

# 多段階学習と誤差に応じた 学習係数に基づく振動効果を利用した ニューラルネットワークの一学習法

田口 功

## A Learning Method for the Layered Neural Networks Based on the Multi-stage Learning and a Dynamic Adjustment of Learning Coefficient by Errors for Using Oscillatory Effect

Isao TAGUCHI

This paper proposes an efficient learning method for the layered neural networks based on the selection of training data and input characteristics of an output layer unit. The multilayer neural network is widely used due to its simple structure. When learning objects are complicated, the problems, such as unsuccessful learning or a significant time required in learning, remain unsolved. The aims of this paper are to suggest solutions of these problems and to reduce the total learning time.

Focusing on the input data during the learning stage, we under-took an experiment to identify the data that makes large errors and interferes with the learning process. Our method divides the learning process into several stages.

In general, input characteristics to an output layer unit show oscillation during learning process for complicated problems. Computational experiments suggest that the proposed method has the capability of higher learning performance and needs less learning time compared with the

conventional method.

## 1. はじめに

ニューラルネットワークの研究の多くは、パルスニューラルネットワーク<sup>(1)</sup>など、複雑なネットワークの研究が盛んに行われている。一方、バックプロパゲーション法（以下：BP法）を用いたシグモイド素子から構成される階層型ニューラルネットワーク（以下：NN）は、構成が容易であるために、現実的には広範囲に利用されている。

学習対象が複雑で、多量の学習データを用いた学習が必要な場合において、論文<sup>(2)</sup>では、学習する前に教師データを学習のしやすさに着目した分類、その分類に基づく多段階学習<sup>(3)</sup>、誤差の大きさに応じた学習係数の動的調整、出力層素子への入力特性（振幅減少条件や目標値捕捉条件の有効利用）を特徴とした総合的な学習方法を提案し、その有効性を示した。

提案手法による積極的な振動特性の利用は、BP法の拡張であるQuickProp法<sup>(4)</sup>（以下：QPROP法）や、弾力性BP法<sup>(5)</sup>（以下：RPROP法）に対し、組み込み効果が期待できる。QPROP法は、学習係数をできるだけ大きく調整しながら、かつ、重み係数の変化量に慣性項を加え振動を抑える方法である。また、重み係数の急激な変更による振動を避けるために最大変化量も導入されている。基本的には、学習係数、最大変化量、重み変更抑制係数の3係数を必要とする。また、弾力性BP法<sup>(5)</sup>は、重み更新量の符号を確認しながら振動を抑え、加速を行う。本来使用されている学習係数などの5個のパラメーターを条件に応じて調整することにより、学習を進行する方法である。両方法ともに振動を抑えながら、しかも、加速するためのパラメーターを持って学習が行われる。

本論文では、特に、組み込みによる誤差、学習時間に対する効果、および、出力層素子への入力特性にどのような変化が起きるのかについても報告する。

振動現象については、論文<sup>(6)</sup>や参考文献<sup>(7)(8)</sup>にも述べられている。論文<sup>(6)</sup>では、出力層に、強制的にしきい値揺らぎを加えることによる効果が報告されており、参考文献<sup>(7)(8)</sup>では、オーバーシュートする場合がありますが述べられている。本論文における振動は、強制的にしきい値揺らぎを加えることはしない。論文<sup>(2)</sup>で述べたように、種々の工夫と学習データ自身の複雑さから生じたものと考えられる。

また、振動現象を解析するために、重み更新量（中間層と出力層間）をBP法の基本となる各パターンに対する誤差の絶対値の大きさの増加、減少を基本とし、増加に対する重み更新量と減少に対する重み更新量をそれぞれ累計加算し、重み更新量を2個に分割した。その、誤差の絶対値の増加、減少に対応する重み更新量の要素から2個のベクトルが構成できる。それらのベクトルの大きさ、角度の動きを調べることにより、学習が良好になる場合と学習が進行しない場合の検討ができる。

学習中、振動特性を持つ学習と振動していない学習に対する平均2乗誤差（以下：*RMSE* [Root Mean Square Error]）の関係、すなわち、どちらが*RMSE*値が少なくなるかは、興味深い問題である。学習係数を小さくすれば、学習の進行が遅くなり、振動が抑えられることは、経験的な数々の実験から期待できる。すると、従来法で学習を行っている場合に、振動が起きているとき、途中で学習係数を小さくすることにより、振動の無い（単調減少特性）学習は、実現できる。したがって、振動を残しながら学習を行った場合と、単調減少特性の場合の誤差に対する比較実験は、可能となる。すなわち、従来から、オーバーシュートが起こることは、学習時間を増加させることで、学習を悪化させるために悪いものと思われるが、振動が生じていたほうが学習が改善されるのか、または、振動がないほうが良いのかは、*RMSE*値をもとに確認できる。本論文では、関数近似問題を対象とした計算機実験により、*RMSE*値と出力層素子への入力特性との関係として報告する。

また、重み係数自身の振動を利用した学習は、あまり知られていない。最終段階の学習の終了直前（終了9回前から学習終了時まで）のすべての重

み係数（入力層と中間層間の重み係数および中間層と出力層間の重み係数）を記憶しておき、すべての重み係数の最大値と最小値の平均値（Mean Weight）を求め、その重み係数を用いることで、*RMSE* 値を減少させることができるかどうかの確認をする。

本論文の構成は、以下のとおりである。第2章では、重み更新量を基にしたベクトルについて述べる。第3章では、関数近似問題を例として、計算機実験により、*QPROP* 法や *RPROP* 法との比較を行い、提案手法の組み込みに対する有効性を示す。また、提案手法で発生する振動は、学習に有効で、学習誤差を減少させる働きがあることを、出力層素子への入力特性の変化、重み係数の更新量ベクトルの動きをもとに実験的に確認する。第4章はまとめである。

