研究速報

並列分散処理ネットワークのための新素子 ——学習距離積素子——

正員 村山 聡敏[†]　　正員 横井 博一[††]

A New Element for Parallel Distributed Processing Network——Learning Distance Product Element

Satoshi MURAYAMA[†], *Associate Member and* Hirokazu YOKOI[††], *Member*

[†] 日本電気株式会社，川崎市
NEC Corporation, Kawasaki-shi, 211 Japan

[††] 九州工業大学工学部，北九州市
Faculty of Engineering, Kyushu Institute of Technology, Kitakyushu-shi, 804 Japan

あらまし　並列分散処理ネットワークの学習における収束速度に影響を与える要因の一つは，素子の入出力特性である．そこで本論文では，従来の学習しきい素子に代わって学習距離積素子を新たに提案し，この素子を用いた2入力1出力の3層ネットワークで計算機シミュレーションを行った．その結果，学習しきい素子を用いた場合に比べて，学習が平均的に加速されることが確認された．

キーワード　ニューラルコンピュータ，ニューラルネット，並列分散処理，学習

1. まえがき

現在盛んに研究されているニューラルネットワークは，学習しきい素子で構成された並列分散処理ネットワークで，階層型の場合には誤差逆伝搬学習則が最もよく用いられる．しかしながら，この学習則には(1)ローカルミニマの存在のために最適解でないところへ収束する，(2)収束に時間がかかる，という二つの問題がある．後者の解決策としては，素子の入出力特性，ネットワークの構造，ネットワークの学習則を改良することなどが考えられる．

そこで筆者らは，素子の入出力特性の改良の方向からネットワークの学習をより高速化することを目的に，入力ベクトルと1個の基準ベクトルとの距離をもとに動作する素子を提案し，これを学習距離素子と名づけた[(1),(2)]．一方，筆者らとは独立に，学習距離素子とほとんど同じ入出力特性をもったRBFユニットが提案されており，これを用いたRBFネットワークの研究も行われている[(3)~(5)]．但し，このネットワークではRBFユニットが用いられるのは一つの層だけである．従って，ネットワークの学習則は筆者らが導いたものとは異なる．

本論文では，学習距離素子を更に発展させ，入力ベクトルと複数個の基準ベクトルとの距離の積をもとに動作する素子を新たに提案し，これを学習距離積素子と名づける．次に，学習距離積素子を用いた階層型ネットワークに確率的降下法を適用することにより，ネットワークの学習則を導き出す．最後に，この学習則により2入力1出力の3層ネットワークに2変数論理関数を学習させたときの学習の収束の様子を計算機シミュレーションにより調べ，学習しきい素子を用いた場合と比較する．

2. 学習距離積素子

2.1 学習しきい素子の入出力特性の一般化

従来用いられている学習しきい素子の入出力特性は，式(1)で示される．

$$z=f(\boldsymbol{w}\cdot\boldsymbol{x}-h) \tag{1}$$

ここで，$\boldsymbol{x}$ は入力ベクトル，$\boldsymbol{w}$ は荷重ベクトル，h はしきい値，$f(\cdot)$ は出力写像，z は出力信号である．出力写像としては，離散情報型では単位ステップ関数，連続情報型ではランプ関数，区分線形関数，ロジスティック関数等が用いられる．

式(1)によれば，学習しきい素子が行う情報処理は，入力ベクトル x を入力ベクトルと荷重ベクトルの内積 $\boldsymbol{w}\cdot\boldsymbol{x}$ に変換する過程と，これを出力信号 z に変換する過程の二つに分けることができる．前者の過程は，入力ベクトルが荷重ベクトルにどの程度類似しているかの判断を行っていると見ることもでき，これを判断過程と呼ぶことにする．後者の過程は，入力ベクトルと荷重ベクトルの類似度が高いと出力信号が大きくなるように素子の動作を決定しており，これを動作過程と呼ぶことにする．

