad \dots\dots\dots (12)$$

$$\Delta w_{ji} = \eta_w \delta_{pj} \frac{2\rho_j^2}{n} (w_{ji} - o_{pi}) \quad \dots\dots\dots (13)$$

$$\Delta \rho_j = \eta_\rho \delta_{pj} \frac{2\rho_j}{n} \sum_{i=1}^n (w_{ji} - o_{pi})^2 \quad \dots\dots\dots (14)$$

ここで、 $\eta_w, \eta_\rho (>0)$ は学習率と呼ばれるフリーパラメータである。

一方、シグモイド素子では、重みの修正値 Δw 、しきい値の修正値 $\Delta \theta$ はそれぞれ以下の式で与えられる。

$$\begin{aligned} \text{(出力層)} \quad \begin{cases} \delta_{pk} = (t_{pk} - o_{pk})o_{pk}(1 - o_{pk}) \\ \Delta w_{kj} = \eta \delta_{pk} o_{pj} + \alpha \Delta w_{kj}^{old} \\ \Delta \theta_k = -\eta \delta_{pk} + \alpha \Delta \theta_k^{old} \end{cases} \quad \dots\dots\dots (15) \end{aligned}$$

図 4 重みの変化に対する出力関数の形状
Fig. 4. Output function of cosine element for its weight change.

$$\begin{aligned} \text{(中間層)} \quad \begin{cases} \delta_{pj} = o_{pj}(1 - o_{pj}) \sum_{k=1}^l \delta_{pk} w_{kj} \\ \Delta w_{ji} = \eta \delta_{pj} o_{pi} + \alpha \Delta w_{ji}^{old} \\ \Delta \theta_j = -\eta \delta_{pj} + \alpha \Delta \theta_j^{old} \end{cases} \quad \dots\dots\dots (16) \end{aligned}$$

なお、(15)、(16)式の $\Delta w, \Delta \theta$ の第 2 項は慣性項と呼ばれ、収束を高速化するために使われている。 $\Delta w^{old}, \Delta \theta^{old}$ はそれぞれ、前回の学習における修正値で、 $\alpha (0 \leq \alpha \leq 1)$ は慣性係数と呼ばれるフリーパラメータである。

〈2・3〉 伸縮度 ρ の初期値設定 距離型 cos ネットワークでは、(3)式からわかるように、各ユニットへの入力信号（前層ユニットの出力）と重みの差の二乗に比例して内部状態 net_{pj} の値が決まる。このため、入力信号と重み（初期値は 0.5 付近の一樣乱数）の差の絶対値が大きい学習初期などでは、出力関数のピーク間が狭い領域で学習が行われるため（図 4）、わずかな重みの修正により出力値が大きく変動して学習が収束しないなどの問題が生じる。

そこで、出力関数がゆるやかな領域で学習が開始されるよう、中間層、出力層の各ユニットの伸縮度 ρ の初期値を適切に設定することを考える。

中間層ユニットへの入力（すなわち入力層ユニットの出力）は、ネットワークに与えられる入力信号 x_{pi} そのものなので、出力値の範囲は問題に応じて変化する。そのため、(17)式で求められる ρ_{ini} を、中間層ユニットの伸縮度 ρ_j の初期値（またはその目安）とする。

$$\rho_{ini} = \frac{2}{|x_{\max} - 0.5| + |x_{\min} - 0.5|} \quad \dots\dots\dots (17)$$

ここで、 $x_{\max}, x_{\min}$ はそれぞれ、入力信号の最大値と最小値である。

一方、出力層ユニットに関しては、中間層ユニットの出力値が(4)式により $[0, 1]$ の範囲に収まるため、伸縮度 ρ_k の初期値は常に 1.0 とする。

3. 距離型 cos 素子の特徴

距離型 cos 素子をネットワークの構成要素とすることにより、シグモイド素子によるネットワークにおける学習の停止や収束率の低下といった問題点が改善される。以下では、その理由について定性的に述べる。

シグモイド素子の重みの修正値 Δw のうち、各層のユ

図 5 シグモイド素子の出力値と重みの修正値の関係

Fig. 5. Relation between output and renewal value of weight (sigmoid element).

