

## EBP-EWLS 学習アルゴリズムの応用 ～正弦関数と文字認識～

山本祥弘・坂本和洋<sup>\*1</sup>・松田直樹<sup>\*1</sup>  
高塚峰人<sup>\*1</sup>・奥谷拓生<sup>\*2</sup>・福本壮志<sup>\*3</sup>

知能情報工学科・<sup>\*1</sup>鳥取大学大学院  
<sup>\*2</sup>松江計算センター・<sup>\*3</sup>富士通関西通信システム

### Applications of the EBP-EWLS Learning Algorithm ～ Sinusoidal Function and Literal Recognitions ～

Yoshihiro YAMAMOTO, Kazuhiro SAKAMOTO<sup>\*1</sup>, Naoki MATSUDA<sup>\*1</sup>  
Mineto TAKATSUKA<sup>\*1</sup>, Takuo OKUTANI<sup>\*2</sup> and Tsuyoshi FUKUMOTO<sup>\*3</sup>

Department of Information and Knowledge Engineering, Faculty of Engineering  
Tottori University, Tottori, 680-8552 Japan

E-mail yamamoto@ike.tottori-u.ac.jp

<sup>\*1</sup> Graduate School of Engineering, Tottori University, 680-8552, Japan

<sup>\*2</sup> Matsue Computing Center, Matsue, Shimane, 690-0835 Japan

<sup>\*3</sup> Fujitsu Kansai Communication Systems, Suita, Osaka, 565-0816 Japan

Abstract : A new error back propagation (EBP) learning algorithm proposed by the authors is a method to assign a fictitious teacher signal for each output of hidden units by an error back propagation method. Many methods can be considered for updating the weight parameters. The EBP-EWLS algorithm using an exponential weighted least squares (EWLS) method for the weight corrections is examined by the examples of the sinusoidal function learning and the pattern recognition of Arabic numerals.

Key words: Error back propagation, Learning algorithm, Multilayer neural network, Exponentially weighted least squares method, Sinusoidal function, Literal recognitions

### 1. はじめに

階層型ニューラルネットワーク(NN)に対する教師あり学習アルゴリズムとしては誤差逆伝搬法(BP法)[1]およびその拡張である慣性項を用いたBP法[2]を始めとして非常に多くの提案[3]～[5]がなされている。しかし、BP法には多くの欠陥が指摘されており、その改良に多くの努力が払われているが、勾配法によるBP法を基にした改良では、限界があると思われる。そこで、筆者らはこの勾配法から脱却した新しい学習アルゴリズム[6]～[10]を提案してきた。これは最小二乗法による解であり、この流れは、適応制御の適応アルゴリズムの歴史とまさに一致する。すなわち、適応制御においても初期の頃には勾配法がその適応アルゴリズムに用いられていたが、勾配法ゆえに派生する局所性の問題から脱皮する一連の研究を経

て、現在では、最小二乗法によるアルゴリズムが知られている[11]。問題は、どのように最小二乗法を適用するかである。

提案する方法は次の二つから構成される。

- A) 各中間層への仮の教師信号を、NNの出力誤差から順に決定する(誤差逆伝搬法: EBP)。
- B) 与えられた仮の教師信号との誤差を減少(零と)するように重みパラメータを修正する(最小二乗法)。

ここで A)のEBP法によりNNの出力のみならず各中間層の出力に対しても仮の教師信号が与えられるので、階層型NNは2層NNの積み重ねと考えられる。この結果、B)の重み修正法には種々の方法が考えられる。仮の教師信号との誤差を活性化関数の逆関数を用いて重みパラメータの空間で考えることによりこの B)の問題は線形問題となる。そこで重みパラメータの現時点での推定

値から解集合を表す超平面への直交射影法を用いる逐次型 EBP-OP 法、および任意個数のデータを一括処理する一括型 EBP-OP 法を提案してきた。しかしながら先に記した適応制御での最小 2 乗法等では、解の評価は一括的であり、そのパラメータの更新処理は逐次的である最小 2 乗法によるアルゴリズムが知られている。学習に対しても同様な特性を持つアルゴリズムが開発されればオンラインでの学習には適切なものとなる。そこで本論文では最小 2 乗法の一つである指数重み付き最小 2 乗(EWLS)法を用いた EBP- EWLS アルゴリズムを提案し、正弦関数の学習と文字認識の学習に応用した結果を述べる。正弦関数の学習に関してはすでに参考文献[12]で発表されているが、ここでは正弦波に高調波が加わった場合の特性を調べている。一方文字認識に関してはまだ基礎的な段階であり、0～9 の数字の学習結果を報告する。

## 2. 学習アルゴリズムの導出

### 2.1 ニューラルネットワーク

本論では、以下の 3 層 NN を考える。

$$\mathbf{c}_k = f(\mathbf{z}_k), \quad \mathbf{z}_k = \bar{\mathbf{W}}_{k-1}^T \bar{\mathbf{a}}_k, \quad (1a)$$

$$\mathbf{a}_k = f(\mathbf{y}_k), \quad \mathbf{y}_k = \bar{\mathbf{V}}_{k-1}^T \bar{\mathbf{b}}_k. \quad (1b)$$

$$\begin{aligned} \bar{\mathbf{a}}_k^T &= (1, \mathbf{a}_k^T), \quad \bar{\mathbf{W}}_k^T = (\mathbf{w}_{0k}, \mathbf{W}_k^T), \\ \bar{\mathbf{b}}_k^T &= (1, \mathbf{b}_k^T), \quad \bar{\mathbf{V}}_k^T = (\mathbf{v}_{0k}, \mathbf{V}_k^T). \end{aligned} \quad (1c)$$

ここに  $\mathbf{c}, \mathbf{b}$  は NN の  $p$  次元出力と  $n$  次元入力であり、 $\mathbf{a}$  は  $m$  次元中間層出力である。また  $\mathbf{z}, \mathbf{y}$  はそれぞれ出力ユニット、中間層ユニットの内部状態を表し、 $\mathbf{W}, \mathbf{V}$  は重みパラメータを表す行列であり、そのサイズは適切なものとする。添え字  $k$  は学習データの番号であるが、動的システムを対象のときは時刻と考えてよい。各ユニットの活性化関数  $f(\cdot)$  は連続単調増加の非線形関数であり、逆関数が存在するものとする。 $\mathbf{w}_{0k}$  と  $\mathbf{v}_{0k}$  はしきい値であり、上付きバーはしきい値を含むことを示している。重みパラメータ  $\mathbf{W}$  と  $\mathbf{V}$  の調整を、 $\mathbf{d}_k$  を教師信号(目標信号)として、

$$E_k = (\mathbf{d}_k - \mathbf{c}_k)^T (\mathbf{d}_k - \mathbf{c}_k) \quad (2)$$

を最小とすることを目的とする場合が逐次処理法であり、

$$E_k = \sum_{j=0}^{M-1} (\mathbf{d}_{k-j} - \mathbf{c}_{k-j})^T (\mathbf{d}_{k-j} - \mathbf{c}_{k-j}) \quad (3)$$