そこで，学習しきい素子の入出力特性を一般化し，並列分散処理ネットワーク用素子の入出力特性を一般的に式(2)，(3)のように表す．

$$u=G(\boldsymbol{x}) \tag{2}$$

$$z=F(u) \tag{3}$$

ここで，u は判断量，$G(\cdot)$ は判断写像，$F(\cdot)$ は動作写像である．

この一般的な素子では，判断写像として類似度写像と距離写像の2通りが存在する．類似度写像は，入力ベクトルと素子のもつ基準ベクトルとの類似度を判断量とする写像で，このとき動作写像は単調増加関数となる．類似度写像の具体例としては，入力ベクトルと基準ベクトルの内積を判断量とする写像のほかに，入力ベクトルと基準ベクトルの方向余弦を判断量とする写像，入力ベクトルと複数個の基準ベクトルそれぞれとの内積や方向余弦の和あるいは積を判断量とする写

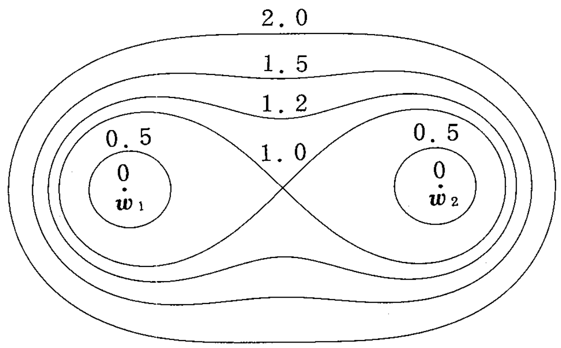

図1 学習距離積素子の判断写像の性質
Fig. 1 The property of judgment mapping of learning distance product element.

像等が考えられる．

一方，距離写像は，入力ベクトルと素子のもつ基準ベクトルとの距離を判断量とする写像で，このとき動作写像は単調減少関数となる．距離写像の具体例としては，入力ベクトルと基準ベクトルのユークリッド距離を判断量とする写像，入力ベクトルと基準ベクトルのマハラノビス距離を判断量とする写像，入力ベクトルと複数個の基準ベクトルそれぞれとのユークリッド距離やマハラノビス距離の和あるいは積を判断量とする写像等が考えられる．

2.2 学習距離積素子の入出力特性

学習距離積素子は，基準ベクトルを複数個もち，判断写像として距離写像を用いる．その入出力特性は，離散情報型の場合には式(4)，(5)で表され，連続情報型の場合には式(4)，(6)で表される．

$$u=\prod_n \|\boldsymbol{w}_n-\boldsymbol{x}\| \tag{4}$$

$$z=\mathbf{1}(-(u-\sigma)) \tag{5}$$

$$z=\exp(-(\rho^2u^2)^A) \tag{6}$$

ここで，$\boldsymbol{w}_n$ は第 n 基準ベクトル，$\|\ \|$ はノルム，$\mathbf{1}(\cdot)$ は単位ステップ関数，σ は選択の幅，ρ は選択の鋭さ，A は正の実数である．式(5)は，式(6)において A を ∞ にした場合に相当し，このとき σ は $1/\rho$ である．

図1は，入力ベクトルが2次元で，基準ベクトルが2個の場合における学習距離積素子の判断写像の性質を示したもので，判断量が同一となる入力ベクトルの集合を一つの曲線で表している．学習距離素子では曲線はすべて同心円になるのに対して，学習距離積素子では曲線は円，レムニスケートまたはだ円のような形になる．図の中の $\boldsymbol{w}_1$ と $\boldsymbol{w}_2$ はそれぞれ第1基準ベクトルと第2基準ベクトル，数値は判断量である．但し，$\boldsymbol{w}_1$ と $\boldsymbol{w}_2$ のユークリッド距離は2としてある．

3. ネットワークの学習則

連続情報型学習しきい素子を用いた階層型ネットワークに確率的降下法を適用したとき得られる学習則が誤差逆伝搬学習則である．本論文では，連続情報型学習距離積素子を用いた階層型ネットワークにも同様に確率的降下法を適用し，ネットワークの学習則を導出する．