ニットの出力値 o_p に依存する分の修正値を ΔR とすると, (15), (16)式より,

$$(\text{出力層}) \quad \Delta R = (t_p - o_p) o_p (1 - o_p) \dots\dots\dots (18)$$

$$(\text{中間層}) \quad \Delta R = o_p (1 - o_p) \dots\dots\dots (19)$$

となる。これらの ΔR と o_p の関係を図 5 に示す。ただし, 出力層については, 教師信号 t_p が 0 および 1 の場合のみ描画している。

この図からもわかるように, シグモイド素子に対する BP 法の学習では, ユニットの出力値が 0, 1 に近づくにつれ, 重みの修正量 $|\Delta R|$ も小さくなる。誤った学習の結果, 中間層ユニットの出力値が 0, 1 に近い飽和領域に入ってしまうと, 重みの修正量が小さくなるため, 飽和領域から脱出することは容易ではない。

また, 出力層のユニットについて考えると, 教師信号との誤差が大きいくほど重みの修正量が大きくなるべきである。しかし, 実際の重みの修正量は(18)式のようにシグモイド関数の導関数 $o_p(1-o_p)$ が含まれるため, 教師信号との誤差が増加するにもかかわらず, 修正量 $|\Delta R|$ が減少してしまう領域が存在する ($t_p=0$ の場合は $o_p > 0.67$, $t_p=1$ の場合は $o_p < 0.33$)。

以上のことが, シグモイド素子に対する BP 法の学習において, 学習の停止や収束率の低下を招く原因であると考えられる⁽³⁾。

一方, 距離型 cos 素子では, 出力関数が周期関数なので, ある一つの出力値に複数の入力値が対応する。そのため, ΔR と o_p の関係は, net_p を媒介変数とした次の式で表される。

$$o_p = \frac{\cos(\pi net_p) + 1}{2} \dots\dots\dots (20)$$

$$(\text{出力層}) \quad \Delta R = -\frac{\pi(t_p - o_p)}{2} \sin(\pi net_p) \dots\dots\dots (21)$$

$$(\text{中間層}) \quad \Delta R = \frac{\pi}{m} \sin(\pi net_p) \sum_{k=1}^l \delta_{pk} \rho_k^2(w_{kj} - o_p) \dots\dots\dots (22)$$

(21), (22)式の o_p に(20)式の右辺を代入して net_p のみの関数とし, net_p を媒介変数として ΔR と o_p の関係をグラフ化すると, 図 6 のような閉曲線となる。ただし, 出力層

図 6 距離型 cos 素子の出力値と重みの修正値の関係

Fig. 6. Relation between output and renewal value of weight (cosine element).

については教師信号 t_p が 0 および 1 の場合のみ, 中間層については, $l=1$, $m=2$, $\delta_{p1} = \pi(1+\sqrt{2})/8$ (最大値), $w_{1j}=0.5$, $\rho_1=1.0$ の場合のみ描画している。

これらの閉曲線の軌道は, $net_p=0$ の $(1, 0)$ から出発し, $net_p=1$ のとき原点 $(0, 0)$ を通り, $net_p=2$ で元の $(1, 0)$ に戻り, 以後は同じ軌道を通る (図中の矢印を参照)。なお, 出力関数では $0 < net_p < 1$ が減少部分, $1 < net_p < 2$ が増加部分に相当する。

この図を見ると, 距離型 cos 素子でも, ユニットの出力値が 0, 1 に近づくにつれて重みの修正量は小さくなる。しかし, シグモイド素子とは違って出力関数に飽和領域が存在しないため, 出力値が 0, 1 に近い領域からの脱出は容易であると考えられる。

また, 出力層の ΔR のグラフは, シグモイド関数のそれよりも左右にかたよった形状をしており, 教師信号との誤差が増加するにもかかわらず修正量 $|\Delta R|$ が減少してしまう領域は, シグモイド素子よりも 0.08 ほど狭くなっている ($t_p=0$ の場合は $o_p > 0.75$, $t_p=1$ の場合は $o_p < 0.25$)。

これらのことから, 距離型 cos 素子をネットワークの構成要素とすることにより, 従来のシグモイド素子によるネットワークにおける学習の停止や収束率の低下といった問題点が改善され, 短時間で学習が収束することが期待できる。

4. N ビットパリティ問題への適用

距離型 cos ネットワークに対して, 2 ビット (XOR) から 5 ビットまでのパリティ問題を学習させ, 従来のシグモイド素子および非シグモイド素子によるネットワークとの比較を行った。

〈4・1〉 実験の条件 以下の実験は, すべて PC 互換機の C 言語で書かれたシミュレータ上で行った。

各パリティ問題の学習で使われるネットワークは, 問題のビット数 N に応じて $N-N-1$ (入力層 N ユニット, 中間層 N ユニット, 出力層 1 ユニット) の 3 層構造とする。