を最小とすることを目的とする場合が一括処理法である。それに対して本論文では

$$E = \sum_{j=0}^{k-1} \rho^j (\mathbf{d}_{k-j} - \mathbf{c}_{k-j})^T (\mathbf{d}_{k-j} - \mathbf{c}_{k-j}) \quad (4)$$

を最小とすることを目的とする場合を新たに提案する。これは指数重み付き最小 2 乗法として知られているが、この結果、評価は一括的でありパラメータ調整は逐次的であることが特徴である。

### 2.2 逐次修正法

(1)式において  $\mathbf{W}, \mathbf{V}$  の修正量をそれぞれ  $\Delta \mathbf{W}, \Delta \mathbf{V}$  とする。このとき(1-2)式から、ベクトル  $\mathbf{a}$  に  $\Delta \mathbf{a}$  の修正がなされ、(1-1)式より

$$\mathbf{c}_k + \Delta \mathbf{c}_k = f(\mathbf{z}_k + \Delta \mathbf{z}_k), \quad (5a)$$

$$\begin{aligned} \mathbf{z}_k + \Delta \mathbf{z}_k \\ = (\bar{\mathbf{W}}_{k-1} + \Delta \bar{\mathbf{W}}_{k-1})^T (\bar{\mathbf{a}}_k + \Delta \bar{\mathbf{a}}_k) \end{aligned} \quad (5b)$$

$$\Delta \bar{\mathbf{a}}_k^T = (0, \Delta \mathbf{a}_k^T) \quad (5c)$$

となる。すなわち、 $\mathbf{z}$ 、従って  $\mathbf{c}$  の値は  $\bar{\mathbf{W}}$  と  $\bar{\mathbf{a}}$  の修正を受けることになる。逆に、出力(5a)式の値が(2)式を最小にするため、すなわち  $\mathbf{d}_k$  と一致するために必要な修正量  $\Delta \mathbf{c}_k$  から、(5b)式を満たす  $\Delta \bar{\mathbf{W}}_{k-1}$  と  $\Delta \bar{\mathbf{a}}_k$  を求めようとするのが、EBP 学習アルゴリズムの考え方である。また本方法では、(2)式を減少させるのではなく、その値をゼロとするように決めるのも一つの特徴である。そのためには、

$$\mathbf{c}_k + \Delta \mathbf{c}_k = \mathbf{d}_k \quad (6)$$

を満たすことであるが、これはまた、

$$\mathbf{z}_k + \Delta \mathbf{z}_k = f^{-1}(\mathbf{d}_k) \quad (7)$$

と等価である。ただし、関数  $f$  の逆関数の存在を仮定している。以上より、

$$(\bar{\mathbf{W}}_{k-1} + \Delta \bar{\mathbf{W}}_{k-1})^T (\bar{\mathbf{a}}_k + \Delta \bar{\mathbf{a}}_k) = f^{-1}(\mathbf{d}_k) \quad (8)$$

が条件式となる。ここで、 $\Delta \bar{\mathbf{W}}$  は任意であるが、 $\Delta \bar{\mathbf{a}}$  には制限が付く。すなわち、新しい  $\mathbf{a}_{Nk}$ 、

$$\mathbf{a}_{Nk}' = \mathbf{a}_k + \Delta \mathbf{a}_k \quad (9)$$

は次の層(1b)式にたいする教師信号の役割をすることになるので、 $\mathbf{a}_{Nk}'$  は非線形関数  $f(\cdot)$  の値域に属することが必要となる。そこでまず仮の教師信号を求めるために、(8)式において  $\Delta \mathbf{W} = \mathbf{0}$

とした

$$\bar{W}_{k-1}^T(\bar{\mathbf{a}}_k + \Delta \bar{\mathbf{a}}_k) = f^{-1}(\mathbf{d}_k) \quad (10)$$

から  $\Delta \bar{\mathbf{a}}_k$  を決定する。(10)式は

$$\mathbf{W}_{k-1}^T \Delta \bar{\mathbf{a}}_k = f^{-1}(\mathbf{d}_k) - \bar{W}_{k-1}^T \bar{\mathbf{a}}_k \quad (11)$$

と表される。ここに(5c)式の関係を利用している。(11)式の右辺はNNの出力誤差であり、仮の教師信号が出力誤差の逆伝搬 (Error back propagation, EBP) で定まることになる。従ってこの方法から得られるアルゴリズムを EBP アルゴリズムと総称している。ちなみに BP 法として良く知られている従来の誤差逆伝搬法は、出力誤差の感度を逆伝搬しているにすぎない。行列  $\mathbf{W}_{k-1}$  は一般に正則でないがフルランクを仮定する。このとき(11)式の解を求めることができる。この結果式(1a),(1b)それぞれに教師信号が与えられ、二組の2層NN問題として、重み  $\mathbf{W}$  と  $\mathbf{V}$  の修正を行うことができる。このとき重みの修正法として直交射影 (Orthogonal Projection: OP) 法を用いるのが以下のアルゴリズムである。

### [逐次型EBP-OP学習アルゴリズム]

$$1) \Delta \bar{\mathbf{a}}_k = \mathbf{W}_{k-1}(\mathbf{W}_{k-1}^T \mathbf{W}_{k-1})^{-1} \mathbf{e}_k \quad (12a)$$

$$\Delta \bar{\mathbf{a}}_k = (\mathbf{W}_{k-1} \mathbf{W}_{k-1}^T)^{-1} \mathbf{W}_{k-1} \mathbf{e}_k \quad (12a')$$

$$\mathbf{e}_k = f^{-1}(\mathbf{d}_k) - \bar{W}_{k-1}^T \bar{\mathbf{a}}_k \quad (12b)$$

$$2) \mathbf{a}_{Nk} = \bar{\mathbf{a}}_k + \Delta \bar{\mathbf{a}}_k \quad (13)$$

$$3) \mathbf{a}_{Nk} = \mathbf{H} * \mathbf{a}_{Nk}' \quad (14)$$

$$4) \Delta \bar{\mathbf{V}}_{k-1} = \bar{\mathbf{b}}_k (\bar{\mathbf{b}}_k^T \bar{\mathbf{b}}_k)^{-1} \mathbf{e}_{vk}^T \quad (15a)$$

$$\mathbf{e}_{vk} = f^{-1}(\mathbf{a}_{Nk}) - \bar{V}_{k-1}^T \bar{\mathbf{b}}_k \quad (15b)$$

$$5) \bar{\mathbf{V}}_k = \bar{V}_{k-1} + \Delta \bar{V}_{k-1} \quad (16)$$

$$6) \Delta \bar{\mathbf{W}}_{p-1} = \bar{\mathbf{a}}_{Nk} (\bar{\mathbf{a}}_{Nk}^T \bar{\mathbf{a}}_{Nk})^{-1} \mathbf{e}_{wk}^T \quad (17a)$$

$$\mathbf{e}_{wk} = f^{-1}(\mathbf{d}_k) - \bar{W}_{k-1}^T \bar{\mathbf{a}}_{Nk} \quad (17b)$$