そのためまず，層数が M の階層型ネットワークを考える．このネットワークの第1層の各素子は，式(7)に示すように，それぞれ異なる一つの入力信号を受け取り，それをそのまま出力する．第2層から第 M 層までの各素子は，すべて連続情報型学習距離積素子で，その入出力特性は式(8)，(9)で表される．

$$z\langle{}^{1}_{i}\rangle=x_i \tag{7}$$

$$u\langle{}^{L}_{j}\rangle=\prod_n\sqrt{\sum_i(w_n\langle{}^{L-1}_{i}\,{}^{L}_{j}\rangle-z\langle{}^{L-1}_{i}\rangle)^2}\qquad(L=2,3,\cdots,M) \tag{8}$$

$$z\langle{}^{L}_{j}\rangle=\exp(-(\rho\langle{}^{L}_{j}\rangle^2u\langle{}^{L}_{j}\rangle^2)^A)\quad(L=2,3,\cdots,M) \tag{9}$$

ここで，x_i はネットワークの入力ベクトルの第 i 成分である．u，ρ，z は，各素子の判断量，選択の鋭さ，出力信号である．括弧$\langle\ \rangle$の上段は層の番号，下段は層内での素子の番号を示す．$w_n\langle{}^{L-1}_{i}\,{}^{L}_{j}\rangle$は第 L 層第 j 素子の第 n 基準ベクトルの第 i 成分である．

次に，以上述べたネットワークの評価関数を式(10)に示す誤差2乗和とし，確率的降下法を適用する．その結果，式(11)，(12)，(13)，(14)で表される学習則が得られる．

$$E=\frac{1}{2}\sum_j(y_j-z\langle{}^{M}_{j}\rangle)^2 \tag{10}$$

$$\begin{aligned}&\Delta w_n\langle{}^{L-1}_{i}\,{}^{L}_{j}\rangle^{(\tau+1)}\\&\quad=2A\alpha r\langle{}^{L}_{j}\rangle\rho\langle{}^{L}_{j}\rangle^{2A}\Bigl[\prod_n\Bigl\{\sum_i(w_n\langle{}^{L-1}_{i}\,{}^{L}_{j}\rangle\\&\qquad-z\langle{}^{L-1}_{i}\rangle)^2\Bigr\}^{A-1}\Bigr]\Bigl\{\sum_i(w_n\langle{}^{L-1}_{i}\,{}^{L}_{j}\rangle\\&\qquad-z\langle{}^{L-1}_{i}\rangle)^2\Bigr\}(w_n\langle{}^{L-1}_{i}\,{}^{L}_{j}\rangle-z\langle{}^{L-1}_{i}\rangle)\\&\qquad+\beta\Delta w_n\langle{}^{L-1}_{i}\,{}^{L}_{j}\rangle^{(\tau)}\end{aligned} \tag{11}$$

$$\begin{aligned}&\Delta\rho\langle{}^{L}_{j}\rangle^{(\tau+1)}\\&\quad=2A\alpha r\langle{}^{L}_{j}\rangle\rho\langle{}^{L}_{j}\rangle^{2A-1}\prod_n\Bigl\{\sum_i(w_n\langle{}^{L-1}_{i}\,{}^{L}_{j}\rangle\\&\qquad-z\langle{}^{L-1}_{i}\rangle)^2\Bigr\}^{A}+\beta\Delta\rho\langle{}^{L}_{j}\rangle^{(\tau)}\end{aligned} \tag{12}$$

$$r\langle{}^{M}_{j}\rangle=-z\langle{}^{M}_{j}\rangle(y_i-z\langle{}^{M}_{j}\rangle) \tag{13}$$

$$r\langle{}^{L}_{j}\rangle=2Az\langle{}^{L}_{j}\rangle$$

$$\cdot \sum_k r\langle {}^{L+1}_{k}\rangle \rho\langle {}^{L+1}_{k}\rangle^{2A} \left[\prod_n \left\{\sum_j (w_n\langle {}^{L}_{j}\,{}^{L+1}_{k}\rangle - z\langle {}^{L}_{j}\rangle)^2\right\}^{A-1}\right] \sum_m (w_m\langle {}^{L}_{j}\,{}^{L+1}_{k}\rangle - z\langle {}^{L}_{j}\rangle) \cdot \prod_{n\neq m}\sum_j (w_n\langle {}^{L}_{j}\,{}^{L+1}_{k}\rangle - z\langle {}^{L}_{j}\rangle)^2 \qquad (14)$$