学習データは 2^N 個で, ネットワークへの入力パターン

は0.1の2値をそのまま使用，教師パターンでは0の代わりに0.1，1の代わりに0.9を使用する。

距離型 cos 素子によるネットワークでは，重み w の初期値は $0.0+0.3$ の一様乱数，中間層の伸縮度の初期値 ρ_{ini} は，(17)式より2.0に設定した。

比較対象としては，以下の3種類の素子を使ったネットワークを使用した。

(i) 積和型シグモイド素子（入力：積和型，出力：単調増加）

BP法による学習のほか，誤差評価関数に Kullback Divergence を用いる学習⁽²⁾もあわせて行った。

(ii) 積和型 sin 素子（入力：積和型，出力：周期関数）
文献⁽⁸⁾では，出力関数の周期をパラメータ制御することにより，積和型 sin 素子を周期カオスニューロンとして用いているが，ここでは比較のため周期を一定にして通常のニューロンとして用いる。また，出力値が $[0, 1]$ となるよう出力関数の sin 関数をアフィン変換している。

(iii) 学習距離素子（入力：距離型，出力：単調減少）
文献⁽⁹⁾⁽¹⁰⁾の学習距離素子をそのまま用いる。

各ネットワークのフリーパラメータや初期値は，本実験の前に行った小規模な予備実験の結果に基づき，適切と思われる値を設定している。

これら4種類のネットワークに対し，フリーパラメータである学習率 η を変化させ，各学習に対してそれぞれ100種類の重みの初期値で学習を行い，その傾向を調べることとする。

誤差の尺度としては，PROBEN 1⁽¹³⁾で提案された二乗誤差パーセンテージ (squared error percentage) を用いる。

$$E_{sep} = 100 \frac{O_{max} - O_{min}}{lP} \sum_{p=1}^P \sum_{k=1}^l (t_{pk} - o_{pk})^2 \dots\dots\dots (23)$$

ここで， l は出力層のユニット数， P は訓練セットのパターン数， O_{max} , O_{min} は（問題表現上の）出力の最大値と最小値である。つまり，通常使われる二乗誤差の総和やその平均に対し，更にパターン数や出力値のレンジを考慮に入れて変形したものである。

学習が収束したかどうかは，この E_{sep} が1(%)未満に収束したかどうかで判断する。なお収束しないまま10,000エポックに達したら，学習失敗とみなして学習を終了させるものとする。

〈4・2〉 実験結果 各ネットワークで得られた結果のうち，もっとも収束率が高いもの（収束率が等しい場合は平均学習回数をもっとも少ないもの）を表1に示す。更に，各ネットワークの XOR 問題の学習における E_{sep} の収束の様子を図7に示す。

なお，表中の (BP) はBP学習，(KD) はKullback Divergence を用いる学習を示しており，平均学習時間とは，各パリティ問題での積和型シグモイド素子 (BP 学習) の学習時間を100%とした相対時間を表している。

表1の学習結果をみると，積和型シグモイド素子 (BP 学習) によるネットワークでは，収束に多くの学習時間を要し，問題が複雑になるにつれて収束率も低下している。Kullback Divergence を導入することにより，学習は高速化されるが，収束率の低下を招くことが多い。

表 1 N ビットパリティ問題の実験結果
Table 1. Simulation results of learning N -bits parity problems.

ビット数 n	素子の種類	学習効率			慣性係数 α	収束率 (%)	平均学習回数 (epochs)	平均相対学習時間 (%)
		η_w	η_θ	η_ρ				
2	積和型シグモイド素子 (BP)	0.5	0.5	—	0.9	95	325	100
	(KD)	0.9	0.9	—	0.9	93	304	94
	積和型 sin 素子	0.1	0	—	—	100	91	24
	学習距離素子	0.1	0.1	0.1	0.5	100	39	14
	距離型 cos 素子	0.1	—	0	—	100	17	5.5
3	積和型シグモイド素子 (BP)	0.9	0.9	—	0.9	100	1327	100
	(KD)	0.7	0.7	—	0.5	97	974	73
	積和型 sin 素子	0.1	0	—	—	100	202	13
	学習距離素子	0.03	0.03	0.03	0.5	99	362	30
	距離型 cos 素子	0.2	—	0	—	100	17	1.4
4	積和型シグモイド素子 (BP)	0.6	0.6	—	0.9	91	4136	100
	(KD)	0.3	0.3	—	0.8	80	2991	72
	積和型 sin 素子	0.04	0	—	—	100	515	11
	学習距離素子	0.009	0.05	0.033	0.5	89	1177	34
	距離型 cos 素子	0.3	—	0.001	—	100	29	0.83
5	積和型シグモイド素子 (BP)	0.8	0.8	—	0.9	14	5861	100
	(KD)	0.7	0.7	—	0.9	49	2930	50
	積和型 sin 素子	0.01	0	—	—	81	699	11
	学習距離素子	0.01	0.05	0.001	0.5	22	2742	57
	距離型 cos 素子	0.4	—	0.01	—	100	35	0.75

図 7 XOR 問題の学習における誤差の収束
Fig. 7. Error convergence of learning XOR problem.