$$7) \bar{\mathbf{W}}_k = \bar{W}_{k-1} + \Delta \bar{W}_{k-1} \quad (18)$$

このアルゴリズムにおいて 1)の(12a)式は  $p \leq m$  の場合であり、(12b)式は  $p \geq m$  の場合である。また、(15),(17)式は現時点のみのデータに基づく場合であり、直交射影法と呼んでいる。すなわち(17)式は、超平面

$$f^{-1}(\mathbf{d}_k) = \bar{W}^T \bar{\mathbf{a}}_k \quad (19)$$

への直交射影となっている。(15)式も同様である。

また、蓄積された過去のデータをも利用する方法として、正規直交化法[4]を提案している。一方、過去のデータを一括して処理する方法は次節で示す。ところで、(11)式の意味は、

$$\mathbf{a}_{Nk} \in \text{range}(\mathbf{f}(\mathbf{y})) \quad (20)$$

とすることである。  $\mathbf{f}(\mathbf{y})$  の値域は通常区間で与えられるので、その方法として、

H1) 値域を越えた分をカットする方法。

H2)  $\mathbf{f}(\mathbf{0})$  を中心にして  $\mathbf{a}_{Nk}$  を縮小する方法。

の二つが考えられる。また、非線形関数  $f$  として、

$$\text{F1) } f(y) = \frac{1}{1 + \exp(-my)} \quad (21)$$

$$\text{F2) } g(y) = \frac{1 - \exp(-my)}{1 + \exp(-my)} \quad (22)$$

を考える。ただし通常は  $m = 1$  とする。(21)式はシグモイド関数と呼ばれており、その値域は開区間(0,1)である。一方(22)式は双曲線関数の一つで  $\tanh(my/2)$  であり、その値域は開区間(-1, 1)である。

### 2.3 一括処理法

(1)式は  $k$  番目の学習データにたいする応答を表しているが、本節では、このデータを  $M$  組一括して処理することを考える。そのために以下の記号を導入する。

$$\mathbf{C}_k = (\mathbf{c}_k, \mathbf{c}_{k+1}, \dots, \mathbf{c}_{k+M-1}) \quad (23a)$$

$$\mathbf{Z}_k = (\mathbf{z}_k, \mathbf{z}_{k+1}, \dots, \mathbf{z}_{k+M-1}) \quad (23b)$$

$$\mathbf{A}_k = (\mathbf{a}_k, \mathbf{a}_{k+1}, \dots, \mathbf{a}_{k+M-1}) \quad (23c)$$

$$\mathbf{Y}_k = (\mathbf{y}_k, \mathbf{y}_{k+1}, \dots, \mathbf{y}_{k+M-1}) \quad (23d)$$

$$\mathbf{B}_k = (\mathbf{b}_k, \mathbf{b}_{k+1}, \dots, \mathbf{b}_{k+M-1}) \quad (23e)$$

$$\mathbf{D}_k = (\mathbf{d}_k, \mathbf{d}_{k+1}, \dots, \mathbf{d}_{k+M-1}) \quad (23f)$$

上付きバーが付いたのも同様である。このとき、(1)式に対応して、

$$\mathbf{C}_k = \mathbf{f}(\mathbf{Z}_k), \mathbf{Z}_k = \bar{W}_{k-1}^T \bar{\mathbf{A}}_k, \quad (24a)$$

$$\mathbf{A}_k = \mathbf{f}(\mathbf{Y}_k), \mathbf{Y}_k = \bar{V}_{k-1}^T \bar{\mathbf{B}}_k \quad (24b)$$

と表され、このときの学習の目的は

$$\mathbf{D}_k = \mathbf{C}_k \quad (25)$$

あるいは等価な

$$\mathbf{f}^{-1}(\mathbf{D}_k) = \mathbf{Z}_k \quad (26)$$

である。このとき逐次学習アルゴリズムと同様にして、以下の結果が導かれる。

### [一括型EBP-OP学習アルゴリズム]

1)  $p \leq m$  のとき

$$\Delta \mathbf{A}_k = \mathbf{W}_{k-1} (\mathbf{W}_{k-1}^T \mathbf{W}_{k-1})^{-1} \mathbf{E}_{a_k} \quad (27a)$$

$p \geq m$  のとき

$$\Delta \mathbf{A}_k = (\mathbf{W}_{k-1} \mathbf{W}_{k-1}^T)^{-1} \mathbf{W}_{k-1} \mathbf{E}_{a_k} \quad (27a)$$

$$\mathbf{E}_{a_k} = \mathbf{f}^{-1}(\mathbf{D}_k) - \bar{\mathbf{W}}_{k-1}^T \bar{\mathbf{A}}_k \quad (27b)$$

$$2) \mathbf{A}_{Nk}' = \mathbf{A}_k + \Delta \mathbf{A}_k \quad (28)$$

$$3) \mathbf{A}_{Nk} = \mathbf{H} * \mathbf{A}_{Nk}' \quad (29)$$

4)  $M \leq n + 1$  のとき

$$\Delta \bar{\mathbf{V}}_{k-1} = \bar{\mathbf{B}}_k (\bar{\mathbf{B}}_k^T \bar{\mathbf{B}}_k)^{-1} \mathbf{E}_{v_k} \quad (30a)$$

$M \geq n + 1$  のとき

$$\Delta \bar{\mathbf{V}}_{k-1} = (\bar{\mathbf{B}}_k \bar{\mathbf{B}}_k^T)^{-1} \bar{\mathbf{B}}_k \mathbf{E}_{v_k} \quad (30a)$$

$$\mathbf{E}_{v_k} = \mathbf{f}^{-1}(\mathbf{A}_{Nk}) - \bar{\mathbf{V}}_{k-1}^T \bar{\mathbf{B}}_k \quad (31)$$

$$5) \bar{\mathbf{V}}_k = \bar{\mathbf{V}}_{k-1} + \Delta \bar{\mathbf{V}}_{k-1} \quad (32)$$

6)  $M \leq m + 1$  のとき

$$\Delta \bar{\mathbf{W}}_{k-1} = \bar{\mathbf{A}}_{Nk} (\bar{\mathbf{A}}_{Nk}^T \bar{\mathbf{A}}_{Nk})^{-1} \mathbf{E}_{w_k} \quad (33a)$$

$M \geq m + 1$  のとき

$$\Delta \bar{\mathbf{W}}_{k-1} = \bar{\mathbf{A}}_{Nk} (\bar{\mathbf{A}}_{Nk} \bar{\mathbf{A}}_{Nk}^T)^{-1} \mathbf{E}_{w_k} \quad (33a)$$

$$\mathbf{E}_{w_k} = \mathbf{f}^{-1}(\mathbf{D}_k) - \bar{\mathbf{W}}_{k-1}^T \bar{\mathbf{A}}_k \quad (33b)$$

$$7) \bar{\mathbf{W}}_{pk} = \bar{\mathbf{W}}_{k-1} + \Delta \bar{\mathbf{W}}_{k-1} \quad (34)$$

(19)式以下の注意がこの場合にも適用される。ただしこの場合には、行列 $\mathbf{B}_k$ のフルランクを仮定している。問題は行列 $\bar{\mathbf{A}}_k$ のフルランクを仮定できないことである。そのために、このフルランク性を満たすように、学習データの選別[13]が必要となる。

