

報 文

変傾共役勾配法による BP 学習の安定化と高速化

後藤 正幸*, 開沼 泰隆**, 俵 信彦***

階層型ニューラルネットワークの学習法である BP 学習において、学習速度とローカルミニマの問題がある。学習の高速化に関しては、共役勾配法が提案されており、その高速性が明らかになっているが、初期値依存性の問題も合わせ持つ。一方、初期値依存性に対しては、変傾法が提案されているが、本質的に最急降下法であるため、学習速度の面で改良の余地がある。そこで、本研究では変傾法の更新を共役方向に実現する変傾共役勾配法を提案する。さらに、その挙動に癖が見られることを指摘し、この癖を利用して高速化を図った改良型変傾共役勾配法を提案し、化学プロセスの制御問題に適用してその有効性を示す。

Variable Slope-Conjugate Gradient Method for BP Learning Algorithm

Masayuki GOTOH,* Yasutaka KAINUMA** and Nobuhiko TAWARA***

BP learning algorithm from Conjugate Gradient method is an effective method for high speed learning. However, in this method, the problem of its dependency on learning initial values should be mentioned. Variable Slope is effective for learning with accuracy. Therefore, in this paper, we propose a learning algorithm that is a combination of the Conjugate Gradient and Variable Slope methods. Moreover, we propose Improved Variable Slope-Conjugate Gradient algorithm that uses the trend of Variable Slope-Conjugate Gradient algorithm. In order to verify this proposal we performed a simulation experiment using a data for the production control of a chemical plant.

1. はじめに

階層型ニューラルネットワークの学習法であるバックプロパゲーション学習(以下 BP 学習)[1]において、①学習速度が遅い、②学習結果が初期値に依存してよい結果が得られないことがある(学習の不安定性:ローカルミニマの問題)などの不備が指摘されている[2]~[11]。

学習の高速化に関しては、共役勾配法が提案されており、最急降下法に比べて高速であることが明らかになっている[2]~[9]。しかし、共役勾配法においても、①学習の初期値に大きく依存する場合がある、②直線探索に多くの計算が必要になる、などの問題点も残っている。そこで、本研究では、共役勾配法を導入した BP 学習について、より高速で安定性の高い学習アルゴリズムを構築することを目的とする。

具体的には、まず、変傾法[9]、[10]の更新を共役方向に実現する変傾共役勾配法(VSCG)を提案して初

期値依存性の問題を解決し、変傾共役勾配法が学習の安定性の面で有効となることを示す。ここで、変傾共役勾配法ではその挙動に一定の癖がみられ、この癖を積極的に利用すれば、直線探索部分の計算量を低減できることを指摘する。そこでさらに、挙動の癖を利用した改良型変傾共役勾配法(改良型 VSCG)を提案する。最後に、これらの学習法を化学プロセスの制御介入ニューラルネットワークモデル構築に適用することにより、改良型変傾共役勾配法が学習の安定性と同時に、学習速度の面でも有効であることを示す。

2. 従来の BP 学習法

2.1 ニューラルネットワークモデル

まず、本研究で対象とするネットワークを示す。 m 層のネットワークにおいて、ある入力パターンが与えられたとき、 k 層の第 j ユニットへの入力の総和を i_j^k 、出力を o_j^k とし、 $k-1$ 層の第 i ユニットから k 層の第 j ユニットへの結合の重み(結合ウェイト)を $w_i^{k-1,j}$ 、しきい値を b_j^k とすると、

$$i_j^k = \sum_i w_i^{k-1,j} o_i^{k-1} + b_j^k \quad (1)$$

$$o_j^k = f(i_j^k, s_j^k) \quad (2)$$

$$f(x, s) = \frac{1}{1 + \exp(-sx)} \quad (3)$$

* 早稲田大学(Waseda University)

** 東京都立商科短期大学(Tokyo Metropolitan College of Commerce)

*** 武蔵工業大学(Musashi Institute of Technology)

受付: 1994年8月1日, 再受付(1回)