積和型 sin 素子によるネットワークでは、5 ビットパリティ問題を除いて収束率は 100% に達しており、学習時間も積和型シグモイド素子 (BP 学習) の 11~24% しか要さない。

学習距離素子を用いたネットワークでは、XOR (2 ビットパリティ) 問題では優秀な結果を示しているが、問題が複雑になるにつれ収束率が低下してしまう。

これらに対し、距離型 cos ネットワークではすべての問題で収束率が 100% に達し、かつ学習時間も積和型シグモイド素子 (BP 学習) の 0.75~5.5% ときわめて短くなっていることがわかる。

5. アヤメ分類問題への適用

距離型 cos ネットワークに対して、フィッシャーのアヤメ分類問題⁽¹²⁾を学習させ、汎化能力に関する比較を行った。

〈5・1〉 実験の条件 フィッシャーのアヤメ分類問題は、がくと花卉それぞれの幅と長さ (いずれも cm 単位) の 4 種類の入力パターンから、3 種類のアヤメ (Iris Setosa, Iris Versicolour, Virginica) を分類する問題である。

データはアヤメの種類ごとに 50 個ずつ (計 150 個) 用意されており、このうち 5 個ずつランダムに抽出した計 15 個を学習データ、残りの 135 個を汎化能力評価用のテストデータとして用いる。

ネットワークは 4-2-3 の 3 層構造とする。出力層のユニット 3 個は、3 種類のアヤメそれぞれに対応しており、出力パターンは (0.9, 0.1, 0.1), (0.1, 0.9, 0.1), (0.1, 0.1, 0.9) のいずれかである。入力パターンは、フィッシャーのデータをそのまま利用する。

距離型 cos 素子によるネットワークでは、重み w の初期値は 0.0 ± 0.3 の一様乱数、中間層の伸縮度の初期値 ρ_{ini} は、(17) 式より 0.22 に設定した。

比較対象のネットワークは、 n ビットパリティ問題と同

表 2 アヤメ分類問題の実験結果
Table 2. Simulation results of fisher's iris classification problem.

素子の種類	学習効率			慣性係数 α	誤差 E_{sep} (%)
	η_w	η_θ	η_ρ		
積和型 (BP)	0.1	0.1	---	0.9	4.3
シグモイド素子 (KD)	0.1	0.1	---	0.5	4.4
積和型 sin 素子	3.0e-4	1.0e-3	---	---	4.5
学習距離素子	---	---	---	---	---
距離型 cos 素子	0.06	---	1.0e-4	---	6.0

じ 3 種類を用いる。フリーパラメータおよび初期値は、本実験の前に行った小規模な予備実験の結果に基づき、適切と思われる値を設定している。

これら 4 種類のネットワークに対して、それぞれ 100 種類の重みの初期値で学習を開始する。過学習を避けるため、(23) 式の二乗誤差パーセンテージ E_{sep} が 1 (%) 未満に収束した時点で学習を打ち切り、テストデータに対する E_{sep} を求めて汎化能力を比較する。

〈5・2〉 実験結果 各ネットワークで得られた結果を表 2 に示す。なお、学習距離素子によるネットワークでは、どのようにパラメータを調整しても学習が進まず、結果を示すことができなかった。

表 2 の距離型 cos ネットワークの結果をみると、シグモイド素子のネットワークより、テストデータに対する E_{sep} が 1.7% 大きくなっており、それだけ汎化能力が劣っていることがわかる。

学習距離素子によるネットワークで学習が進まなかったことも考え合わせると、この結果は、実測値のような誤差を含んだ入力データに対して、積和型の入力関数よりも距離型の入力関数のほうが鋭敏に反応するためと考えられる。