## 2. 効率的学習法

提案手法の誤差に応じた学習係数の値の動的調整、出力層素子への入力特性、教師データの選択、それらに基づく効率的学習法については、論文<sup>(2)</sup>で詳しく述べた。*QPROP* 法と *RPROP* 法の詳細についても、それぞれ、論文<sup>(4)</sup> および、論文<sup>(5)</sup> で述べられている。ここでは、絶対誤差と重み更新量を基にしたベクトルの構成法について述べる。

### 2.1 絶対誤差と重み更新量を基にしたベクトルについて

ここでは、学習中、すべてのパターンに対する中間層と出力層間の重みに対し、1エポックごとの重み更新量をベクトルの要素とする。1エポックの学習で、すべてのパターンに対し、重み更新量を前回の学習に対する絶対誤差と現在の学習に対する絶対誤差が増加する場合と減少する場合で2分割し、それぞれ、加算した。それぞれのベクトルの要素は、中間層と出力層間の重みの個数に等しい。その要素をもとに、増加（絶対誤差増加）ベクトル、減少（絶対誤差減少）ベクトル、さらに、合計ベクトルを構成できる。また、特定の誤差が最後まで残るベクトル要素に対す

る重み更新量も選択でき、表示することもできる。本論文では、振動現象が止まり、学習が進行しない場合のベクトルの動き（単調減少）や振動が継続していく場合の基本的なベクトルの動きを調べる。

ここでは、学習中、絶対誤差が増加、減少する重み更新量のベクトルの要素数を  $p$  とし、 $i = 1$  は、入力層および中間層間の重み係数とし、 $i = 2$  は、中間層と出力層間の重み係数とする。また、 $t$  は、学習回数とし、 $inc$  は絶対誤差の増加、 $dec$  は絶対誤差の減少を意味する。絶対誤差の増加に対する重み更新量の要素ごとの合計を、

$$\frac{\partial E^{inc}(t)}{\partial W_{j2}} = \{ \Delta W_{12}^{inc}(t), \Delta W_{22}^{inc}(t), \dots, \Delta W_{p2}^{inc}(t) \} \quad (1)$$

とする。また、同様に、絶対誤差の減少に対する重み係数更新分のベクトル要素は、

$$\frac{\partial E^{dec}(t)}{\partial W_{j2}} = \{ \Delta W_{12}^{dec}(t), \Delta W_{22}^{dec}(t), \dots, \Delta W_{p2}^{dec}(t) \} \quad (2)$$

となる。このとき、2個のベクトルの内積は、

$$\begin{aligned} \frac{\partial E^{inc}(t)}{\partial W_{j2}} \cdot \frac{\partial E^{dec}(t)}{\partial W_{j2}} &= \Delta W_{12}^{inc}(t) \Delta W_{12}^{dec}(t) \\ &+ \Delta W_{22}^{inc}(t) \Delta W_{22}^{dec}(t) + \dots + \Delta W_{p2}^{inc}(t) \Delta W_{p2}^{dec}(t) \end{aligned} \quad (3)$$

となり、2個のベクトルの角度を  $\theta$  とすると、

$$\cos \theta = \frac{\frac{\partial E^{inc}(t)}{\partial W_{j2}} \cdot \frac{\partial E^{dec}(t)}{\partial W_{j2}}}{\left| \frac{\partial E^{inc}(t)}{\partial W_{j2}} \right| \left| \frac{\partial E^{dec}(t)}{\partial W_{j2}} \right|} \quad (4)$$

となる。

### 3. 関数近似問題を例とした計算機実験

本論文で提案している多段階学習法の有効性を示すために、ここでは関数近似問題を例として計算機実験を行う。関数近似問題を選択した理由は、学習する際の難易度を関数の選択によって設定でき、未学習デー

表 1 Learning Functions

(a) Schwefels Function $f(x, y) = -x \sin \sqrt{ x } - y \sin \sqrt{ y }$ $(-15.0 \leq x \leq 15.0, -15.0 \leq y \leq 15.0)$ $(-20.04 \leq f(x, y) \leq 20.04)$
(b) Rastrigin Function $f(x, y) = x^2 - 10 \cos(2\pi x) - y^2 - 10 \cos(2\pi y)$ $(-0.8 \leq x \leq 0.8, -0.8 \leq y \leq 0.8)$ $(-20.0 \leq f(x, y) \leq 20.5)$
(c) Ridge Function $f(x, y) = 2x^2 + 2xy^2 + y^2$ $(-6.0 \leq x \leq 6.0, -6.0 \leq y \leq 6.0)$ $(0.0 \leq f(x, y) \leq 180.0)$

タに関して実験の検証がしやすい。また、いくつかの関数は、最適化問題において、よく知られている Schwefels 関数、Rastrigin 関数および Ridge 関数を使用し、それらをまとめて表 1 に示す。

### 3.1 教師データ

表 1(a)では、学習すべき領域は  $-15.0 \leq x, y \leq 15.0$  として示されている。さらに、教師データは、 $x, y$  方向の刻み幅を 3.0 として領域を  $11 \times 11$  の格子に分割し、それらの格子点 121 点に対する値を教師データの集合  $D$  とする。

