

# ディレイ素子を含むFIRニューラルネットワーク における共役勾配法の一検討

西 正明 生活科学教育講座

## 1 まえがき

ニューラルネットワークは、新しい情報処理技術である。この情報処理技術は、現在のノイマン型コンピュータと異なる原理のコンピュータの誕生、あるいは、ニューラルネットワークと現在のコンピュータ技術との融合を喚起し、新たな情報処理技術の形態を作り出す。ニューラルネットワークはパターンマッチング、パターン完全化、パターン分類等のパターンに関連した問題において特に優れた能力を発揮する。例えば、財務データから財務予測を行なったり、視覚イメージからロボットへどんな命令を出すかを決めたり、イメージをキーワードに翻訳したりするのに使うことができる。他にも音声認識や制御など、応用可能な分野は広い。本論文では、一般的な階層型バックプロバケーションニューラルネットワークを基本とし、その出力フィードバックにディレイ素子を含むFIRニューラルネットワークを用いる。その学習方法には、一般的に最急降下法が使われるが、ここでは共役勾配法を用い、出力フィードバックにディレイ素子を含むFIRニューラルネットワークにおける共役勾配法の効果を検討する。適用問題として、AM波とFSK波および矩形波におけるノイズ除去問題をとりあげ、ノイズを付加した波形を最急降下法と共役勾配法でそれぞれ学習させて、ノイズを付加した別の波形の認識出力波形と学習過程を比較検討する。

## 2 ニューラルネットワークの構成

バックプロバケーションニューラルネットワークは入力層、中間層、出力層からなるネットワークであり、中間層の数は任意である。本論文では、このネットワークの出力フィードバックにディレイ素子を含んだニューラルネットワーク<sup>1)</sup>を基本として、さらに図1に示すFIRシナプスをユニット直前に組み込んだ構成のニューラルネットワーク<sup>2)</sup>を扱う。3層の場合のディレイ素子を含むFIRニューラルネットワークの構成を図2に示す。ネットワークは前後の層間ユニットで完全結合しており、結合荷重が設定されている。この結合荷重を修正することによって、教師信号(望ましい出力)とネットワークの出力の誤差を減少させるように学習することができる。

## 3 学習アルゴリズム

バックプロバケーションの学習アルゴリズムは、入力層から中間層を経て出力層へと進行する順伝播と出力層から入力層へと進行する逆伝播の処理からなる。以下に、最急降下法と共役勾配法<sup>3)</sup>の学習アルゴリズムを示す。

