ニューラルネットの統一的教師あり学習法

4L－1

山本祥弘
鳥取大学工学部知能情報工学科

1．はじめに

階層型ニューラルネットワーク(NN)の学習アルゴリズムとして、勾配法を用いない新しい誤差逆伝搬法(EBP 学習アルゴリズム)[1)~3)]が提案されている。その方法は

A) 各中間層出力に対する仮の教師信号を出力誤差の逆伝搬により決定する。

B) 仮の教師信号との出力誤差を最小とするように各重みパラメータを修正する。

の２点から構成される。さらにB) の重み修正に指数重み付き最小２乗法(EWLS)を用いた EBP-EWLS アルゴリズム[4),5)]を先に提案しているが、本論では、これが、簡単な工夫により相互結合型ＮＮに対しても適用出来ることを示す。

2．階層型ＮＮ

簡単のため次の３層ＮＮを考える。

$$y_k = f(x_k), \quad x_k = \bar{W}_{k-1}^T \bar{a}_k \tag{1a}$$

$$a_k = f(z_k), \quad z_k = \bar{V}_{k-1}^T \bar{u}_k \tag{1b}$$

ここに添字ｋは学習パターンを表し、y_kはｐ次元出力、u_kはｎ次元入力、a_kはｍ次元の中間層出力、fはシグモイド関数のような連続単調増加関数で逆関数が存在するものとする。上付きバーはしきい値を含むことを示す。

すでに発表されている EBP 学習アルゴリズムは以下の通りである。

［ＥＢＰアルゴリズム（逐次型）］

1) 各入力 u_k.に対して(1)式を計算する。

2) $\triangle a_k = W_{k-1}(W_{k-1}^T W_{k-1})^{-1} e_{wk}^T$ or (2a)

$$\triangle a_k = (W_{k-1} W_{k-1}^T)^{-1} W_{k-1} e_{wk}^T \tag{2a'}$$

$$e_{wk} = f^{-1}(d_k) - \bar{W}_{k-1}^T \bar{a}_k \tag{2b}$$

$$a_{tk}' = a_k + \triangle a_k, \tag{3}$$

$$a_{tk} = \mathrm{H} a_{tk}' \tag{4}$$

3) $\triangle \bar{V}_{k-1} = \bar{u}_k(\bar{u}_k^T \bar{u}_k)^{-1} e_{vk}^T$ (5a)

$$e_{vk} = f^{-1}(a_{tk}) - \bar{V}_{k-1}^T \bar{u}_k \tag{5b}$$

$$\bar{V}_k = \bar{V}_{k-1} + \triangle \bar{V}_{k-1} \tag{5c}$$

4) $\triangle \bar{W}_{k-1} = \bar{a}_k^*(\bar{a}_k^{*T} \bar{a}_k^*)^{-1} e_{wk}^T$ (6a)

$$e_{wk} = f^{-1}(d_k) - \bar{W}_{k-1}^T \bar{a}_k^* \tag{6b}$$

$$\bar{W}_k = \bar{W}_{k-1} + \triangle \bar{W}_{k-1} \tag{6c}$$

ここに 2)が A)仮の教師信号決定法であり、(2a)式は出力誤差を零とするように決められる。(4)式のＨは a_{tk}が Vの修正により実現可能となるために、非線形関数 fの値域に属するように修正する操作を表し、具体的には値域を越えた分をカットする操作、あるいは原点を内部に含むときはベクトルの縮小を表す。また、a_k^*は V修正後の中間層の実際の出力であるが、ｐ≦ｍの場合には a_{tk}と一致している。次に、3)が B)重み修正を表す。本手法では中間層にも（仮の）教師信号が与えられるので、各層ごとに２層ＮＮとして重みが修正できる。実際、(5), (6)式は