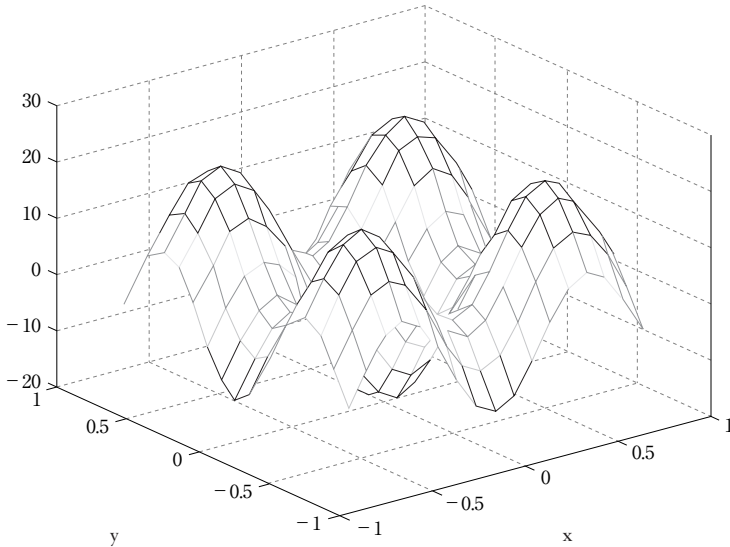
表 1(b)を図 1 に示す。図 1 において、隣り合う教師データは、実線で結んで表示している。表 1(b)では、学習すべき領域は  $-0.8 \leq x, y \leq 0.8$  である。教師データは、 $x, y$  方向の刻み幅を 0.1 として領域を  $17 \times 17$  の格子 (289 格子点) に分割し、教師データの集合  $D$  とした。

同様に、表 1(c)では、学習すべき領域は  $-6.0 \leq x, y \leq 6.0$  である。教師データは、 $x, y$  方向の刻み幅を 1.2 として領域を  $11 \times 11$  の格子 (121 格子点) に分割し、教師データの集合  $D$  とした。

### 3.2 教師データの選択

ここでは、本論文で、学習困難なために、多くの実験を行った表 1(b)

図 1 An example of characteristics of the target function



についての教師データの選択方法について述べる。表 1(b) による教師データの選択は、 $x, y$  両方向の偏微分値  $f_x(x, y)$  または  $f_y(x, y)$  で代用できる。

次に、教師データを 3 段階に分け ( $s = 3$ )、多段階学習を行う。 $D_i \cup D'_i$  は、全体の約 30% の教師データを用いる。 $D_i$  および  $D'_i$  [ $f_x(x, y)$  または  $f_y(x, y)$  の値を利用して選択] は、関数の概観を把握するという意味で重要である。 $D_i \cup D'_i$  に含まれるデータ数は、最終的に 31.1% (90 個) になった。これらのデータには、 $|f(x, y)| \geq 0.90$  を満たす教師データ 24 個、 $|f_x(x, y)| \geq 60.3$  を満たす教師データ 34 個、 $|f_y(x, y)| \geq 60.3$  を満たす教師データ 34 個が含まれ、そのうち 2 個が重複しているため、全体で 90 個となる。

第 2 段階では、 $D_i^2 \cup D_i'^2$  に含まれるデータ数は、最終的に 58.1% (168 個) になった。これらのデータには、 $|f(x, y)| \geq 0.80$  を満たす教師データ 52 個、 $|f_x(x, y)| \geq 60.1$  を満たす教師データ 68 個、 $|f_y(x, y)| \geq 60.1$  を満たす教師データ 68 個が含まれ、そのうち 20 個が重複しているため、全体で 168 個となる。

表 2 The Method of Learning (Schwefles Function)

Structure Learning data	Initial weight vector is modified 5 times. 2 input layer units. 9 middle layer units. 1 output layer unit. Updating a weight vector should be done every 1 epoch. Learning domains are $-15.0 \leq x \leq 15.0, -15.0 \leq y \leq 15.0$ . Learning data is conditioned equally as the reference table1 (a). All the domain is divided into $11 \times 11$ grid. All the learning data ( $\mathbf{D}$ ) is divided into 3 parts before learning. $0 < D^i < 1, D^i \in \mathbf{D} (D^i = f(x_i, y_i); i = 1, \dots, 121)$
1st step	$D_1^1 = \{f(x_i, y_i) \mid  f_x(x_i, y_i)  \geq 1.67 \text{ or }  f_y(x_i, y_i)  \geq 1.67\}$ (44 learning data) $ \mathbf{D}_1^1 \cup \mathbf{D}_2^1  = 2$ (2 overlapped) 42 learning data are selected. (34.7%)
2nd step	$D_2^2 = \{f(x_i, y_i) \mid  f_x(x_i, y_i)  \geq 1.22 \text{ or }  f_y(x_i, y_i)  \geq 1.22\}$ (21 learning data) $ \mathbf{D}_2^2 \cup \mathbf{D}_1^2  = 8$ (8 overlapped) 80 learning data are selected. (about 66%)
3rd step	121 learning data are selected. (100%)
Each step	Every step are learned for 2,334 times.
Conventional method	Learn 7,000 times with all the learning data and learning coefficient 0.8. Initial weight vector is modified 5 times.