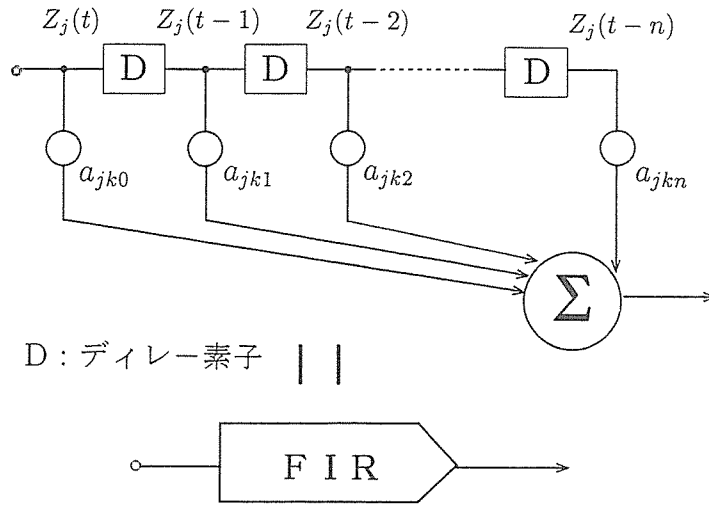


図1 FIR シナプス

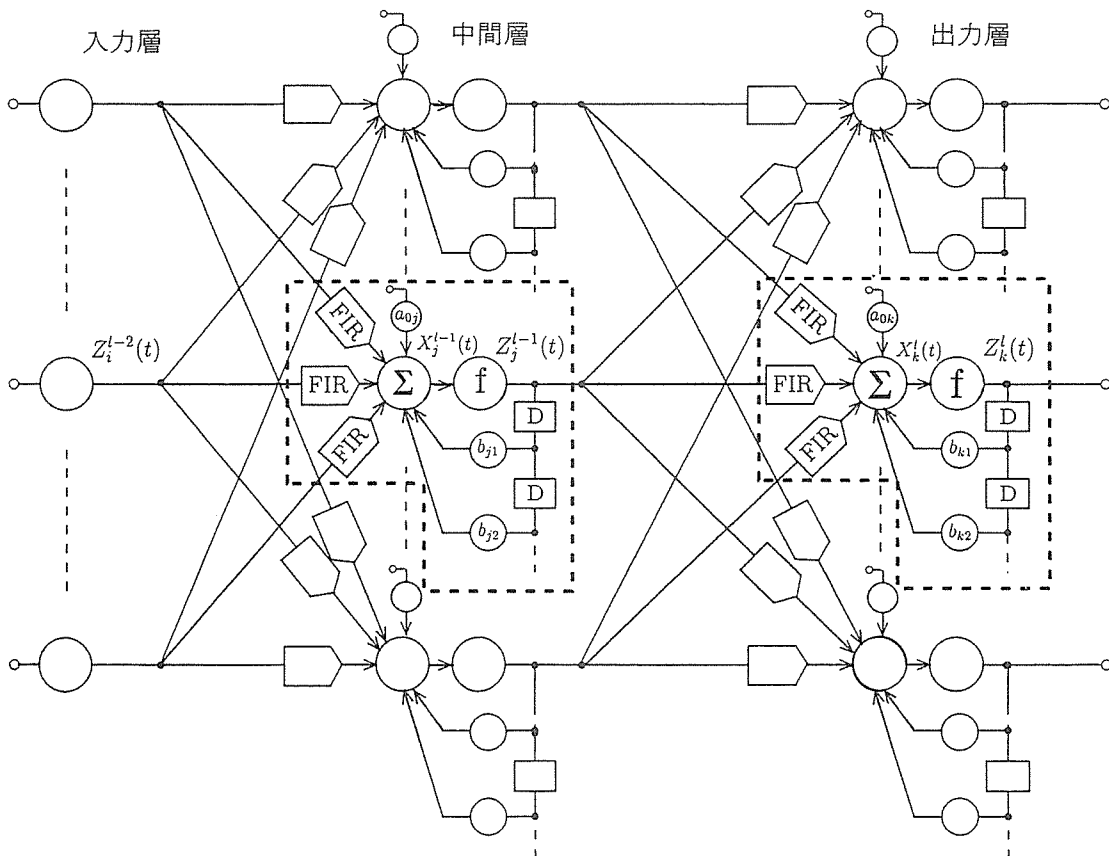


図2 出力フィードバックにディレー素子を含む FIR ニューラルネットワーク

### 3.1 最急降下法

最急降下法はバックプロパゲーション学習アルゴリズムにおいて従来からよく用いられている方法であり、以下に簡単に示す。

第  $l$  層の  $k$  番目のニューロンの内部状態  $X_k^l(t)$  と出力  $Z_k^l(t)$  を次式で求める。

$$X_k^l(t) = a_{0k} Z_0^{l-1}(t) + \sum_{j=1}^{N_j} \sum_{p=0}^L a_{j k p} Z_j^{l-1}(t-p) + \sum_{m=1}^M b_{k m} Z_k^l(t-m) \quad (1)$$