$$J_{vk} = \| f^{-1}(a_{tk}) - \bar{V}_{k-1}^T \bar{u}_k \| \tag{7}$$

$$J_{wk} = \| f^{-1}(d_k) - \bar{W}_{k-1}^T \bar{a}_k^* \| \tag{8}$$

を最小（＝０）とするように定められる。これらは直交射影の考え方であり、各学習データごとに修正される逐次修正法である。この考え方は一括処理法[3)]にも拡張される。ところで、例えば適応アルゴリズムのような他のデータ処理では、パラメータの修正は逐次的であるが、その評価は一括的である。そこで、(7), (8)式に相当する過去の評価すべてを一括して、

$$J_{wk} = \frac{1}{2}\sum_{j=0}^{k-1} \rho^j e_{w,k-j}^T e_{wk-j} \tag{9a}$$

$$e_{wk-j} = f^{-1}(d_{k-j}) - \bar{W}^T \bar{a}_{k-j}^* \tag{9b}$$

$$J_{vk} = \frac{1}{2}\sum_{j=0}^{k-1} \rho^j e_{vk-j}^T e_{vk-j} \tag{10a}$$

$$e_{vk-j} = f^{-1}(a_{k-j}^*) - \bar{V}^T \bar{u}_{k-j} \tag{10b}$$

を評価とするように $\bar{W}_k$, $\bar{V}_k$を求めることを考える。ここに、ｊ＝０以外の過去の評価分はすでに行った重みの修正により正確なものでない。このために、指数重み ρ（$0<\rho<1$）が導入されている。これらの評価を最小とする $\bar{W}_k$, $\bar{V}_k$の逐次型修正法はよく知られており、以下のようになる。

［ＥＢＰ-ＥＷＬＳアルゴリズム（階層型）］

1), 2) は EBP と同じ。

3) $\triangle \bar{V}_{k-1} = \dfrac{P_{k-1}\bar{u}_k}{\rho + \bar{u}_k^T P_{k-1} \bar{u}_k} e_{vk}^T$ (11a)

$$e_{vk} = (f^{-1}(a_{tk}) - \bar{V}_{k-1}^T \bar{u}_k) \tag{11b}$$

$$P_k = \frac{1}{\rho}\left(P_{k-1} - \frac{P_{k-1}\bar{u}_k \bar{u}_k^T P_{k-1}}{\rho + \bar{u}_k^T P_{k-1} \bar{u}_k}\right) \tag{12}$$

$$\bar{V}_k = \bar{V}_{k-1} + \triangle \bar{V}_{k-1} \tag{13}$$

4) $\triangle \bar{w}_{k-1} = \dfrac{Q_{k-1}\bar{a}_k^*}{\rho + \bar{a}_k^{*T} Q_{k-1} \bar{a}_k^*} e_{wk}^T$ (14a)

$$e_{wk} = (f^{-1}(d_k) - \bar{W}_{k-1}^T \bar{a}_k^*) \tag{14b}$$

$$Q_k = \frac{1}{\rho}\left(Q_{k-1} - \frac{Q_{k-1}\bar{a}_k^* \bar{a}_k^{*T} Q_{k-1}}{\rho + \bar{a}_k^{*T} Q_{k-1} \bar{a}_k^*}\right) \tag{15}$$

$$\bar{W}_k = \bar{W}_{k-1} + \triangle \bar{W}_{k-1} \tag{16}$$

ただし、$P_0 = Q_0 = \alpha I$　、$\alpha = 10^{6\sim9}$ (17)

3．相互結合型ＮＮ

完全相互結合型の次式を考える。ただし、その

Unified supervised learning algorithm of neural networks
Yoshihiro Yamamoto
Faculty of Engineering, Tottori University
Koyama, Tottori, 680-0945, Japan

他の多くのタイプに対しても適用可能である。

$$y_O(k+1) = f_O(s_O(k+1)), \quad (18a)$$

$$s_O(k+1) = \bar{W}_{OO}{}^T \bar{y}_O(k) + W_{OH}{}^T y_H(k) + W_{OI}{}^T u(k) \quad (18b)$$

$$y_H(k+1) = f_H(s_H(k+1)), \quad (19a)$$

$$s_H(k+1) = \bar{W}_{HO}{}^T \bar{y}_O(k) + W_{HH}{}^T y_H(k) + W_{HI}{}^T u(k) \quad (19b)$$