第3段階における  $D_3^3$  は、全教師データを使用する。ここで、第1段階では、全データの31.1%、第2段階では、全データの58.1%、最後の第3段階で100%の教師データが用いられることになる。

表1(a)を学習させるための教師データの選択方法については、表2に、また、表1(c)を学習させるための教師データの選択方法については、表3にまとめる。表1(c)に示す Ridge 関数の特徴としては、 $|f_x(x_i, y_i)|$  または  $|f_y(x_i, y_i)|$  の値が大きい場合は、 $|f(x_i, y_i)|$  の値が大きくなるために、 $|f_x(x_i, y_i)|$  または  $|f_y(x_i, y_i)|$  の値をもとに教師データを選択した。

### 3.3 比較対象手法と実験の諸設定

計算機実験では提案手法の有効性を検証するために、従来手法との比較を行う。提案手法、従来法、QPROP法、RPROP法および提案手法を



表 3 The Method of Learning (Ridge Function)

Structure Learning data	<p>Initial weight vector is modified 5 times.                  2 input layer units. 9 middle layer units.                  1 output layer unit.                  Updating a weight vector should be done every 1 epoch.                  Learning domains are <math>-6.0 \leq x \leq 6.0, -6.0 \leq y \leq 6.0</math>.                  Learning data is conditioned equally as the reference table1 (c).                  All the domain is divided into <math>11 \times 11</math> grid.                  All the learning data (<math>D</math>) is divided into 3 parts before learning.  <math>0 &lt; D^i &lt; 1, D^i \in D (D^i = f(x_i, y_i); i = 1, \dots, 121)</math></p>
1st step	<p><math>D_1^1 = \{f(x, y) \mid  f_c(x, y)  \geq 1.67 \text{ or }  f_s(x, y)  \geq 20.0\}</math> (38 learning data)  <math> D_1^1 \cup D_1^2  = 6</math> (6 overlapped)                  32 learning data are selected. (26.4%)</p>
2nd step	<p><math>D_2^1 = \{f(x, y) \mid  f_c(x, y)  \geq 16.0 \text{ or }  f_s(x, y)  \geq 16.0\}</math> (70 learning data)  <math> D_2^1 \cup D_2^2  = 20</math> (20 overlapped)                  50 learning data are selected. (about 41.3%)</p>
3rd step	<p>121 learning data are selected. (100%)</p>
Each step	<p>Every step are learned for 2,334 times.</p>
Conventional method	<p>Learn 7,000 times with all the learning data and learning coefficient 0.8. Initial weight vector is modified 5 times.</p>

組み込んだ学習法における NN の構成は、シグモイド素子を用い、入力層素子数 2 個、中間層素子数 9 個、出力層素子数 1 個からなるフィードフォワード型の 3 層構造のネットワークとする。中間層素子数は予備実験より決定した。さらに、重み係数の更新は、両手法とも 1 エポック毎の一括更新方式を用いた。提案手法の学習回数は、1 段階目 2,333 エポック、2 段階目 2,333 エポック、3 段階目 2,334 エポックの合計 7,000 エポックとし、従来法は 7,000 エポックの学習で全教師データを常に用いる。実験に用いる計算機は、OS: Windows XP, CPU: Pentium 4, 3.0GHz, RAM: 2GB である。また、学習係数は、提案手法における基準の学習係数を  $\eta = 0.8$  (基準) とし、従来法は学習係数 0.8 を基準とした。学習結果を評価する RMSE に関しては、教師データに対する値を用い、重み係数の初期値を 5 通り設定した場合の平均値を用いる。重み係数の初期値は  $[-0.01, 0.01]$

の範囲でランダムに設定する。

また、数々の予備実験から、QPROP法に対する重み変更抑制係数 $\lambda = 0.005$ 、最大変化量 $= 0.95$ 、学習係数 $1\eta = 0.80$ 、重みの修正方向によって、学習係数 $2\eta = 0.12$ （文献では、 $0.0$ としてある）を決定し、RPROP法に対しても、 $\Delta_{max} = 5.0$ 、 $\Delta_{min} = 0.0025$ 、 $\eta^+ = 0.97$ （文献では $1$ 以上）、 $\eta^- = 0.89$ 、 $\eta^0 = 0.61$ とし実験を行った。

### 3.4 計算機実験結果

表1(c)に対して、提案手法および単独でQPROP法やRPROP法を用いた場合、さらに、QPROP法やRPROP法に提案手法を組み込んだ場合の

表4 RMSE for Proposed Methods and Traditional Methods in Learning of the Ridge Function

Proposed method	RMSE			
	Mean	Maximum value	Minimum value	Mean Weight value
	0.107	0.235	0.038	0.117
Learning time	Mean 6 minutes 59 seconds			
Proposed method + QPROP method	RMSE			
	Mean	Maximum value	Minimum value	Mean Weight value
	0.058	0.113	0.024	0.038
Learning time	Mean 7 minutes 10 seconds			
QPROP method	RMSE			
	Mean	Maximum value	Minimum value	Mean Weight value
	0.172	0.241	0.108	0.129
Learning time	Mean 10 minutes 24 seconds			
RPROP method	RMSE			
	Mean	Maximum value	Minimum value	Mean Weight value
	0.050	0.080	0.023	0.050
Learning time	Mean 13 minutes 52 seconds			
Proposed method + RPROP method	RMSE			
	Mean	Maximum value	Minimum value	Mean Weight value
	0.038	0.051	0.030	0.040
Learning time	Mean 7 minutes 38 seconds			

学習結果を表4および表5に示す。表4および表5に示す表中の数値は重みの初期値をランダムに設定した5回分の実験結果の平均の値である。

表4に示すように、Ridge関数の学習では、RPROP法およびQPROP法ともに提案手法を組み込んだ場合、RMSE (Mean) 値が改良され、学習時間も短縮された。

さらに、表5は、Schwefels関数を用い、提案手法および提案手法にQPROP法を組み込んだ場合の結果を示す。この場合も表4同様に、RMSE (Mean) 値が単独と比較して改良され、学習時間も約4分程度減少した。