ここで，y_j は第 M 層第 j 素子に与えられる教師信号，$\Delta w_n\langle {}^{L-1}_{i}\,{}^{L}_{j}\rangle^{(\tau)}$ は，$w_n\langle {}^{L-1}_{i}\,{}^{L}_{j}\rangle$ の τ 回目の修正量，$\Delta\rho\langle {}^{L}_{j}\rangle^{(\tau)}$ は $\rho\langle {}^{L}_{j}\rangle$ の τ 回目の修正量，$r\langle {}^{L}_{j}\rangle$ は第 L 層第 j 素子の強化信号である．α は強化係数，β はモーメント係数でともに正の実数である．

4. 計算機シミュレーション

4.1 方　　法

2入力1出力の3層ネットワークにおいて，学習距離積素子を用いた場合と学習しきい素子を用いた場合とについて学習の計算機シミュレーションを行い，式(10)に示した誤差2乗和を求める．但し，学習距離積素子の基準ベクトルは2個，A は1とする．ネットワークの入力ベクトルの各成分と教師信号は0または1とする．第2層の素子数は1から5まで変化させる．

学習は，入力ベクトルの集合を定義域とする16個のすべての論理関数について行う．一つの論理関数の学習においては，入力ベクトルと教師信号の対を全入力ベクトルについて一定の順序で提示する．この過程を1学習サイクルとし，これを10,000回繰り返す．誤差2乗和は100サイクルごとに求める．但し，各素子のパラメータの初期値は一つの論理関数について5回変えるので，最終的な誤差2乗和はそれら5回の平均となる．

パラメータの初期値は，学習距離積素子の第1基準ベクトルの第1成分については−0.2から0.2までの一様乱数，第2成分については0.5に−0.2から0.2までの一様乱数を加えた値，第2基準ベクトルの第1成分については1に−0.2から0.2までの一様乱数を加えた値，第2成分については0.5に−0.2から0.2までの一様乱数を加えた値とする．また，選択の鋭さについては1.6とする．学習しきい素子の基準ベクトルの各成分としきい値についてはすべて−0.3から0.3までの一様乱数とする．

強化係数 α は，学習距離積素子では0.00001，0.0001，0.001，0.01，0.1の5通り，学習しきい素子では0.0001，0.001，0.01，0.1，1の5通り変え，10,000サイクル目の誤差2乗和の全関数についての総和が最小となる α の値を求める．次に，その値を中心に α を更に細かく変化させ，それぞれの場合の誤差2乗和の総和を求める．モーメント係数 β については0.9に固定する．

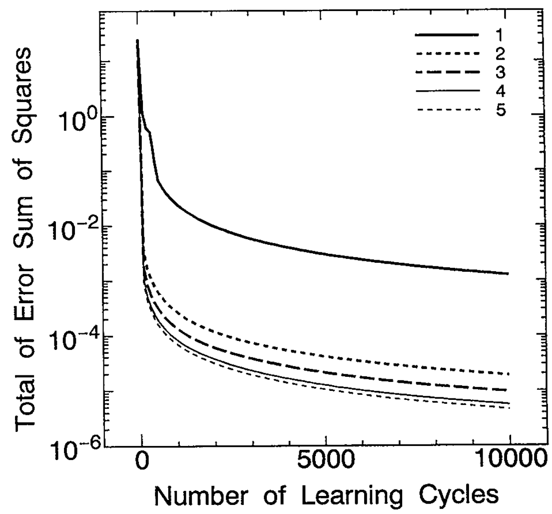

図2　学習距離積素子を用いた場合における学習サイクル数の関数として表された誤差2乗和の総和（数字は第2層の素子数）

Fig. 2　Total of error sum of squares as a function of the number of learning cycles for the network constructed from learning distance product element. Numerals indicate the number of elements of the second layer.