## 2.4 EBP-EWLSアルゴリズム

2.2節で述べた逐次型修正法では各学習時にその時点でのデータのみを用いているので、データ処理としては一般に非効率である。また、2.3節の一括型はその処理方法が煩雑であり、またオンラインでの処理には向いていない。そこで、両者の利点を兼ね備えたものとして、学習パラメータの修正に指数重み付き最小2乗法を用いる方

法を以下に述べる。これは、学習の評価に関しては一括的であり、学習パラメータの更新は逐次的となっている。各学習毎に中間層出力に対する仮の教師信号が(12)～(14)式で与えられたとする。このとき3層NNは二組の2層NNから構成されるので、NNの出力誤差と中間層の出力誤差に対する評価を以下の式で与える。まずNNの出力誤差に対しては(4)式と等価な

$$\mathbf{E}_{w,k} = \sum_{j=0}^{k-1} \rho^j \mathbf{e}_{w,k-j}^T \mathbf{e}_{w,k-j} \quad (35)$$

$$\mathbf{e}_{w,k} = \mathbf{f}^{-1}(\mathbf{d}_k) - \bar{\mathbf{W}}^T \bar{\mathbf{a}}_k^* \quad (36)$$

であり、中間層出力に対しては

$$\mathbf{E}_{v,k} = \sum_{j=0}^{k-1} \rho^j \mathbf{e}_{v,k-j}^T \mathbf{e}_{v,k-j} \quad (37)$$

$$\mathbf{e}_{v,k} = \mathbf{f}^{-1}(\mathbf{a}_{Nk}) - \bar{\mathbf{V}}_{k-1}^T \bar{\mathbf{b}}_k \quad (38)$$

を評価とする。ただし $\mathbf{a}_k^*$ はV修正後の実際の中間層出力とする。このとき逐次型EBP-OP学習アルゴリズムと同様にして、以下のアルゴリズムが得られる。

### [EBP-EWLS学習アルゴリズム]

$$1) \Delta \mathbf{a}_k = \mathbf{W}_{k-1} (\mathbf{W}_{k-1}^T \mathbf{W}_{k-1})^{-1} \mathbf{e}_k \quad (39a)$$

$$\Delta \mathbf{a}_k = (\mathbf{W}_{k-1} \mathbf{W}_{k-1}^T)^{-1} \mathbf{W}_{k-1} \mathbf{e}_k \quad (39a)$$

$$\mathbf{e}_k = \mathbf{f}^{-1}(\mathbf{d}_k) - \bar{\mathbf{W}}_{k-1}^T \bar{\mathbf{a}}_k \quad (39b)$$

$$2) \mathbf{a}_{Nk}' = \mathbf{a}_k + \Delta \mathbf{a}_k \quad (40)$$

$$3) \mathbf{a}_{Nk} = \mathbf{H} * \mathbf{a}_{Nk}' \quad (41)$$

$$4) \Delta \bar{\mathbf{V}}_{k-1} = \frac{\mathbf{P}_{k-1} \mathbf{b}_k}{\rho + \bar{\mathbf{b}}_k^T \mathbf{P}_{k-1} \bar{\mathbf{b}}_k} \mathbf{e}_{v,k}^T \quad (42a)$$

$$\mathbf{P}_k = \frac{1}{\rho} (\mathbf{P}_{k-1} - \frac{\mathbf{P}_{k-1} \bar{\mathbf{b}}_k \bar{\mathbf{b}}_k^T \mathbf{P}_{k-1}}{\rho + \bar{\mathbf{b}}_k^T \mathbf{P}_{k-1} \bar{\mathbf{b}}_k}) \quad (42b)$$

$$\mathbf{e}_{v,k} = \mathbf{f}^{-1}(\mathbf{a}_{Nk}) - \bar{\mathbf{V}}_{k-1}^T \bar{\mathbf{b}}_k \quad (42c)$$

$$5) \bar{\mathbf{V}}_k = \bar{\mathbf{V}}_{k-1} + \Delta \bar{\mathbf{V}}_{k-1} \quad (43)$$

$$6) \Delta \bar{\mathbf{W}}_{k-1} = \frac{\mathbf{Q}_{k-1} \mathbf{a}_{Nk}}{\rho + \bar{\mathbf{a}}_{Nk}^T \mathbf{Q}_{k-1} \bar{\mathbf{a}}_{Nk}} \mathbf{e}_{w,k}^T \quad (44a)$$

$$\mathbf{Q}_k = \frac{1}{\rho} (\mathbf{Q}_{k-1} - \frac{\mathbf{Q}_{k-1} \bar{\mathbf{a}}_{Nk} \bar{\mathbf{a}}_{Nk}^T \mathbf{Q}_{k-1}}{\rho + \bar{\mathbf{a}}_{Nk}^T \mathbf{Q}_{k-1} \bar{\mathbf{a}}_{Nk}}) \quad (44b)$$

$$\mathbf{e}_{w,k} = \mathbf{f}^{-1}(\mathbf{d}_k) - \bar{\mathbf{W}}_{k-1}^T \bar{\mathbf{a}}_{Nk} \quad (44c)$$

$$7) \bar{\mathbf{W}}_k = \bar{\mathbf{W}}_{k-1} + \Delta \bar{\mathbf{W}}_{k-1} \quad (45)$$

#### 4. 正弦関数の学習

本節では、先に記したEBP-EWLS学習アルゴリズムを用いて正弦関数の学習を行った結果を述べる。これは、提案するアルゴリズムの有効性を調べると同時に、学習の特性を調べることを目的としている。単に正弦関数を学習した結果は参考文献[12]に詳しく述べられているので、本論文では、正弦波に高調波が加わった場合の特性を調べてみる。すなわち教師信号として

$$y_k = \sin(x_k) + K \sin(n x_k) \quad (46)$$

を考える。ただし、 $K = 0.05, 0.1, 0.2, 0.4$  および  $n = 6, 12, 18, 24$  としている。用いるNNはスカラー入出力であり、種々の中間層ユニット数  $m$  を試みる。出力ユニットの活性化関数は線形であり、中間層ユニットは双曲線関数とする。また、学習データとしての入力  $x_k$  は閉区間  $[0, 2\pi]$  を等間隔に  $T-1$  分割した

$$x_k = \frac{2\pi i}{T-1}, \quad i = 0, 1, \dots, T-1 \quad (47)$$

の  $T$  点とし、これを学習点数と呼ぶ。ただし、 $T = 10, 12, 14, 16, 18, 20$  とする。学習の評価は  $\hat{y}_k$  を NN の出力として

$$E = \frac{1}{2T} \sum_{k=0}^{T-1} (\hat{y}_k - y_k)^2 \quad (48)$$

を最小とすることであり、(47)式の入力とそのときの(46)式の出力を学習データとしている。Kあるいは  $n = 0$  のときの正弦関数の学習において、学習点数が正弦波の一周期に対して最低7が必要であることが文献[12]で調べられている。このことは有名なサンプリング定理からも当然のことである。それでは0でないKおよび  $n$  の値に対してはNNが何を学習するのかが興味の対象となる。また学習点数は多ければ多いほど良いのかも問題となる。従って、(48)式とは別に

$$E_0 = \frac{1}{2T} \sum_{k=0}^{T-1} (\hat{y}_k - \sin(x_k))^2 \quad (49)$$

を設定し、学習が  $E$  および  $E_0$  のどちらを先に小さくするかを考える。すなわち、学習の終了条件を

$$E \text{ or } E_0 < \varepsilon = 0.001 \quad (50)$$

とする。次に、学習結果の汎化能力をみるた

めのテスト誤差を(48)および(49)式で  $T = 100$  とした  $E_t$  および  $E_{t0}$  を定義する。以上の準備のもとで行った種々の数値計算結果を以下に記す。ただし、用いたアルゴリズムはEBP-EWLS学習アルゴリズムであり、(42b), (44b)式の初期値を  $P_0 = Q_0 = \alpha I$ 、 $\alpha = 10^9$  および  $\rho = 0.9$  としている。次のTable 1は中間層ユニット数  $m = 50$ ,  $T = 12$ ,  $n = 12$ ,  $\varepsilon = 0.001$  の場合の結果である。最後の列は学習回数を表している。