また、表1(b)のRastrigin関数に対して、提案手法と従来法を用い、学習係数変化に対する学習誤差をRMSE、未学習データに対するRMSE (Non-Learning RMSE)、および、重み係数の振動を考慮した重み係数の使用によるRMSE (Mean Weight RMSE value) の結果を、それぞれ、表6および表7に表示した。表6および表7において、最終段階で振動が生じていた場合には、Non-Learning RMSEの右に○印を描き、振動がなく単調特性の場合には、空白とした。従来法においても、7,000回の学習に対して、最後の2,334回の学習中に振動が発生している場合には、同様に、○印を描き、振動がなく単調特性の場合には、空白とした。

表5 RMSE for Proposed Methods and Traditional Methods in Learning of the Schwefels Function

Proposed method	RMSE			
	Mean	Maximum value	Minimum value	Mean Weight value
	0.093	0.113	0.073	0.071
Learning time	Mean 6 minutes 28 seconds			
Proposed method + QPROP method	RMSE			
	Mean	Maximum value	Minimum value	Mean Weight value
	0.088	0.092	0.081	0.058
Learning time	Mean 6 minutes 35 seconds			
RPROP method	RMSE			
	Mean	Maximum value	Minimum value	Mean Weight value
	0.090	0.192	0.041	0.058
Learning time	Mean 10 minutes 29 seconds			

表 6 RMSE for Proposed Methods in Learning of the Rastrign Function

Proposed method	RMSE			
Learning time	Mean 15 minutes 56 seconds			
Learning coefficient	Mean Weight RMSE value	RMSE value	Non-Learning value	
0.9	0.046	0.050	0.049	○
0.8	0.044	0.126	0.069	○
0.7	0.033	0.037	0.030	○
0.6	0.027	0.027	0.028	○
0.5	0.028	0.026	0.030	○
0.4	0.034	0.039	0.047	○
0.3	0.066	0.081	0.066	○
0.2	0.097	0.096	0.092	○
0.1	0.115	0.138	0.105	○
0.05	0.135	0.135	0.125	

表 7 RMSE for Traditional Methods in Learning of the Rastrign Function

Traditional method	RMSE			
Learning time	Mean 28 minutes 17 seconds			
Learning coefficient	Mean Weight RMSE value	RMSE value	Non-Learning value	
0.9	0.618	0.618	0.627	
0.8	0.202	0.233	0.278	○
0.6	0.228	0.185	0.180	○
0.5	0.179	0.199	0.243	○
0.4	0.161	0.183	0.246	○
0.3	0.162	0.162	0.163	○
0.2	0.045	0.045	0.047	○
0.1	0.025	0.025	0.025	○
0.05	0.035	0.035	0.034	○
0.04	0.053	0.053	0.053	○
0.02	0.112	0.112	0.099	

表6のように、提案手法を表1(b)のRastrign関数に適用すると、学習係数(提案手法の時は基準)が、0.9から0.2までは、すべてのRMSE値は、0.1以下となった。学習係数が、0.1、0.05の時、RMSE値は、0.1以上となるが、学習係数値の使用範囲は、非常に広がっている。

表 8 RMSE for Traditional Methods in Learning of the Rastrign Function

Traditional method	RMSE				
	Learning coefficient in the final stage	Mean Weight RMSE value	RMSE value		Non-Learning value
	0.05	0.050	0.050	0.051	○
	0.04	0.053	0.053	0.053	○
	0.035	0.056	0.056	0.056	○
	0.032	0.057	0.057	0.058	
	0.030	0.059	0.059	0.059	

一方、表7に示すように、従来法に対しても、学習係数を0.9から、0.02まで変化させて学習を行った。0.9から、0.3での範囲では、*RMSE*値が0.1以上となり学習が行われていない。0.2から0.04の学習率を用いた場合には、*RMSE*値は、0.1以下となり、さらに、0.02まで減少させた場合には、逆に、*RMSE*値が増加した。

学習係数が0.04で一定の時、従来法を用いて学習を行った場合の学習結果(表7)は、7,000回の学習に対して、最後の2,334回の学習中に振動が発生し、*RMSE*値は、0.053である。この時の出力層素子への入力特性を図2に示す。ここで、最後の2,334回の始めの40回から学習学習係数を0.035に下げた時の出力層素子への入力特性を図3に示す。およそ、1,400回から振動が発生していることがわかる。さらに、40回から学習係数を0.050に上げた時の出力層素子への入力特性を図4に示し、学習学習係数を0.032、0.030に下げた時の出力層素子への入力特性を、それぞれ、図5、図6に表示する。40回から学習学習係数を0.050に上げた時の*RMSE*値は、0.051となり、学習係数が0.04で一定の時と比較して減少した。また、0.032、0.030に下げた場合には、0.057、0.059となり、振動はなくなる特性となったが、逆に、*RMSE*値は、増加した。これらのまとめた結果を表8に表示する。

### 3.5 実験結果の考察

計算機実験に基づき、従来からNN学習において使われているQPROP

图 2 Input characteristics for the traditional methods in the output layer

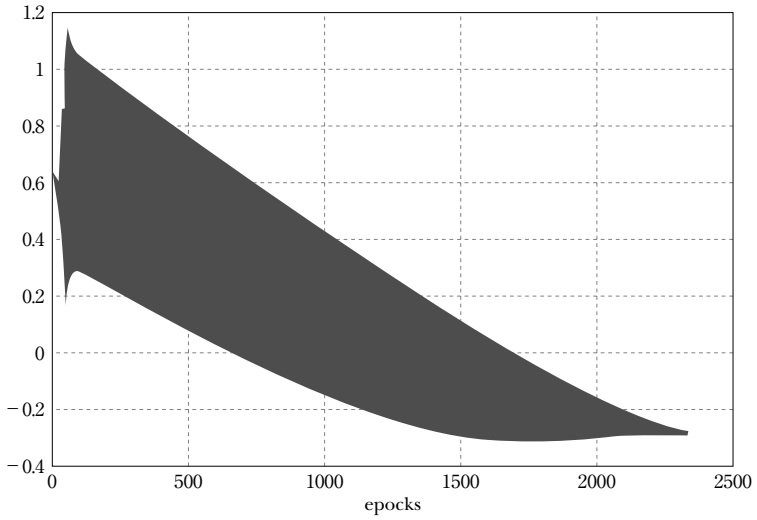


图 3 Input characteristics for the traditional methods in the output layer (0.035)