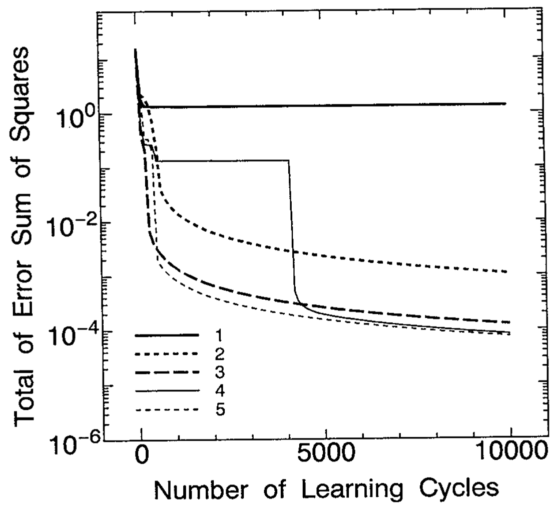

図3　学習しきい素子を用いた場合における学習サイクル数の関数として表された誤差2乗和の総和（数字は第2層の素子数）

Fig. 3　Total of error sum of squares as a function of the number of learning cycles for the network constructed from learning threshold element. Numerals indicate the number of elements of the second layer.

4.2 結果と考察

10,000サイクル目の誤差2乗和の総和は，学習距離積素子では α が0.01のとき，学習しきい素子では α が

表1 第2層の素子数に対する α の値

第2層の素子数	α	
	学習距離積素子	学習しきい素子
1	0.04	0.7
2	0.06	0.3
3	0.06	2
4	0.08	3
5	0.08	3

1のとき最小となった．そこで，前者の場合には α を0.002から0.009まで0.001間隔で8通り，0.02から0.09まで0.01間隔で8通り変化させ，後者の場合には α を0.2から0.9まで0.1間隔で8通り，2から9まで1間隔で8通り変化させ，それぞれの α の値における誤差2乗和の総和を求めた．

図2と図3は，それぞれ学習距離積素子を用いた場合と学習しきい素子を用いた場合とについて，誤差2乗和の総和を学習サイクル数の関数として表したものである．図の中の数字は第2層の素子数である．α の値は，表1に示すように，第2層の素子数それぞれに対して10,000サイクル目の誤差2乗和の総和が最小になるものを選んでいる．

図4と図5は，それぞれ学習距離積素子を用いた場合と学習しきい素子を用いた場合とについて，α の値を固定したときの誤差2乗和の総和を学習サイクル数の関数として表したものである．但し，学習距離積素子では α を0.04，学習しきい素子では2とした．

2から5までの図はすべて，第2層の素子数が同じであれば学習距離積素子を用いた場合の方が学習しきい素子を用いた場合より10,000サイクル目の誤差2乗和の総和が小さくなる，言い換えれば前者は後者に比べて学習が平均的に加速されることを示している．特に第2層の素子数が1個の場合には両者の違いは大きい．このことは，個々の関数ごとに比較しても全く同様である．ただ，パラメータの数が A を除いて数えれば学習距離積素子では5個，学習しきい素子では3個と異なっているので，以上述べたことだけから両者の優劣について直ちに結論を下すことはできない．

そこで，第2層の素子数にこだわらず，どこまで学習が加速されるかという観点から，両者を比較してみる．まず学習しきい素子を用いた場合には，第2層の素子数が4個程度になると素子数をそれ以上増やしても学習の加速はほとんど見られなくなる．それに対し

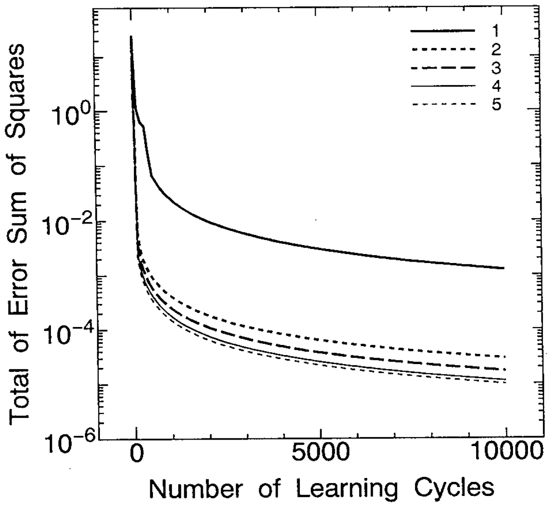

図4 学習距離積素子を用いた場合における学習サイクル数の関数として表された誤差2乗和の総和（数字は第2層の素子数，α は0.04に固定）

Fig. 4 Total of error sum of squares as a function of the number of learning cycles for the network constructed from learning distance product element. Numerals indicate the number of elements of the second layer. α is fixed at 0.04.