Table 1 Simulation Result 1 of Sin Learning  
 $m = 50, T = 12, n = 12, \varepsilon = 0.001$

K	E	$E_0$	$E_t$	$E_{t0}$	
0.05	0.0006	0.0023	0.3779	0.2544	21
0.1	0.0005	0.0042	1.1692	0.6729	33
0.2	0.0008	0.0223	4.3887	2.4097	21
0.4	0.0006	0.0788	16.437	8.5114	31

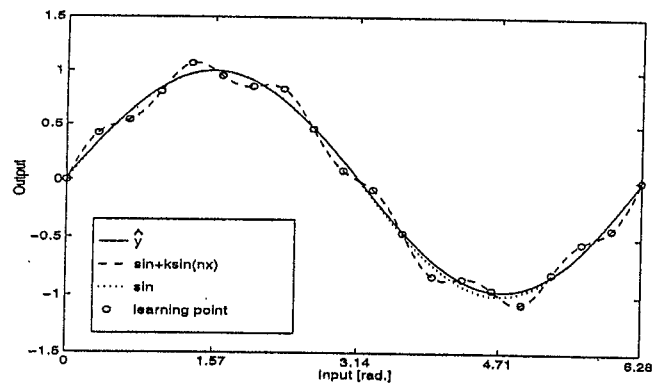


Fig.1 The Case :  $n = 6, k = 0.1, T = 20, m = 50$

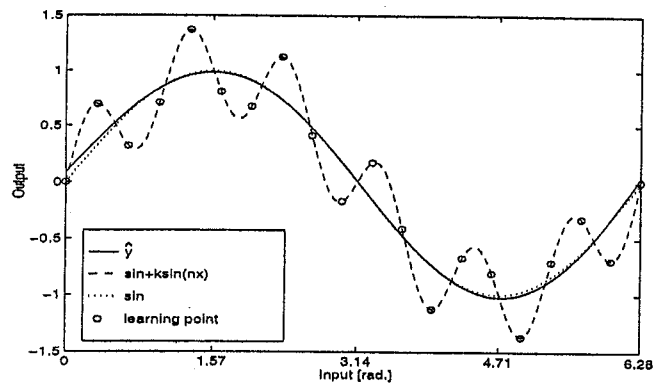


Fig.2 The Case :  $n = 6, k = 0.4, T = 20, m = 50$

この場合には学習目的である  $E$  を打ち切り誤差 0.001 以下として学習が終了しているが、関数全体を眺めることになる汎化能力に対しては、高調波を含まない  $E_{t_0}$  のほうがより小さくなっている。その他の場合には、学習は  $E_0$  の値により終了する場合も多々見られるが、汎化能力に関しては多くの場合  $E_{t_0}$  のほうが小さい場合が多い。Table 2 は中間層ユニット数を 25 とし、他は Table 1 と同じ条件である。この結果を Table 1 と比較すると、

Table 2 Simulation Result 2 of Sin Learning  
 $m=25, T=12, n=12, \epsilon=0.001$

K	E	$E_0$	$E_t$	$E_{t_0}$	
0.05	0.0005	0.0016	0.2939	0.1693	31
0.1	0.0005	0.0040	0.9293	0.4309	31
0.2	0.0003	0.0184	3.9710	1.9888	31
0.4	0.0004	0.0723	15.742	7.8050	31

すべての評価の値が小さくなっているが、学習回数は増加している。しかし、この傾向は他の例では必ずしも正しいわけではないので一般的傾向とは言えない。一方、 $\epsilon = 0.01$  とし、他は Table 1 と同じ条件の場合が Table 3

Table 3 Simulation Result 3 of Sin Learning  
 $m=50, T=12, n=12, \epsilon=0.01$

K	E	$E_0$	$E_t$	$E_{t_0}$	
0.05	0.0089	0.0139	1.6307	1.5078	19
0.1	0.0031	0.0108	1.5647	1.0396	20
0.2	0.0045	0.0327	5.5176	3.5402	20
0.4	0.0071	0.1043	19.199	11.288	20

である。このとき評価の値がすべて悪化しているのは基準値によるもので当然である。これら何れの Table においても  $K$  の増加に対して学習はうまくいくものの、汎化能力は大きく後退している。このとき汎化能力を良くするためには学習点数の増加が必要である。さらに種々の場合の学習結果については文献 [14] を参照されたい。なお、これらの数値結果の一部ををグラフ化したものが Fig.1 ~ Fig.5 である。図中丸印(○)が学習点を表している。これらのグラフから、高調波成分の学習点における値が、正弦波のまわりをばらついていれば、それらを内挿する形で学習結果は正弦波を構成している(cf. Fig.1, 2)。すなわち、学習点が偏りなくほぼ平均零で正弦波のまわりをばらついていれば、高調波を含む正弦波の学習は、高調波を含まない正弦波により近くなっていると言える。言い換えれば、NN