受理: 1994年12月16日

で与えられる。ここで、 x はユニットへの入力、 s_j^k は傾きを決定する値である。入力 x 、傾き s による出力関数の偏微分は、それぞれ次のようになる。

$$\frac{\partial f(x, s)}{\partial x} = sf(x, s)\{1 - f(x, s)\} \quad (4)$$

$$\frac{\partial f(x, s)}{\partial s} = xf(x, s)\{1 - f(x, s)\} \quad (5)$$

2.2 BP 学習法

Rumelhart[1]によるBP学習における、ウェイトの修正は次式で与えられる。これは、本質的には二乗誤差に対する最急降下法であり、傾き s_j^k は定数として与えられる。学習の評価関数である二乗誤差 E は、 y_{jp} をサンプル p の実測値(正解)として、

$$E = \sum_p E_p \quad (6)$$

$$E_p = \frac{1}{2} \sum_j (o_{jp}^m - y_{jp})^2 \quad (7)$$

で定義される。 E_p は学習サンプル p に対する二乗誤差である。また、 o_{jp}^m や y_{jp} の添字 p はサンプル p に対するものであることを示すが、以下では混同することがないため、この添字は省略する。

t 回目のウェイトの更新においてサンプル p が与えられたとき、ウェイトの修正式は次式で与えられる。

$$w_i^{k-1j}(t) = w_i^{k-1j}(t-1) + \Delta w_i^{k-1j}(t-1) \quad (8)$$

$$\begin{aligned} \Delta w_i^{k-1j}(t) &= -\varepsilon \frac{\partial E_p}{\partial w_i^{k-1j}} + \alpha \Delta w_i^{k-1j}(t-1) \\ &= -\varepsilon d_j^k o_i^{k-1} + \alpha \Delta w_i^{k-1j}(t-1) \end{aligned} \quad (9)$$

ε 、 α は小さな正定数とする。 $-\frac{\partial E_p}{\partial w_i^{k-1j}}$ の部分をすべてのウェイトについてまとめてベクトル表現したものが、サンプル p の二乗誤差に対するウェイトの最急降下方向であり、これを $-\nabla E_{wp}$ で表す。また、逆伝搬誤差は次式で計算される。

$$\begin{aligned} k=m \text{ のとき} \\ d_j^m &= (o_j^m - y_j) \\ s_j^m f(i_j^m, s_j^m) \{1 - f(i_j^m, s_j^m)\} \end{aligned} \quad (10)$$

$$\begin{aligned} k \neq m \text{ のとき} \\ d_j^k &= \left(\sum_t w_{jt}^{k+1} d_t^{k+1} \right) \\ s_j^k f(i_j^k, s_j^k) \{1 - f(i_j^k, s_j^k)\} \end{aligned} \quad (11)$$

2.3 変傾法(VS法: Variable Slopes)

変傾法は、各ユニットの入出力関数の傾き s についても学習される方法である。変傾法では、探索空間が広がるため、誤差を小さくする機会が増えると考えら

れる。

ウェイトの修正は、式(9)~式(11)で行う。また、ウェイトの学習で計算した d_j^k を利用して、傾き s の最急降下方向を生成することができる。

$$\frac{\partial E_p}{\partial s_i^k} = d_i^k \frac{i_i^k}{s_i^k} \quad (12)$$

したがって、傾き s に対する慣性項付き修正式は、以下のようなになる。 β 、 ρ は正定数である。

$$\Delta s_i^k(t) = -\beta d_i^k \frac{i_i^k}{s_i^k} + \rho \Delta s_i^k(t-1) \quad (13)$$

変傾法は学習の安定性が向上するが、学習パラメータが多いという問題点がある。

2.4 共役勾配法(CG法: Conjugate Gradient)

共役勾配法は共役方向に直線探索する方法であり、以下の式でウェイトを更新する。

$$\begin{aligned} \Delta w(t) &= \alpha(t) \left\{ -\nabla E_w(t) + \beta_w(t) \frac{\Delta w(t-1)}{\alpha(t-1)} \right\} \\ &\stackrel{\text{def}}{=} \alpha(t) d_w(t) \end{aligned} \quad (14)$$

ただし、 $\alpha(t)$ は適当な直線探索により求める歩み幅であり、 $\beta_w(t)$ はPolak-Ribiere-Polyakの公式により求める。

$$\beta_w(t) = \frac{\nabla E_w(t)^T \{\nabla E_w(t) - \nabla E_w(t-1)\}}{\|\nabla E_w(t-1)\|^2} \quad (15)$$