6. む す び

本論文では、入力と重みの差の二乗和を入力、cos 関数を出力関数とする距離型 cos 素子を提案した。この素子により構成される階層型ネットワークは、学習が高速で収束率が高いという特徴を持つ。実際、5 ビットパリティ問題に対して平均 35 エポックで 100% 学習が収束するなど、従来のシグモイド素子や非シグモイド素子によるネットワークよりも極めて少ない学習回数で高速に収束するという結果が得られた。一方、アヤメの分類問題のような誤差を含んだデータに対しては、シグモイド素子によるネットワークより汎化能力がいくぶん劣るとの結果が得られており、何らかの対策が必要であると考えている。

今後は、この距離型 cos ネットワークをより大規模なパターン認識問題や関数近似問題などに対して応用し、その能力を更に明らかにしていきたい。また、振動系の出力関数を導入したことによる系の安定性などについての検討も行っていきたい。

(平成 9 年 1 月 20 日受付, 同 9 年 5 月 29 日再受付)

文 献

- (1) D.E. Rumelhart, J.L. McClelland and the PDP Research Group: *Parallel distributed processing*, MIT Press (1986)
- (2) 丹・加藤・江島:「誤差評価関数による PDP モデルの改良」, 信学論(D-II), **J73-D-II**, 12, 2022~2028 (平2)
- (3) 猿田・孫・安部・根元:「係数変化学習法とその手書き文字認識への応用」, 信学論(D-II), **J78-D-II**, 6, 973~981 (平7)
- (4) S. Fahlman: "An empirical study of learning in back-propagation networks", *Technical Report of Carnegie Mellon University*, CMU-CS-88-162 (1988)
- (5) N. Baba, et al: "A hybrid algorithm for finding the global minimum of error function of neural networks and its applications", *Neural Networks*, **7**, 1253~1265 (1994)
- (6) 小谷・松本・金川:「非単調素子を持つ階層型ニューラルネットワークの学習能力」, 計測自動制御学会論文集, **29**, 12, 1465~1473 (平5)
- (7) 小谷・松本:「非単調なユニットから構成される階層型ニューラルネットワーク」, 第21回知能システムシンポジウム, 235~240 (平7)
- (8) 岡田・中川:「周期活性化関数を用いた誤差逆伝播法に関する研究」, 信学技報, **NC96-23** (平8-07)
- (9) 永田・横井:「距離情報を基にした素子のネットワーク」, 信学技報, **NC91-45** (平3-09)
- (10) 村山・永田・横井:「学習距離素子によるネットワークの学習能力」, 信学技報, **NC92-51** (平4-09)
- (11) 大池・小坪・平田:「距離型 cos ネットワークの学習能力」, 信学技報, **NC95-157** (平8-03)
- (12) R. A. Fisher: "The use of multiple measurements in taxonomic problems", *Annual Eugenics*, **7** (2), 179~188 (1936)
- (13) L. Prechelt: "PROBEN 1—A Set of Benchmarks and Benchmarking Rules for Neural Network Training Algorithms", *Technical Report 21/94*, Fakultät für Informatik, Universität Karlsruhe, (1994), Anonymous FTP: /pub/papers/techreports/1994/1994-21. ps. Z on ftp. ira. uka. de

大 池 浩 一 (非会員) 1967 年 3 月 6 日生。92 年千葉大学大学院工学研究科電子工学専攻修士課程修了。現在, 同大学院自然科学研究科生産科学専攻博士課程在学中。ニューラルネットワークによる最適化および VLSI レイアウト設計の研究に従事。電子情報通信学会会員。

小 坪 成 一 (正員) 1964 年 7 月 7 日生。92 年千葉大学大学院自然科学研究科生産科学専攻博士課程修了。同年, 千葉大学工学部助手, 97 年同助教授, 同年同大学院自然科学研究科助教授となり現在に至る。1994~95 年, カリフォルニア大学サンタクルズ校客員研究員。VLSI レイアウト設計, 確率的最適化手法, ニューラルネットワーク, 学習オートマトンの研究に従事。博士(工学)。

IEEE, INNS, 電子情報通信学会, 計測自動制御学会会員。

平 田 廣 則 (正員) 1948 年 6 月 2 日生。76 年東京工業大学大学院電気工学専攻博士課程修了。同年, 千葉大学工学部助手, 81 年同助教授, 94 年同教授, 97 年同大学院自然科学研究科教授となり現在に至る。大規模システムのモデリング, 解析と設計, 特に生態システム, VLSI レイアウト, 並びに分散型コンピュータの基礎的理論に興味をもつ。工学博士。IEEE, ISEM, INNS, 電子情報通信学会, 情報処理学会, 計測自動制御学会, システム制御情報学会, 人工知能学会会員。