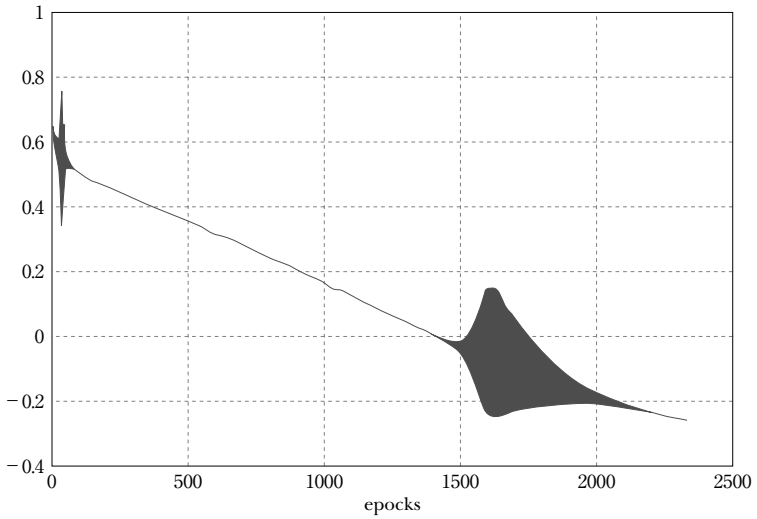


図 4 Input characteristics for the traditional methods  
in the output layer (0.05)

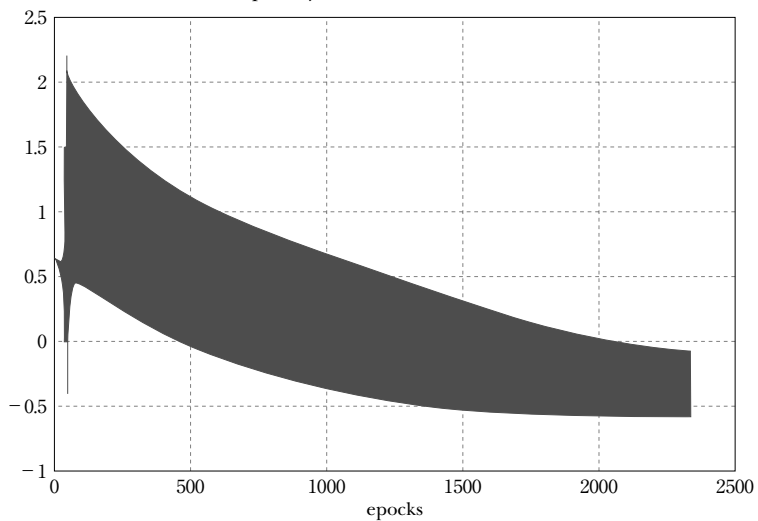


図 5 Input characteristics for the traditional methods  
in the output layer (0.032)

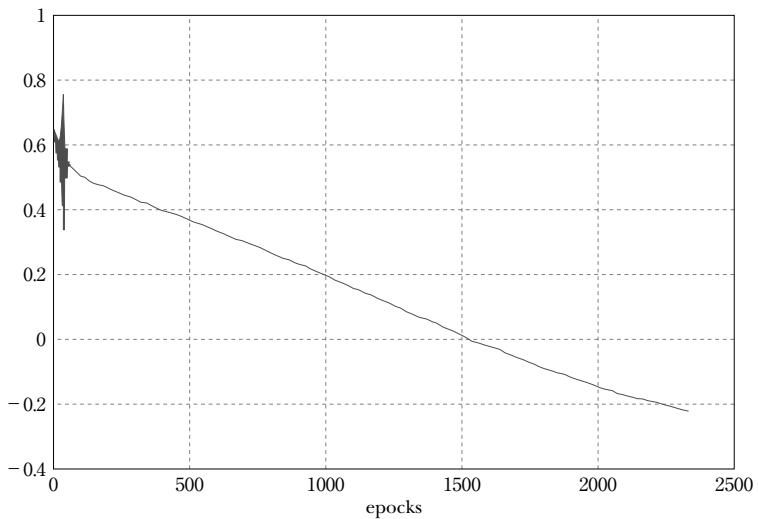
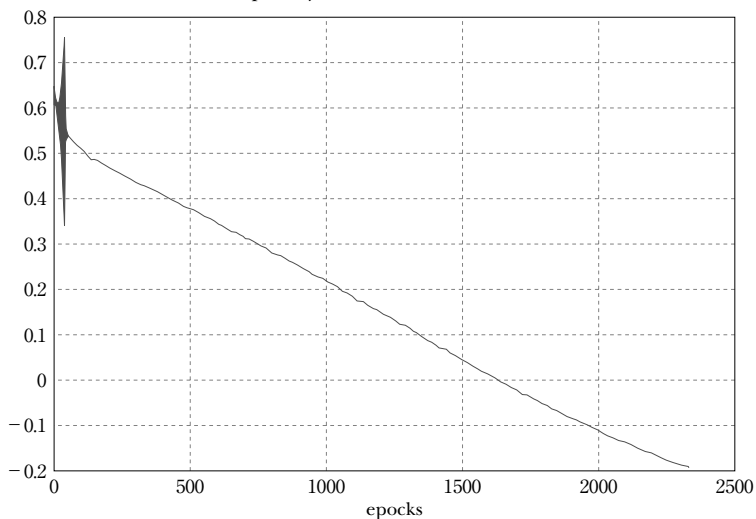