また、 $-\nabla E_w$ は各サンプルの二乗誤差に対するウェイトの最急降下方向 $-\nabla E_{wp}$ のベクトル和である。

$$-\nabla E_w(t) = -\sum_p \nabla E_{wp}(t) \quad (16)$$

本研究では直線探索として、次の囲い込みと2次補間法を使用する。これらは最も基本的でかつ簡潔であり、適用範囲も広い直線探索法と考えられる。

(囲い込み 2次補間法)

十分小さい値から α を少しずつ増加させていき、

$$\begin{cases} E(\alpha_1) \geq E(\alpha_2) \leq E(\alpha_3) \\ \alpha_1 < \alpha_2 < \alpha_3 \end{cases} \quad (17)$$

となる三点 $\alpha_1, \alpha_2, \alpha_3$ を見いだす。この三点において、 $E_1 = E(\alpha_1)$ 、 $E_2 = E(\alpha_2)$ 、 $E_3 = E(\alpha_3)$ を計算すれば、次式から二乗誤差 E を最小とする α の推定値 $\hat{\alpha}$ を得ることができる。

$$\hat{\alpha} = \frac{\{a_2^2 - a_3^2\}E_1 + \{a_3^2 - a_1^2\}E_2 + \{a_1^2 - a_2^2\}E_3}{2\{(a_2 - a_3)E_1 + (a_3 - a_1)E_2 + (a_1 - a_2)E_3\}} \quad (18)$$

共役勾配法は、大次元の探索問題に対して特に有効となる方法である。

3. 変傾共役勾配法による学習の安定化

3.1 変傾共役勾配法(VSCG法: Variable Slope-Conjugate Gradient)

変傾法の探索を共役方向で実現すれば、高速で安定性の高い学習が期待できる。さらに、共役勾配法は大次元の探索問題で有効となるため、探索次元を増やせる変傾法との相乗効果も期待できる。また、変傾法では設定する学習パラメータの数が多いという欠点があるが、この点も共役勾配法を用いれば解決できる。そこで、本節では、変傾法と共役勾配法を組み合わせた変傾共役勾配法を提案し、その有効性を検証する。

変傾共役勾配法の修正式を以下に示す。

$$\begin{aligned} \Delta w(t) &= \alpha(t) \left\{ -\nabla E_w(t) + \beta_w(t) \frac{\Delta w(t-1)}{\alpha(t-1)} \right\} \\ &\stackrel{\text{def}}{=} \alpha(t) d_w(t) \end{aligned} \quad (19)$$

$$\begin{aligned} \Delta s(t) &= \alpha(t) \left\{ -\gamma \nabla E_s(t) + \beta_s(t) \frac{\Delta s(t-1)}{\alpha(t-1)} \right\} \\ &\stackrel{\text{def}}{=} \alpha(t) d_s(t) \end{aligned} \quad (20)$$

ただし、 $\alpha(t)$ は適当な直線探索により求める歩み幅であり、 $\beta_w(t)$ と $\beta_s(t)$ は次の公式により求める。

$$\beta_w(t) = \frac{\nabla E_w(t)^T \{ \nabla E_w(t) - \nabla E_w(t-1) \}}{\| \nabla E_w(t-1) \|^2} \quad (21)$$

$$\beta_s(t) = \frac{\nabla E_s(t)^T \{ \nabla E_s(t) - \nabla E_s(t-1) \}}{\| \nabla E_s(t-1) \|^2} \quad (22)$$

また、 $-\nabla E_s$ は各サンプルの二乗誤差に対する傾きの最急降下方向 $-\nabla E_{sp}$ のベクトル和である。

変傾共役勾配法では傾きとウェイトで別々に共役方向を計算して探索する方法をとり、その際ウェイトは通常どおりに探索するのに対し、傾きの最急降下ベクトルには1以下のパラメータ γ をかけて制御する。これは、ウェイトに比べて傾きの変化が激しくなってしまうことを抑えるためのものである。