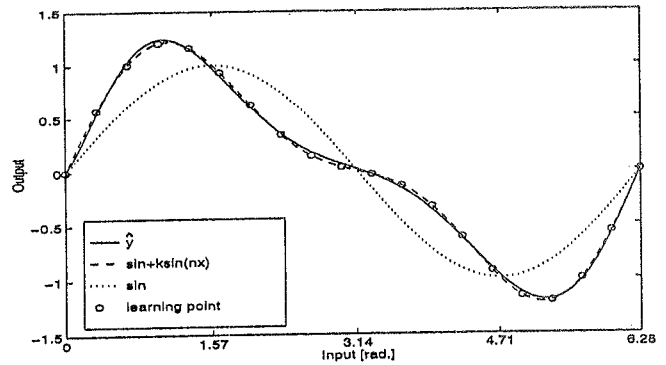


Fig.3 The Case :  $n = 2, k = 0.4, T = 20, m = 50$

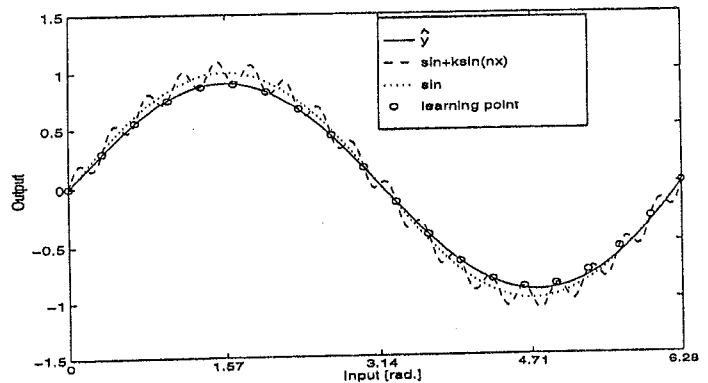


Fig.4 The Case :  $n = 18, k = 0.1, T = 20, m = 50$

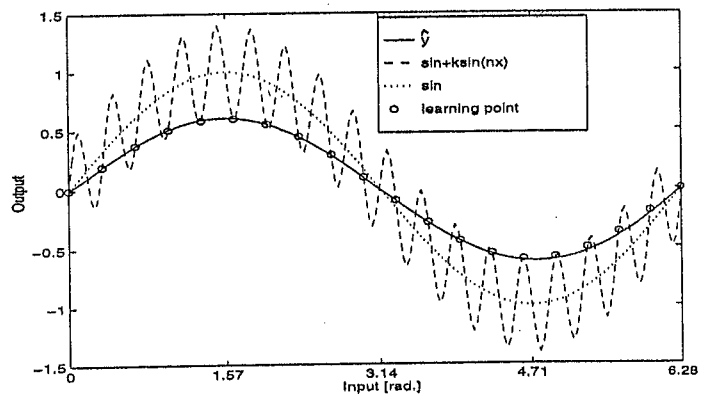


Fig.5 The Case :  $n = 18, k = 0.1, T = 20, m = 50$

による学習は、平均零の雑音を炉波する働きをしていることになる。ただし、このことは学習点数に大きく依存するものであり、あまり大きくし過ぎると、雑音成分の影響を受けることになる。あるいは、高調波成分を雑音としてではなく、意味のある確定信号と見なされる場合には、その高調波に応じた学習点数の選択が必要である。例えば、Fig.3 は  $n = 2$  の第2調波の場合であり、 $T = 20$  で精度良く学習できることを示している。一方、学習点に高調波の影響による偏りがある場合にもそれら学習点を内挿することになるが、学習結果は、学習点から近似される最も近い正弦波を構成することになる(cf. Fig.4, 5)。

5. 文字認識

本節では、提案するEBP-EWLS学習アルゴリズムを用いて、0~9のアラビア数字の学習を行った結果を述べる。学習する文字をTable 4の第1行に示す。それぞれの文字は5×5のビットマップで表される。第2行以降は汎化能力のためのテストデータであり、学習データに対して、1~5ビット反転させたものである。従ってNNの入力は25入力となり、黒のブロックを1、白のブロックを0として入力している。出力は0~9の識別のために4出力とし、それぞれの文字を

表す10個の25次元入力ベクトルに対して、10個の異なる4次元教師信号ベクトルを対応させる。この対応関係にも種々の方法が考えられ、ここではTable 5に示す3パターンを採用した。Pattern 1は2進法表示であり、Pattern 2は文献[15]で用いられたものである。Pattern 3が新たに筆者らが考えたものであり、0-1の反転度数を多くしたものである。以上の準備のもとに、 $\alpha = 10^6$ ,  $\varepsilon = 0.001$  および  $\rho = 0.3, 0.6, 0.9$  とし、重みの初期値を(0, 1)の乱数100通りに対して行った結果をTable 6に記す。中間層ユニット数は  $m = 3 \sim 10, 20$  に対して行った。

Table 5 Patterns of Teacher Signals

	Pattern 1	Pattern 2	Pattern 3
0	0 0 0 0	0 0 0 1	0 0 0 0
1	0 0 0 1	0 0 1 0	0 0 1 1
2	0 0 1 0	0 1 0 0	0 1 0 1
3	0 0 1 1	1 0 0 0	0 0 0 1
4	0 1 0 0	0 0 1 1	0 1 1 0
5	0 1 0 1	0 1 1 0	1 0 0 1
6	0 1 1 0	1 1 0 0	1 1 0 0
7	0 1 1 1	1 0 0 1	1 0 1 0
8	1 0 0 0	0 1 0 1	1 1 1 0
9	1 0 0 1	1 0 1 0	1 1 1 1

Table 4 Bitmaps of Arabic Numerals (5 × 5)