図 6 Input characteristics for the traditional methods  
in the output layer (0.03)



法や RPROP 法に提案手法を組み込んだ場合の学習の性能と計算時間および学習における振動の必要性について考察する。

#### QPROP 法や RPROP 法に提案手法を組み込んだ場合

表 4 より、Ridge 関数学習に対し、QPROP 法や RPROP 法に提案手法を組み込むことによって、*RMSE* の平均は、組み込む前と比較すると、ともに減少した。組み込む前の単独学習では、RPROP 法、提案手法、QPROP 法の順に誤差の平均値は、0.050、0.107、0.172、となるが、RPROP 法に提案手法を組み込むことによって、0.036 となり、最小となった。QPROP 法に組み込むことによる誤差も 0.172 から 0.058 まで減少した。QPROP 法や RPROP 法単独での出力層素子への入力特性は、振動がほとんど現われないが、提案手法を組み込むと、QPROP 法や RPROP 法単独の特性に、わずかな振動が加わった特性となった。わずかな振動の発生によって、QPROP 法や RPROP 法の *RMSE* の平均が改善されたのは、提案手法の振動発生の効果と言える。時間に関しても、論文<sup>(2)</sup> で述べたように、教師データの選択効果が発揮されている。また、表 5 より、



Schwefels 関数の学習に対しても、QPROP 法に提案手法を組み込んだ場合  $RMSE$  の平均が最小になり、時間も QPROP 法の約 62.8% となり、改善が行われている。

### 提案手法と振動の必要性

表 6、表 7、表 8 において、○印がつけてある場合は、学習の最終段階において、振動現象が存在することを示している。一般的な学習においては、学習率を小さくすることによって、振動を抑えることができる。学習の最終段階において、表 6、表 7、表 8 ともに、共通して言えることは、 $RMSE$  値が最小となる出力層素子への入力特性は、振動していることである。表 6 (提案手法) においては、学習係数が 0.6 の時、 $RMSE$  値 0.0269 となり、最小となっている。また、従来法を用い Rastrign 関数に対する学習を行った表 7 の結果も同様で、学習係数が 0.1 の時、 $RMSE$  値は、0.0246 となり、最小となっている。

表 8 は、従来法で学習係数を小さくし、Rastrign 関数に対する学習を行ったものである。学習率が 0.04 から 0.05 に増加すると、振幅が増加し、 $RMSE$  値は、0.053 から 0.050 に減少している。逆に、学習率が、0.04 から 0.032 に減少させた場合、0.04 から 0.030 に減少させた場合は、ともに、 $RMSE$  値は、0.058、0.059 となり増加している。このときの出力層素子への入力特性は、図 5 および図 6 に示してあるように、単調減少特性となり、振動はしていない。図 3 に示す出力層素子への入力特性は、約 1,400 回学習終了後から振動が生じているが、単調減少特性となる学習率 (0.03、0.032) に比較すると、 $RMSE$  値は、0.056 であり、減少している。これらの結果は、出力層素子への入力特性に対し、従来法および提案手法で振動現象が存在する場合は、単調減少特性となる学習よりも  $RMSE$  値が減少することを意味する。

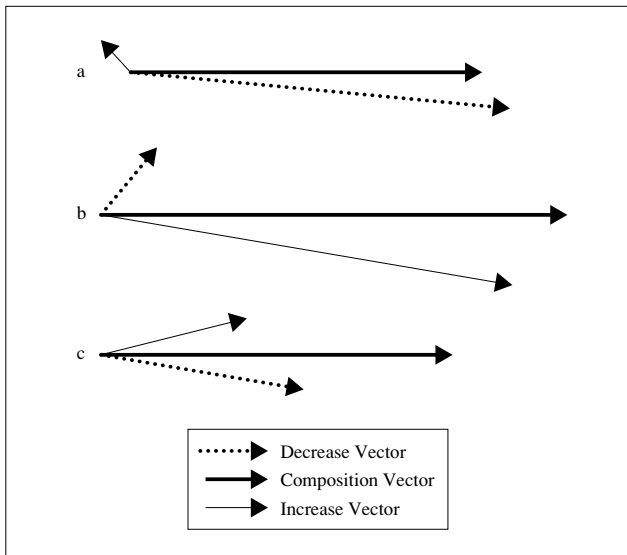
次に、表 1(c) で表わされる Ridge 関数に対して、重み更新量 (中間層と出力層間の重み更新量) をもとにして振動波形のベクトル的な解析を試みた。1 エポックの学習で、すべてのパターンに対し、重み更新量を前回の学習に対する絶対誤差と現在の学習に対する絶対誤差が、増加する

場合と減少する場合で、2分割し増加ベクトル、減少ベクトルを求め、それらの合成ベクトルも求めた。振動現象が止まり、学習が比較的進行しない場合のベクトルの代表的な関係を図7aに示す。減少ベクトルの大きさが、合成ベクトルにほぼ等しく、増加ベクトルが合成ベクトルに比較的影響を与えない状況が持続する場合には、出力層素子への入力特性は、振動特性にならない(単調減少特性)。