変傾共役勾配法において、ウェイトと傾きで別々に共役方向を生成した理由は、①二乗誤差に対する変化率がウェイトと傾きで大きく異なる、②ウェイトの共役方向は傾きを固定して実現できるものであり、傾きが同時に大きく変化してしまうと、共役方向が生成できなくなるためである。したがって、ウェイトと傾きで共役方向を別々に計算し、その際傾きの変化にはパラメータ γ をかけて制御する方法が有効と考えられる。この場合には、傾きの変化方向は共役方向ではなく、ローカルミニマ

に捕まり、ウェイトがあまり更新されなくなってくると、傾きの変化方向は共役方向になり、傾きを更新することで誤差を減少させようとする。このため、本方法はローカルミニマの回避に有効と考えられる。

3.2 学習シミュレーション実験

変傾共役勾配法の有効性の検証のため、(0,1)で乱数発生させたデータを実際に学習させた。使用したデータは、入力5、出力1、20種類のデータである。中間層素子数はすべて4の3層ネットワークとし、学習回数は300回で実験をおこなった。

学習パラメータ γ については、 $1.0, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}, 10^{-7}, 10^{-8}$ の9通りを実験した。結合ウェイト w の初期値は $-0.30 \sim 0.30$ の範囲で一様乱数により発生させ、入出力関数の傾き s の初期値は、 $0.95 \sim 1.05$ の範囲で乱数発生させた。囲い込みの初期値は0.0001とした。以下に、シミュレーションの結果を示す。

表1は、300回の学習を終了した時点の二乗誤差の平均値と標準偏差である。変傾共役勾配法では、共役勾配法よりも二乗誤差の平均値は小さく、初期値によるばらつきも小さい。また、 γ はある程度小さい値の方が二乗誤差も小さく、ばらつきも小さいという傾向がみられる。これは、パラメータ γ を適切に設定することにより、変傾共役勾配法の安定性が向上することを示している。

以上の結果より、変傾法を導入する際には、パラメータ γ が重要な働きをすることがわかる。つまり、傾きの変化方向にパラメータ γ を掛けて大きな変化を抑えた本方法が有効であるといえる。この結果は、傾きを変化させることによって誤差を減じようとするのではなく、ウェイトの学習を助けていることを示している。つまり、基本的にはウェイト空間上で探索を行い、ローカルミニマに陥りそうになると、傾きの更

表1 学習終了時二乗誤差平均値まとめ

γ	平均値	標準偏差
共役勾配法	0.01930	0.03281
1.0	0.02504	0.01638
10^{-1}	0.00894	0.01139
10^{-2}	0.00412	0.00552
10^{-3}	0.00593	0.01029
10^{-4}	0.00586	0.00998
10^{-5}	0.00553	0.01042
10^{-6}	0.00360	0.00396
10^{-7}	0.00257	0.00333
10^{-8}	0.00387	0.00457

新によって誤差を減じようとする。このため、傾きが変化すると、ウェイト空間上ではローカルミニマを脱出しており、再びウェイト空間での探索が行えるものと考えられる。

また、表2は変傾共役勾配法の挙動の一例である。表3は各学習回数を組として、それらの相関係数を求めた表である。 $\beta_w(t)$ および $\|d_w(t)\|^2$ と $\alpha(t)$ との