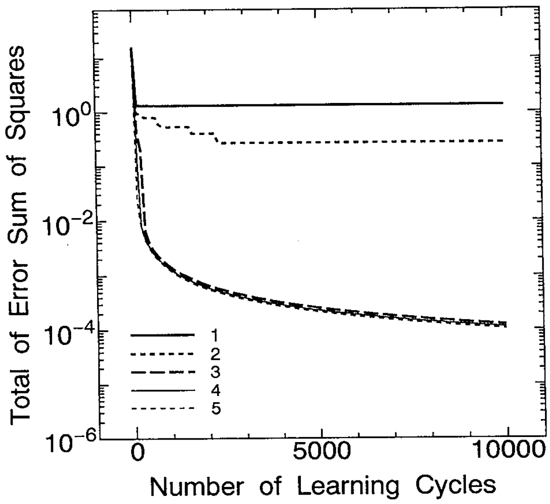

図5 学習しきい素子を用いた場合における学習サイクル数の関数として表された誤差2乗和の総和（数字は第2層の素子数，α は2に固定）

Fig. 5 Total of error sum of squares as a function of the number of learning cycles for the network constructed from learning threshold element. Numerals indicate the number of elements of the second layer. α is fixed at 2.

学習距離積素子を用いた場合には，やはり同じような傾向が現れはするが，第2層の素子数を2個以上にすれば，第2層の素子数にかかわらず学習しきい素子を用いた場合に比べて学習は必ず加速される．このことは，個々の関数ごとに調べても全く同様である．従って，学習の高速化の点からは，階層型ネットワークに

は学習しきい素子よりも学習距離積素子を用いた方がよい.

ただ，以上述べた結論はあくまでも2入力1出力の3層ネットワークに関して得られたもので，階層型ネットワーク一般に対して言えるかどうかはまだ明らかではない．今後，入力ベクトルの次元や層数をもっと多くして，同様の比較を行う必要がある．また，学習距離積素子のパラメータの初期値については，より最適な値が存在する可能性があるので，今後更に検討していく必要がある．なお，学習距離積素子を用いてネットワーク構成した場合，学習しきい素子を用いた場合よりも汎化能力が向上する可能性がある．この点についても調べる必要があろう．また，学習の高速化にしろ汎化能力向上の可能性にしろ，計算機シミュレーションだけでなく，理論的な面からも検討を行っていかなければならない.

5. むすび

本論文では，素子の入出力特性の改良の方向から並列分散処理ネットワークの学習をより高速化することを目的に，従来の学習しきい素子に代わって学習距離積素子を新たに提案した．次に，この素子を用いた階層型ネットワークに確率的降下法を適用することにより，ネットワークの学習則を導き出した．最後に，学習距離積素子を用いた2入力1出力の3層ネットワークで計算機シミュレーションを行った．その結果，学習しきい素子を用いた場合に比べて，学習が平均的に加速されることが確認された.

今後の課題としては，階層型ネットワークの入力ベクトルの次元や層数をもっと多くして，学習距離積素子を用いた場合と学習しきい素子を用いた場合との比較を行うことが第1に挙げられる．第2は，学習距離積素子のパラメータの初期値についての検討である．第3は，学習距離積素子を用いることによる階層型ネットワークの学習の高速化と汎化能力向上の可能性について，理論的な面から検討を行っていくことである.

文 献

(1) 永田明徳，横井博一：“距離情報を元にした素子のネットワーク”，信学技報，**NC91**-45 (1991).

(2) 村山聡敏，永田明徳，横井博一：“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 (1992).

(3) Moody J. and Darken C. J.: “Fast learning in network of locally-tuned processing units”, Neural Computation, **1**, pp. 281-294 (1989).

(4) Poggio T. and Girosi F.: “Regularization algorithms for learning that are equivalent to multilayer networks”, Science, **247**, pp. 978-982 (1990).

(5) Poggio T. and Girosi F.: “Networks for approximation and learning”, Proc. IEEE, **78**, 9, pp. 1481-1497 (1990).

(平成5年4月12日受付，7月19日再受付)