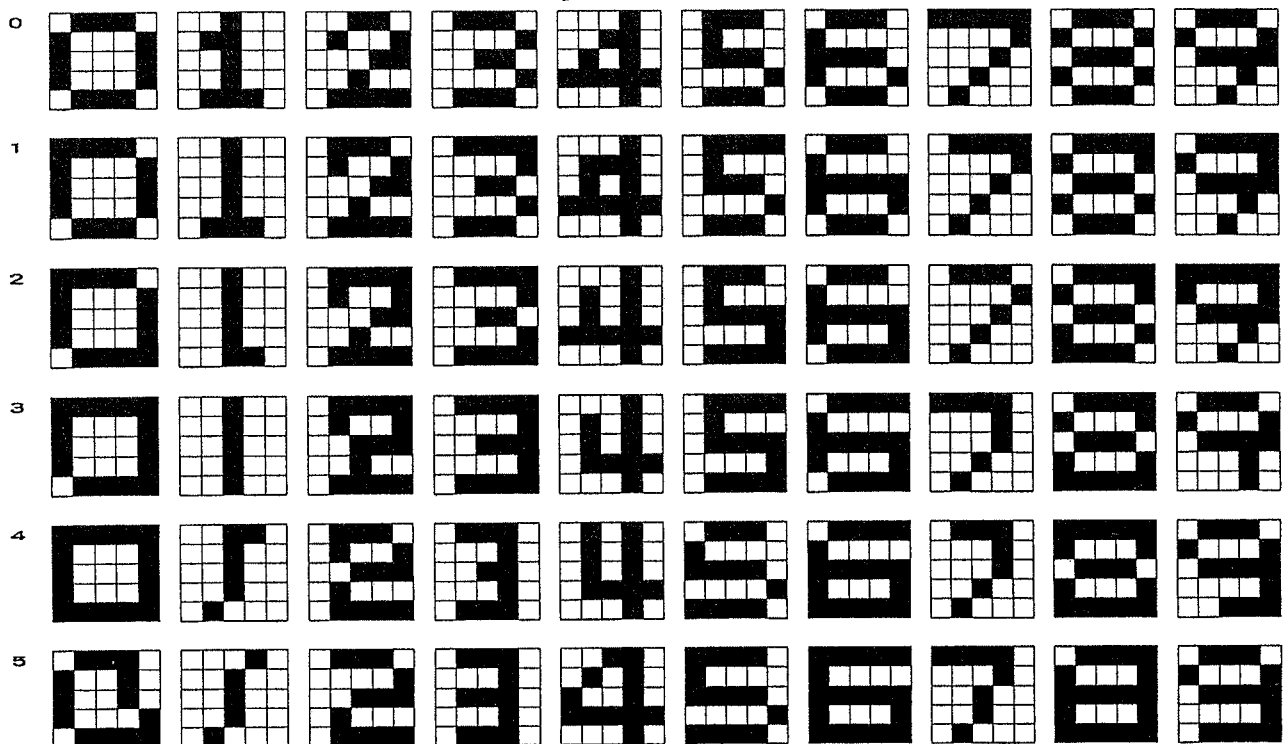


Table 6 Simulation Results of Arabic Numerals Learning  
(a) Pattern 1

m	$\rho = 0.3$				$\rho = 0.6$				$\rho = 0.9$			
	C	A	L	U	C	A	L	U	C	A	L	U
3	10	95.3	27	201	5	272.0	138	455	0	-----	-----	-----
4	34	53.6	13	230	41	110.3	10	589	34	249.7	10	955
5	74	56.1	10	259	67	147.8	10	679	39	303.7	10	997
6	92	37.8	11	105	96	78.8	13	678	79	186.1	10	978
7	99	19.3	10	80	100	40.9	10	354	95	111.0	10	910
8	100	13.7	10	11	100	15.9	10	69	100	30.8	10	474
9	100	10.0	10	10	100	10.0	10	12	100	10.0	10	10
10	100	10.0	10	10	100	10.0	10	10	100	10.0	10	10
20	100	10.0	10	10	100	10.0	10	10	100	10.0	10	10

(b) Pattern 2

m	$\rho = 0.3$				$\rho = 0.6$				$\rho = 0.9$			
	C	A	L	U	C	A	L	U	C	A	L	U
3	4	98.5	74	149	19	280.1	41	654	9	516.0	225	891
4	59	52.1	10	251	70	85.4	10	581	58	177.5	10	790
5	76	43.2	10	220	90	99.9	10	649	69	149.8	10	871
6	93	27.6	10	119	97	54.1	10	510	87	107.3	10	730
7	100	17.7	10	240	99	19.5	10	84	99	61.0	10	906
8	100	13.8	10	120	100	16.1	10	166	100	36.8	10	619
9	100	10.0	10	12	100	10.1	10	14	100	10.1	10	20
10	100	10.0	10	10	100	10.0	10	10	100	10.0	10	10
20	100	10.0	10	10	100	10.0	10	10	100	10.0	10	10

(c) Pattern 3

m	$\rho = 0.3$				$\rho = 0.6$				$\rho = 0.9$			
	C	A	L	U	C	A	L	U	C	A	L	U
3	0	-----	-----	-----	0	-----	-----	-----	0	-----	-----	-----
4	46	72.2	10	243	66	108.9	10	658	41	240.5	10	827
5	80	37.4	10	201	91	69.5	10	631	87	164.6	10	679
6	98	24.1	10	120	100	37.2	10	257	98	125.0	10	822
7	100	18.0	10	163	100	25.3	10	290	100	62.1	10	579
8	100	11.7	10	17	100	12.8	10	27	100	18.3	10	118
9	100	10.0	10	12	100	10.0	10	10	100	10.0	10	10
10	100	10.0	10	10	100	10.0	10	10	100	10.0	10	10
20	100	10.0	10	10	100	10.0	10	10	100	10.0	10	10

これらの結果から、学習回数は中間層ユニット数  $m$ 、指数重み  $\rho$  の値に大きく依存していることが分かる。指数重み  $\rho$  はこの問題に対しては  $\rho = 0.6$  あたりが良いようである。 $m$  の値は 10 くらいは必要である。これらの値を適切に選ぶと、任意の重みの初期値からすべて 10 回の学習で収束していることは大きな特徴である。BP 法は勿論のこと、EBP-OP アルゴリズムでもこのような結果は得られていない。教師信号に対するパター

ンによる違いは、あまり大きく影響していないが、すべての初期値に対して 10 回で収束する  $m$ 、あるいは収束率が 100% となる  $m$  を比較すると、Pattern 3 が若干良いようである。

次に、学習して得られた重みを用いて、Table 4 に示すテストデータに対する汎化能力を調べてみた。ただし、汎化能力の判定基準として、NN の出力  $c$  の値により、



(A)  $c \geq 0.7 \rightarrow 1, c \leq 0.3 \rightarrow 0$

(B)  $c \geq 0.5 \rightarrow 1, c < 0.5 \rightarrow 0$

の2種類で行った。(A)の方がより正しい(厳しい)判定であるが、0~9のどれでもないという判定不可能な場合を含んでいる。これを避け、とにかく何かの数字に判定しようとするのが(B)である。しかし何れの場合にも4次元出力ベクトルには、教師信号ベクトルに含まれないものがある。

判定結果を Table 7 に示す。表中の数字はNNの出力結果であり、誤り判定には薄い黒幕が掛けである。また、数字の左側の星印(\*)は(A)によっては判定されず、(B)による判定結果である。それでも判定不可能なものは、×が記してある。学習する文字により大きく異なるが、全体として、

Table 7 Generalization Using Test Numerals

	$\rho = 0.3$			$\rho = 0.6$			$\rho = 0.9$		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
0	P1	P2	P3	P1	P2	P3	P1	P2	P3
1	0	0	0	0	0	0	0	0	0
2	*2	0	*0	*2	0	*0	*2	0	*0
3	*2	*0	*×	*2	*0	*×	*2	*0	*×
4	2	*0	*×	2	*0	*×	2	*0	*×
5	*2	*3	×	*2	*3	×	*2	*3	×
1	P1	P2	P3	P1	P2	P3	P1	P2	P3
1	*3	*9	1	*3	*9	1	*3	*9	1
2	*1	*9	1	*1	*9	1	*1	*9	1
3	*3	*9	1	*3	*9	1	*3	*9	1
4	7	*9	*×	7	*9	*×	7	*9	*×
5	*7	9	*1	*7	9	*1	*7	9	*1
2	P1	P2	P3	P1	P2	P3	P1	P2	P3
1	*2	*×	*3	*2	*×	*3	*2	*×	*3
2	*3	*1	*2	*3	*1	*2	*3	*1	*2
3	3	*1	*2	3	*1	*2	3	*1	*2
4	3	*1	2	3	*1	2	3	*1	2
5	3	3	2	3	3	2	3	3	2
3	P1	P2	P3	P1	P2	P3	P1	P2	P3
1	3	3	3	3	3	3	3	3	3
2	3	3	2	3	3	2	3	3	2
3	3	*1	2	3	*1	2	3	*1	2
4	*7	*9	3	*7	*9	3	*7	*9	3
5	*7	*×	*×	*7	*×	*×	*7	*×	*×
4	P1	P2	P3	P1	P2	P3	P1	P2	P3
1	4	4	4	4	4	4	4	4	4
2	*4	4	4	*4	4	4	*4	4	4
3	*5	*4	×	*5	*4	×	*5	*4	×
4	5	4	*1	5	4	*1	5	4	*1
5	*3	*9	*×	*3	*9	*×	*3	*9	*×