表2 変傾共役勾配法の挙動の一例

回数	二乗誤差	β_w	$\ d_w\ ^2$	β_s	$\ d_s\ ^2$	α
1	0.7222	0.000	0.1820	0.000	0.0050	0.403
2	0.7208	-0.132	0.0002	-0.000	0.0000	2.976
3	0.7184	-0.011	0.0001	3.160	0.0000	47.396
4	0.7173	26.425	0.0699	3.083	0.0001	0.660
5	0.6714	0.130	0.0017	3.992	0.0007	46.353
6	0.6565	84.123	12.5167	29.082	0.6195	0.236
7	0.6364	1.071	14.5032	1.478	1.3617	0.281
8	0.6197	0.006	0.0974	5.015	34.4418	0.296
9	0.5984	-0.096	0.0037	0.000	0.0002	8.099
10	0.5924	4.102	0.0875	27.545	0.1318	0.461
50	0.0875	1.090	0.0078	0.047	0.0069	0.513
51	0.0872	0.000	0.0010	0.000	0.0066	0.532
52	0.0841	-0.741	0.0017	-0.027	0.0059	5.702
53	0.0821	6.202	0.0704	0.038	0.0063	0.650
54	0.0810	0.495	0.0195	-0.038	0.0049	1.034
55	0.0794	0.132	0.0008	0.280	0.0113	6.181
56	0.0769	12.306	0.1298	0.025	0.0084	0.718
57	0.0753	0.297	0.0132	0.043	0.0091	1.825
58	0.0726	1.008	0.0152	0.384	0.0230	3.112
59	0.0697	6.078	0.5721	-0.078	0.0084	0.535
60	0.0645	0.082	0.0052	-0.078	0.0075	7.942
61	0.0636	1.961	0.0229	-0.164	0.0018	0.566
62	0.0627	0.571	0.0095	0.068	0.0032	0.914
63	0.0607	0.413	0.0075	0.205	0.0055	3.354
64	0.0594	3.238	0.0828	0.457	0.0124	0.604
65	0.0587	0.395	0.0149	-0.067	0.0043	0.651
66	0.0580	0.307	0.0020	-0.062	0.0040	1.975
67	0.0562	3.875	0.0338	-0.061	0.0031	1.349
68	0.0552	1.185	0.0502	0.006	0.0035	0.756
69	0.0549	0.978	0.0022	0.177	0.0059	0.845
70	0.0543	0.432	0.0055	0.057	0.0055	1.439

表3 変傾共役勾配法の挙動 (相関係数) の一例

	β_w	$\ d_w\ ^2$	β_s	$\ d_s\ ^2$	α
β_w	1.000	0.644	0.364	0.269	-0.439
$\ d_w\ ^2$		1.000	0.465	0.522	-0.344
β_s			1.000	0.634	0.026
$\ d_s\ ^2$				1.000	-0.321
α					1.000

間には相関係数で-0.344, -0.439という相関があり、独立ではないことがわかる。また、挙動を詳しく調べてみると、表2のように、学習の初期段階では乱れることがあるが、 $\beta_w(t)$ や $\|d_w(t)\|^2$ が前回より大きくなると $\alpha(t)$ は小さくなり、 $\beta_w(t)$ や $\|d_w(t)\|^2$ が前回より小さくなると $\alpha(t)$ は大きくなるというように、挙動に一定の癖が見られる。確認のために他の様々なデータについても実験をした結果、すべてのケースでこの挙動の癖がみられたことから、この挙動は一般的に成り立つものと推測できる。

3.3 変傾共役勾配法の計算量に関する考察

変傾共役勾配法は、学習1回当たりの計算量は共役勾配法よりも増加する。本節では、変傾共役勾配法の計算量について考察する。

1回の代入を計算量の基本単位とし、ニューラルネットワークは入力層素子数 m 、中間層素子数 n 、出力層素子数1の3層ネットワークとする。学習サンプル数は p とする。また、直線探索においては、 s 回目の囲い込みで式(17)を見いだしたものと仮定する。

また、ウェイトのみの共役勾配法の結果をまとめると、学習1回当たりの計算量は次のようになる。

- 1) ネットワークの出力計算

$$m(m-1)+n(n-1)+4 \quad (23)$$

- 2) 最急降下ベクトルの計算

$$p[(4n+1)(n+1) + \{mn(2n+1)+1\}(mn+n)] + p-1 \quad (24)$$

- 3) β の計算

$$5n(m+2)+4 \quad (25)$$

- 4) 共役方向生成の計算

$$2n(m+2)+2 \quad (26)$$

- 5) 直線探索の計算

$$s[p\{m(m-1)+n(n-1)+4\} + p+2] + s+19 \quad (27)$$

傾きの更新が追加された変傾共役勾配法では、この計算量に、次の計算量が追加される。

$$(9+2p)(n+1)+p+n \quad (28)$$

式(28)より、学習サンプル数 p や中間層素子数 n が大きくなってくると、計算量の増加分が大きくなるが、式(24)のウェイトの最急降下方向の計算量に比べれば、少ないことがわかる。これは、ウェイトの最急降下方向生成時に計算した逆伝搬誤差を利用して、傾きの最急降下方向も計算できるためである。また、式(27)より、直線探索の計算量が大きく、直線探索部分の計算量を低減すれば、効果的に学習の高速化が図れることがわかる。