ここにO, H, Iはそれぞれ出力、隠れ、入力ユニットを表し、$u(k)$、$y_O(k)$はNNの入出力、$y_H(k)$は隠れユニットの出力である。各ユニットのしきい値を便宜上$y_O(k)$に含ませており、

$$\bar{y}_O{}^T(k) = (1,\ y_O{}^T(k)),\quad \bar{W}_{OO}{}^T = (\theta_O,\ W_{OO}{}^T),$$
$$\bar{W}_{HO}{}^T = (\theta_H,\ W_{HO}{}^T) \quad (20)$$

である。ここで、(19a,b)式の時間をシフトした次式を考える。

$$y_O(k+1) = f_O(s_O(k+1)), \quad (21a)$$

$$s_O(k+1) = \bar{W}_{OO}{}^T \bar{y}_O(k) + W_{OH}{}^T y_H(k) + W_{OI}{}^T u(k) = \bar{W}_O{}^T \bar{y}(k) \quad (21b)$$

$$y_H(k) = f_H(s_H(k)), \quad (22a)$$

$$s_H(k) = \bar{W}_{HO}{}^T \bar{y}_O(k-1) + W_{HH}{}^T y_H(k-1) + W_{HI}{}^T u(k-1) = \bar{W}_H{}^T \bar{y}(k-1) \quad (22b)$$

ただし、

$$\bar{W}_O{}^T = (\bar{W}_{OO}{}^T,\ W_{OH}{}^T,\ W_{OI}{}^T)$$
$$\bar{W}_H{}^T = (\bar{W}_{HO}{}^T,\ W_{HH}{}^T,\ W_{HI}{}^T) \quad (23)$$
$$\bar{y}(k) = (\bar{y}_O(k),\ y_H(k),\ u(k))$$

である。するとネットワークとして(18),(19)式と(21),(22)式は別物となり、学習過程も異なるが、学習結果の重みを固定すれば、同じ入力に対して同じ出力を与える意味で、両式は同じ働きをする。そこで(21),(22)式の重みの学習方式を考えることになるが、これらは形式的に３層NNの形をしていることに着目すると、NNの出力誤差から、$y(k)$、従って、隠れユニットの出力$y_H(k)$に対する仮の教師信号$y_{Ht}(k)$を求め、次に、(20a,b)と(20c,d)それぞれの重みを最小２乗法で求める方式が考えられる。これはまさに前節の方法と同じであり、以下の結果を得る。

［ＥＢＰ-ＥＷＬＳアルゴリズム(相互結合型)］

1) $y_O(k+1) = f_O(s_O(k+1))$,
   $s_O(k+1) = \bar{W}_O(k-1)^T \bar{y}(k)$

2) $\triangle y(k) = W_O(k-1)(W_O(k-1)^T W_O(k-1))^{-1} e(k)$
   $e(k) = f_O{}^{-1}(d(k+1)) - \bar{W}_O(k-1)^T \bar{y}(k)$
   $y_{Ht}{}'(k) = y_H(k) + \triangle y_H(k)$
   $y_{Ht}(k) = \mathrm{H}\, y_{Ht}{}'(k)$

3) $$\triangle \bar{W}_H(k-1) = \frac{P_H(k-1)\,\bar{y}(k-1)}{\rho + \bar{y}(k-1)^T P_H(k-1)\,\bar{y}(k-1)}\, e_H(k)$$
   $e_H(k) = f_H{}^{-1}(y_{Ht}(k)) - \bar{W}_H(k-1)^T \bar{y}(k-1)$
   $$P_H(k) = \frac{1}{\rho}\left(P_H(k-1) - \frac{P_H(k-1)\,\bar{y}(k-1)\,\bar{y}(k-1)^T P_H(k-1)}{\rho + \bar{y}(k-1)^T P_H(k-1)\,\bar{y}(k-1)}\right)$$
   $\bar{W}_H(k) = \bar{W}_H(k-1) + \triangle \bar{W}_H(k-1)$