また、合成ベクトルに寄与する増加ベクトルの大きさや角度、減少ベクトルの大きさ、角度が学習回数に対して変化する(増加ベクトルが合成ベクトルに大きく寄与するか、または、減少ベクトルが合成ベクトルに大きく寄与するかのどちらか、そして、周期は不規則)場合には、出力層素子への入力特性は、雑音を含むような不規則な振動特性(図7b、c)ベクトルのような変化が常に起こり、繰り返される)となることも実験において確認された。増加パターン数、減少パターン数の変化が多く生じる場合には、不規則な振動特性となる。

単調減少特性の場合には、増加パターン数および減少パターン数の変

図7 The weight renewal vector for the proposed methods



動数が学習回数の変化に対して非常に少ない。また、良好な学習が行われるためには、減少ベクトルだけではなく、増加ベクトルの存在が必要で、論文<sup>(2)</sup>で示した良好な振動が生じている場合に、誤差が減少することとなる。

本論文において、提案手法、RPROP法やQPROP法（単独では振動は発生しにくい）に組み込んだ場合において、発生する振動は、学習が複雑になればなるほど発生しやすくなるために、複雑な学習においては、提案手法はより有効（誤差および時間）な方法であると考えられる。

## 4. おわりに

本論文では、NNを用いた学習において、学習する前に教師データを学習のしやすさに着目した分類、その分類に基づく多段階学習、誤差の大きさに応じた学習係数の動的調整を特徴とした総合的な学習方法を利用し、その方法をQPROP法やRPROP法に組み込み実験を行った。その結果、RMSE値が減少し、学習時間も改善された。

さらに、従来から、オーバーシュートが起こることは、学習時間を増加させることで、学習を悪化させるために悪いものと思われていた。本論文では、振動が生じていたほうが単調減少特性に比較して学習が改善されることをRMSE値および出力層素子への入力特性をもとに確認した。

ここでは、さらに、中間層と出力層間の重み更新量をもとにした絶対値誤差の増加、減少をもとにしたベクトルを導入し、ベクトル変化から出力層素子への入力特性の違い（振動している場合と振動していない場合）も検討し、ベクトルの動きが活発な場合（増加ベクトル、減少ベクトルの大きさ、角度が頻繁に変わる）には、雑音性の振動が生じ、学習が進んでいることを示した。雑音性の振動は、提案手法の特徴で、2個の工夫から生じたものであり、この工夫が、複雑な学習を可能にしたと考えられる。

今後は、本手法に加え、振動特性を積極的に利用し、パターン認識問題に対する効果に対しても提案手法を拡大し、さらなる誤差の減少も検

討したい。

(参考文献)

- (1) 元木誠、小坏成一、平田廣則「パルスネットワークのための入出力パルスのタイミングを調節する教師あり学習則」『信学論 (D-II)』Vol. J89-D-II, No. 12, 2006年、726-734ページ。
- (2) 田口功、須貝康雄「教師データの選択と出力層素子への入力特性に基づくニューラルネットワークの効率的学習法」『電学論C』Vol. 129-C, No. 4, 2009年、1208-1213ページ。
- (3) 田口功、須貝康雄「出力層素子の入力特性とそれに基づくニューラルネットワークの学習の効率化」『電気学会電子・情報・システム部門講演論文集』、2004年、931-934ページ。
- (4) D. E. Felleman, "An Empirical Study of Learning Speed in Back-Propagation Network," Technical Report CMU-CS-88-162, Carnegie-Mellon University, Computer Science Dept., 1988.
- (5) M. Riedmiller and H. Braun, "A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm," Proc. ICNN, San Francisco, 1993.
- (6) 松井伸之、石見憲一「しきい値ゆらぎをもつニューロンモデルを用いた階層型ニューラルネットワーク」『電学論C』Vol. 114-C, No. 11, 1994年、1208-1213ページ。
- (7) 熊沢逸男『学習とニューラルネットワーク』、電子情報通信シリーズ、森北出版株式会社、1998年、57ページ。
- (8) 八名和男監訳『ニューラルコンピューティング入門』、海文堂、1993年、77-79ページ。
- (9) 金丸隆志、関根優年「ニューラルネットワークの canonical model がみせる振動同期現象」『信学技報』NC2003-138、2004年、17-22ページ。
- (10) Ting Wang、須貝康雄「非線形多変数関数近似のためのウェーブレットニューラルネットワーク」『電学論C』Vol. 120-C, No. 2, 2000年、185-193ページ。
- (11) 須貝康雄、堀部浩、川瀬太郎「基準需要を利用したニューラルネットワークによる翌日最大電力需要予測」『電学論B』Vol. 117-B, No. 6, 1997年、872-879ページ。
- (12) 梅原宗一、山崎輝、須貝康雄「サポートベクタマシンとニューラルネットワークに基づく降水量推定システム」『信学論 (D-II)』Vol. J86-D-II, No. 7, 2003年、1090-1098ページ。
- (13) Charles K. Chui, *An Introduction to Wavelets*, Academic Press, 1992.
- (14) Ting Wang and Yassui Sugai: "A Wavelet Neural Network for the Approximation of Nonlinear Multivariable Functions," Proc. of IEEE International Conference on System, Man, and Cybernetics, III, 1999, pp. 378-383.
- (15) L. K. Jones, "Constructive Approximations for Neural Networks by Sigmoidal

- Functions,” Proc. IEEE, Vol. 78, No. 10, 1990.
- (16) B. Irie and S. Miyake, “Capabilities of Three Layered Perceptrons,” Proc. ICNN, Vol. 1, 1988, pp. 641–648.
- (17) K. Funahashi, “On the Approximate Realization of Continuous Mapping by Neural Networks,” Vol. 2, No. 3, 1989, pp. 183–192.