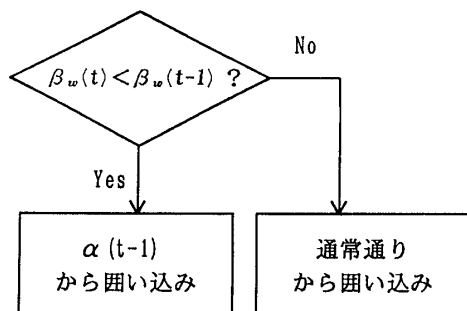


図1 改良型変傾共役勾配法

4. 改良型変傾共役勾配法(改良型 VSCG 法)による安定化と高速化

先に示した変傾共役勾配法では、表2に示したように、 $\beta_w(t)$ や $\|d_w(t)\|^2$ が前回より大きくなると $\alpha(t)$ は小さくなり、 $\beta_w(t)$ や $\|d_w(t)\|^2$ が前回より小さくなると $\alpha(t)$ は大きくなるという挙動の癖が見られる。この挙動の癖は、 β_w が計算された時点で $\alpha(t)$ の存在する領域におおよその見当をつけられることを示しており、この癖を積極的に利用すれば、直線探索の計算量低減が可能である。本節では、このような癖を利用して囲い込みの計算量を軽減するアルゴリズムとして、改良型変傾共役勾配法を提案する。

具体的には、 $\beta_w(t) < \beta_w(t-1)$ の場合には $\alpha(t) > \alpha(t-1)$ と考えられるので、次のように囲い込みの部分を改良する(図1参照)。

【改良型・囲い込み】

もし、 $\beta_w(t) < \beta_w(t-1)$ ならば、 $\alpha(t-1)$

から囲い込みを始め、それ以外は、十分小さな定数(通常通り)から囲い込みを始める。

この方法により、 $\beta_w(t) < \beta_w(t-1)$ の場合に、 $\alpha(t-1)$ までの探索過程を省略することができる。

また、提案方法では、 $\|d\|^2$ は新たに計算を追加する必要があることから、共役方向を生成する際にすでに得られている β_w を用いている。

5. 化学プロセス制御介入データの学習シミュレーション実験

5.1 構築モデル

本研究で対象とするのは、化学プロセスの制御問題である。このプロセスでは自動制御が導入されているが、ときとして不安定な状態に陥る場合もあるため、

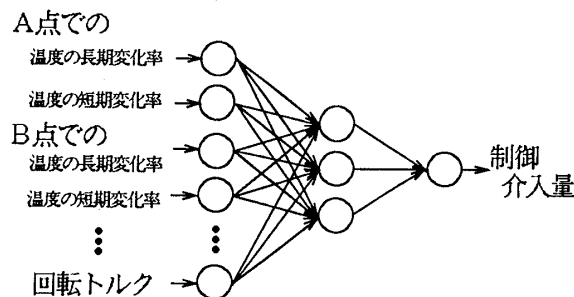


図2 制御介入ニューラルネットワークモデル

熟練オペレータがプロセスの状況を絶えず監視し、経験と勘で操作(介入という)を加えているのが現状である。プロセス制御の完全な自動化のために、このオペレータの介入回数を低減することが望まれている。そのため本研究では、この化学プロセスの制御介入を行うオペレータの知識モデルをニューラルネットで構築することを試みる。オペレータの介入をニューラルネットワークで代用できるならば、熟練オペレータ抜きで、安定した制御が可能となる。

具体的には、図2に示すようにプロセスの状態を示す変数を入力とし、オペレータによる制御介入量を出力するモデルを構築する。教師データとしてオペレータが実際に行った介入量を与えることで、このオペレータの知識を獲得したエキスパートシステムが構築できる。また、プロセスの状態を表す状態変数は主に各設置点での温度であり、これらは時系列データとして得られるが、この温度の時系列データを長期的変化率と短期的変化率に変換し、これらをニューラルネットワークへの入力変数とした。