4) $y_H{}^*(k) = f_H(s_H{}^*(k))$,
   $s_H{}^*(k) = \bar{W}_H(k)^T \bar{y}(k-1)$
   $\bar{y}^*(k)^T = (\bar{y}_O(k)^T,\ y_H{}^*(k)^T,\ u(k)^T)$

5) $$\triangle \bar{W}_O(k-1) = \frac{P_O(k-1)\,\bar{y}(k-1)}{\rho + \bar{y}(k-1)^T P_O(k-1)\,\bar{y}(k-1)}\, e_O(k)$$
   $e_O(k) = (f_O{}^{-1}(d(k+1)) - \bar{W}_O(k-1)^T \bar{y}^*(k)$
   $$P_O(k) = \frac{1}{\rho}\left(P_O(k-1) - \frac{P_O(k-1)\,\bar{y}(k-1)\,\bar{y}(k-1)^T P_O(k-1)}{\rho + \bar{y}(k-1)^T P_O(k-1)\,\bar{y}(k-1)}\right)$$
   $\bar{W}_O(k) = \bar{W}_O(k-1) + \triangle \bar{W}_O(k-1)$

6) $y_H(k+1) = f_H(s_H(k+1))$,
   $s_H(k+1) = \bar{W}_H(k)^T \bar{y}(k)$

7) k = k+1, 1)へ

４．例題

ここでは次の遅延のあるＸＯＲ問題を考える。

$$d(k+1) = u(k) \oplus u(k-1) \quad (24)$$

入力列を以下に与える。

$u(k)$={0 0 0 1 1 0 1 1 | 0 0 0 1 1 0 1 1 | ..}
$d(k)$={* * 0 0 1 0 1 1 | 0 1 0 0 1 0 1 1 | ..}

スカラ入出力NNで学習した結果を示す。

表１：遅延のあるＸＯＲ問題の学習結果

	中間層ユニット数＝５				同＝１０			
ρ	C	A	L	U	C	A	L	U
0.1	88	681.9	20	1984	69	652.7	28	1948
0.2	95	726.0	20	1979	97	698.6	28	1966
0.3	94	629.1	20	1847	98	572.2	28	1579
0.4	100	528.9	20	1963	99	517.8	28	1998
0.5	100	347.5	20	1692	100	434.7	28	1559
0.6	100	274.1	20	1602	100	364.3	28	1343
0.7	100	220.1	20	760	100	231.8	28	586
0.8	100	219.7	20	795	100	192.4	28	538
0.9	98	387.7	20	1535	100	254.0	28	1646
1.0	7	592.4	20	1291	11	1018.0	28	1803

C:収束率、A:平均 L:最小 U:最大学習回数

５．おわりに

本論文では、すでに提案している階層型NNに対する EBP-EWLS アルゴリズムを、相互結合型NNに適用できることを示した。完全相互結合型だけを示したが、Hopfield型、Jordan型、Elman型などに対しても可能であることを確認している。

参考文献

1) 山本：ﾆｭｰﾛ回路の学習規則と適応ｱﾙｺﾞﾘｽﾞﾑ、ｼｽﾃﾑ制御情報学会論文誌、Vol.7, 47/49, 1994,
2) 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学的性質、同上、Vol.9, 1/9, 1996,
3) 山本：一般化誤差逆伝搬法、第 50 回情処全大, 2-237/238, 1996.
4) 山本：階層型ＮＮに対するEBPアルゴリズムとその応用、第 56 回情処全大,2-351/352,1998.
5) 山本・坂本：指数重み付き最小２乗法によるＥBP学習ｱﾙｺﾞﾘｽﾞﾑ, SICE 論文集,34-8,1104/1111, 1998.