5	P1	P2	P3	P1	P2	P3	P1	P2	P3
1	5	5	5	5	5	5	5	5	5
2	5	5	*×	5	5	*×	5	5	*×
3	5	5	×	5	5	×	5	5	×
4	*6	6	*×	*6	6	*×	*6	6	*×
5	*7	6	*6	*7	6	*6	*7	6	*6
6	P1	P2	P3	P1	P2	P3	P1	P2	P3
1	6	*6	6	6	*6	6	6	*6	6
2	6	6	*6	6	6	*6	6	6	*6
3	*6	*×	6	*6	*×	6	*6	*×	6
4	2	*9	*6	2	*9	*6	2	*9	*6
5	*2	*1	*6	*2	*1	*6	*2	*1	*6
7	P1	P2	P3	P1	P2	P3	P1	P2	P3
1	7	*7	7	7	*7	7	7	*7	7
2	*7	*3	7	*7	*3	7	*7	*3	7
3	*3	*3	*7	*3	*3	*7	*3	*3	*7
4	3	3	*×	3	3	*×	3	3	*×
5	3	*9	*×	3	*9	*×	3	*9	*×
8	P1	P2	P3	P1	P2	P3	P1	P2	P3
1	8	8	8	8	8	8	8	8	8
2	8	*8	8	8	*8	8	8	*8	8
3	8	*8	*8	8	*8	*8	8	*8	*8
4	*8	*×	8	*8	*×	8	*8	*×	8
5	*×	*×	*×	*×	*×	*×	*×	*×	*×
9	P1	P2	P3	P1	P2	P3	P1	P2	P3
1	9	*9	9	9	*9	9	9	*9	9
2	9	*4	9	9	*4	9	9	*4	9
3	9	9	9	9	9	9	9	9	9
4	*8	*2	*9	*8	*2	*9	*8	*2	*9
5	*8	2	*×	*8	2	*×	*8	2	*×

ある程度の汎化能力が認められる。テストデータにおいて同じ1ビットの違いでも、どの1ビットを反転させるかで結果が異なることは当然である。特に数字1, 2のように1ビットの違いに対して正しい判定ができない理由の一つとして、この文字認識が5×5のビットマップに対して行われていることが考えられる。0~9に対して正しく判定した個数を Table 8 に示す。\*印のついた

Table 8 A List of Corrections

	$\rho = 0.3$			$\rho = 0.6$			$\rho = 0.9$		
	P1	P2	P3	P1	P2	P3	P1	P2	P3
正	17	13	21	17	13	21	17	13	21
正*	23	21	31	23	21	31	23	21	31
誤	9	6	5	9	6	5	9	6	5
誤*	27	29	19	27	29	19	27	29	19

のは(B)による判定であり、正解も増えるが、それ以上に誤解答が増えている。これらの結果からは Pattern 3 が一番良い教師信号であると言える。一方、指数重み  $\rho$  による違いは見られない。このことは、学習回数の早さと、学習結果による汎化能力とは関係がないことを示している。さらにこのことは、どの学習アルゴリズムを用いるかに対しても言えるのではないと思われる。また、汎化能力に対する結果は、学習で用いた 100 通りの重みの初期値に対する結果の一つの例であり、従って、絶対的な強い主張は言えない。

## 6 おわりに

本論文では、階層型ニューラルネットワークの学習アルゴリズムとして新しく開発している EBP 学習アルゴリズムの一つとして EBP-EWLS 学習アルゴリズムを提案し、その有効性を検討するために正弦関数と 0～9 の文字認識を行った。EBP-EWLS 学習アルゴリズムは、従来の逐次処理法および一括処理法とは異なり、学習の評価に関しては一括的であり、一方、重みの更新は逐次的である。従って、実用的に使い易く、特に、オンライン学習に向いている。

例題として用いた正弦関数の学習に関しては、非常に高速な学習が実行されている。特にここでは、高調波成分が加わった場合の学習について検討した。一方、文字認識に対しても学習が良好に行われることが分かった。さらにその汎化能力についても検討した。ここで用いた学習データは  $5 \times 5$  のビットマップで表される場合であり、まだ基礎的な段階である。より実用化を目指して  $10 \times 10$  のビットマップに拡張し、さらに種々の検討を現在行っている。

ここに示した二つの例題からも分かるように、EBP-EWLS 学習アルゴリズムは指数重み最小 2 乗法を用いているので、指数重み  $\rho$  の決定が重要である。線形システムに対するパラメータ推定では  $\rho$  の値を 0.9～0.95 あたりが良いとされているが、学習問題の多くは非線形であり、適切な  $\rho$  の値の選択は、さらに多くの例題から決定されなければならない。現時点では、 $\rho$  の値は学習データの長さに比例することは確かなようである。

## 参考文献

1) D.E.Rumelhart and J.L.McClelland Eds. : Parallel

- Distributed Processing, Vol.1, MIT Press, 1986.
- 2) R.A.Jacobs : Increased Rates of Convergence through Learning Rate Adaptation, Neural Networks, 1, pp295-307, 1988
  - 3) 石川：忘却を用いたコネクショニストモデルの構造学習アルゴリズム、人工知能学会誌、5 巻, 5 号, pp595-603, 1991.
  - 4) 渡辺 他：学習係数の適応調整によるニューラルネットワーク学習の高速化、計測自動制御学会論文集、30 巻, 9 号, pp1093-1099, 1994.
  - 5) A.J.Shepherd : Second-Order Methods for Neural Networks, Springer, 1996.
  - 6) 山本：ニューロ回路の学習規則と適応アルゴリズム、システム制御情報学会論文誌、7 巻, 12 号, pp47-49, 1994.
  - 7) 山本：多層ニューロ回路の新しい誤差逆伝搬法とその代数的性質、システム制御情報学会論文誌、9 巻, 5 号, pp1-9, 1996.
  - 8) 山本：正規直交化法による多層ニューラルネットワークの学習アルゴリズム、システム制御情報学会論文誌、9 巻, 10 号, pp33-39, 1996.
  - 9) 山本、坂本：指数重み付き最小 2 乗法による EBP 学習アルゴリズム、第 8 回計測自動制御学会中国支部学術講演会論文集、pp26-27, 1997.
  - 10) 山本、松田：EBP-EWLS 学習アルゴリズムによる論理関数の学習、第 8 回計測自動制御学会中国支部学術講演会論文集、pp28-29, 1997.
  - 11) K.J.Astrom and B.Wittenmark : Adaptive Control, Addison-Wesley, 1989.
  - 12) 山本、坂本：指数重み付き最小 2 乗法による EBP 学習アルゴリズム、計測自動制御学会論文集、34 巻, 8 号, 1998.
  - 13) 山本、佐藤：EBP 学習アルゴリズムの逐次修正法と一括処理法、第 17 回適応制御シンポジウム資料、pp37-40, 1997.
  - 14) 奥谷：EBP-EWLS 学習アルゴリズムによる非線形関数の学習、鳥取大学工学部平成 9 年度卒業論文、1998.
  - 15) 巽、福島：逐次射影法としての誤差逆伝搬法、システム制御情報学会論文誌、8 巻, 5 号, pp204-211, 1996.
  - 16) 福本：EBP 学習アルゴリズムによる文字認識の基礎研究、鳥取大学工学部平成 9 年度卒業論文、1998.

(受理 平成 10 年 8 月 25 日)