5.2 シミュレーションの条件

入力変数にはプロセスの状態を示す13変数、出力には介入量を取り、学習データは350サンプルを使用した。ニューラルネットワークは、入力層素子数13、中間層素子数3、出力層素子数1の3層ネットワークとした。ウェイトの初期値は $-0.10 \sim 0.10$ の範囲で乱数発生させた。囲い込みの初期値は0.0001、再出発の回数は20回とした。学習回数は300回とし、学習データには入力変数を平均0、分散1に規準化、出力を(0,1)に変数変換したものを使用した。 γ については、ある程度小さい値で6段階を設定して実験した。

5.3 シミュレーション結果および考察

300回の学習終了時における二乗誤差の初期値による平均値と標準偏差、学習300回の演算時間の初期値による平均値、標準偏差を表4から表6に示す。表

表4 共役勾配法

学習終了時二乗誤差		演算時間 (秒)	
平均値	標準偏差	平均値	標準偏差
0.8013	0.0422	2069.2	48.15

表5 学習終了時二乗誤差比較

γ	VSCG		改良型 VSCG	
	平均値	標準偏差	平均値	標準偏差
10^{-5}	0.7849	0.0268	0.7670	0.0230
10^{-6}	0.7958	0.0272	0.7665	0.0255
10^{-7}	0.7864	0.0276	0.7623	0.0208
10^{-8}	0.7915	0.0288	0.7657	0.0238
10^{-9}	0.7870	0.0267	0.7638	0.0222
10^{-10}	0.7855	0.0296	0.7659	0.0270

表6 演算時間比較 (秒)

γ	VSCG		改良型 VSCG	
	平均値	標準偏差	平均値	標準偏差
10^{-5}	2156.6	61.84	1634.0	34.23
10^{-6}	2089.4	59.43	1644.3	42.76
10^{-7}	2089.1	58.30	1625.1	49.09
10^{-8}	2089.1	59.13	1636.5	42.94
10^{-9}	2093.5	61.69	1621.1	38.55
10^{-10}	2084.4	56.79	1638.9	48.15

4 は共役勾配法の結果である。表5、表6 は VSCG, 改良型 VSCG の結果である。

表5 の二乗誤差の平均値の結果は、標準偏差を考慮すると有意な差がある。二乗誤差の平均値は、改良型変傾共役勾配法が最も小さく、共役勾配法が最も大きい。また、共役勾配法は初期値によるばらつきが大きく、変傾共役勾配法、改良型共役勾配法ではばらつきが抑えられており、学習の安定性の面で優れていることがわかる。また表6 にみられるように、改良型変傾共役勾配法が最も演算時間の平均値が小さく、学習速度の面でも優れている。 $\beta_w(t)$ が前回より小さくなると、 $\alpha(t)$ は大きくなるという性質が成り立っており、この挙動の癖が有効的に利用されているということが理解できる。

改良型変傾共役勾配法では、変傾共役勾配法よりもさらに小さい二乗誤差が得られたが、これは明らかに大きめの歩み幅で探索したことに起因する。通常の囲い込みによる直線探索では、十分小さな値から大きくなる方へ探索するため、極小が複数存在する場合には、歩み幅 $\alpha(t)$ の値が最も小さいものが得られる。しかし、この場合に、改良型変傾共役勾配法では大き

めの歩み幅が得られ、これが何らかの形でローカルミニマ回避を促進し、小さな二乗誤差が得られたものと考えられる。

大きめの歩み幅で探索する方法により、学習の安定性が向上する理由としては、以下の2点が考えられる。一つめは、ニューラルネットワークの学習においては、ウェイトの絶対値を大きくすることで二乗誤差を小さくしようとする傾向があること[12]である。大きめの歩み幅 $\alpha(t)$ で探索されれば、ウェイトの更新量が大きくなり、ウェイトの絶対値が大きい、つまり学習の初期点から遠いところまで探索が行き届くものと考えられる。改良型変傾共役勾配法では、このことが二乗誤差を小さくするという面で効果的に働いたものと考えられる。

もう一つは、 $\alpha(t)$ が次の共役方向に影響を与えることである。共役勾配法はもともと二次関数の最小化問題に対して導出された方法であり、BP 学習のような一般の関数の探索問題においては局所的に二次関数で近似できるという考えから、そのまま適用されているにすぎない。したがって、共役勾配法による BP 学習では、局所的に二次近似した二次関数上での共役方向を生成していることになる。 $\alpha(t)$ が大きいということは、探索点間の距離が離れることを意味しており、より広い範囲を二次関数と考えて次の共役方向を生成することにつながる[13]、[14]。このように大きめの $\alpha(t)$ によって、より大域的な情報が加味された共役方向が生成され、より小さな二乗誤差まで学習が進行したものと考えられる。

以上より、改良型変傾共役勾配法が安定、かつ高速な学習を達成していることが明らかとなった。また、本モデルにおいては、実験範囲では γ の大小による傾向は見られなかった。これは、探索曲面の性質に依存することは明らかであるが、直線探索のパラメータ(囲い込みの初期値、幅)によっても結果が異なるため、今後 γ の影響についてさらに調べる必要がある。

6. ま と め

変傾法を共役勾配法によって実現する変傾共役勾配法を提案し、その挙動に癖が見られることを指摘した。その上で、この挙動の癖を積極的に利用した改良型変傾共役勾配法を提案した。化学プロセスの制御介入データの学習シミュレーションの結果、改良型変傾共役勾配法が学習速度、学習の安定性の面で優れていることを確認した。

学習パラメータ γ の影響について調べることを、および挙動の癖が一般的であることを解析的に示すこと

が今後の課題である。

日頃よりご指導を戴く早稲田大学工業経営学科、平澤茂一教授、ならびに平澤研究室の皆様、俵研究室の皆様深く感謝致します。

参 考 文 献

- [1] Rumelhart, D., Hinton, G. and Williams, R.: "Parallel Distribution Processing, Exporation in the Microstructure of Cognition", MIT Press, Cambridge, MA, (1986)
- [2] Charalambous, C.: "Conjugate Gradient Algorithm for Efficient Training of Neural Networks", *IEE, Proc.*, pp. 301-310, Vol. 139, No. 3, (1992)
- [3] Barnard, E.: "Optimizaton for Training Neural Nets", *IEEE Trans. on Neural Networks*, pp. 232-240, Vol. 3, No. 2, (1992)
- [4] 吉田利信: "2次元探索共役勾配法を用いた多層神経回路網の高速学習法", *情報処理学会論文誌*, pp. 988-995, Vol. 31, No. 7, (1990)
- [5] 市橋秀友: "逆伝播誤差を用いる共役勾配法による階層型ファジィモデルの学習", *日本機械学会 FAN シンポジウム講演論文集*, No. 910-70, pp. 235-240, (1991)
- [6] 瀬戸山 徹, 川人光男, 鈴木良次: "多層神経回路モデルの新学習アルゴリズムと運動制御への応用", *電子情報通信学会技術研究報告*, MBE 88-170, (1988)
- [7] 山品 元, 熊本博光, 奥村 進, 池崎孝裕: "方向制御弁のニューラルネットワークによる故障診断", *精密工学会誌*, pp. 2291-2296, Vol. 56, No. 12, (1990)
- [8] Aryee, E. B. and Poehlman, W. F. S.: "A Neural-network-based System to Recognize Parasites/Seal Worms on Cod Fish Images", *Eng. Applic. Artif. Intell.*, pp. 341-350, Vol. 4, No. 5, (1991)
- [9] Soucek, B. and The IRIS Group: "Neural and Intelligent Systems Integration", John Wiley & Sons, Inc., (1991)
- [10] Sperduti, A. and Starita, A.: "Speed Up Learning and Network Optimization with Extended Back Propagation", *Neural Networks*, pp. 365-383, Vol. 6, (1993)
- [11] 後藤正幸, 開沼泰隆, 俵 信彦: "共役勾配法によるBP学習について", *日本経営工学会平成5年秋季大会予稿集*, pp. 138-139, (1993)
- [12] 芳我尚秀, 石川真澄: "各種構造学習法の構造化および汎化能力の比較", *電子情報通信学会技術研究報告*, NC93-20, (1993)
- [13] 後藤正幸, 俵 信彦: "共役勾配法を導入したBP学習における安定化に関する研究", *日本経営工学会誌*, pp. 70-77, Vol. 46, No. 1, (1994)
- [14] 後藤正幸, 俵 信彦: "二次近似範囲を調整する共役勾配法とBP学習への適用", *電子情報通信学会技術研究報告*, NC94-37, (1994)