$$Z_k^l(t) = f(X_k^l(t)) = \frac{1}{1 + \exp(-2X_k^l(t)/u_0)} \quad (2)$$

式 (1) の第 1 項はオフセット部分、第 2 項は FIR シナプス部分、第 3 項は出力からのフィードバック部分の算出項である。式 (2) の  $u_0$  は応答関数の傾きを規定するパラメータである。結合荷重の修正量は、教師信号を  $d_k(t)$  とし平均自乗誤差を式 (3) とすれば、式 (4) で求められる。ここで、 $\gamma$  はオフセットの結合荷重の学習係数で、 $\alpha$  は FIR シナプスの学習係数で、 $\beta$  は出力からのフィードバックでの学習係数である。

$$E_k(t) = \frac{1}{2} \sum_k (d_k(t) - Z_k^l(t))^2 \quad (3)$$

$$\Delta a_{0k} = -\gamma \frac{\partial E_k(t)}{\partial a_{0k}}, \quad \Delta a_{j k p} = -\alpha \frac{\partial E_k(t)}{\partial a_{j k p}}, \quad \Delta b_{k m} = -\beta \frac{\partial E_k(t)}{\partial b_{k m}} \quad (4)$$

### 3.2 共役勾配法

前述の最急降下法では、パラメータを固定してしまうため、学習の経過で誤差に振動が生じ、不安定で学習効率が悪化する場合がある。共役勾配法は、一回の学習ごとに、パラメータを適切な値に変更して学習を進める方法である。適切なパラメータの値を求めるには、囲い込みと黄金分割法を用いた直線探索法を用いる。このような学習方法を用いることによって、誤差の不安定性が改善されて学習効率の向上が期待できる。以下に共役勾配法を簡単に示す。

まず、式 (3) を時間について正規化した正規化誤差を  $f(W)$  とし式 (5) とする。次に、この正規化誤差関数  $f(W)$  の  $W$  に関する微分を  $g(W)$  とし式 (6) とする。

$$f(W) = \frac{1}{N} \sum_{t=1}^N E_k(W, t) = \frac{1}{2N} \sum_t \sum_k (d_k(t) - Z_k^l(t))^2 \quad (5)$$

$$g(W) = \nabla f(W) = \frac{1}{N} \sum_t \nabla E_k(W, t) \quad (6)$$

以下に直線探索法を用いた共役勾配法を [1] に示す。直線探索をするのに用いる囲い込みと黄金分割法をそれぞれ [2], [3] に示す。

#### [1] 直線探索法

直線探索法のアルゴリズムを示す。

##### STEP0 [初期化]

目的関数  $f(W)$  を最小にする  $W$  の初期値  $W^{(0)}$  を与える。

$W^{(0)}$  を  $-0.5$  から  $+0.5$  の間の値でランダムに初期化する。ただし、ゲインは  $0$  から  $1$  の範囲の値で初期化する。

式 (6) で求められる  $g(W^{(0)})$  を用いて、初期方向ベクトルを  $D^{(0)} = -g(W^{(0)})$  とする。

## STEP1 [直線探索]

$f(W^{(k)} + \alpha_k D^{(k)})$  を最小にする  $\alpha_k$  を求める。

直線探索の方法としては、まず  $f(W)$  の最小値を含む区間  $[a^0, b^0]$  を [2] で示す“囲い込み”によって求め、次に [3] で示す“黄金分割法”によって最小値を与える  $\alpha_k$  を求める。

## STEP2 [W の更新]

$W^{(k+1)} = W^{(k)} + \alpha_k D^{(k)}$  を計算する。

## STEP3 [収束の判定基準]

$g_{k+1} = g(W^{(k+1)}) = \nabla f(W^{(k+1)})$  を求める。

次の2つのいずれかの条件を満たしたときには、学習を終了させる。

- ・ベクトル  $\nabla f(W^{(k+1)})$  の各要素の大きさ(絶対値)が  $\varepsilon$  以下である。
- ・ベクトル  $\nabla f(W^{(k+1)})$  の各要素の二乗の和の平方根が  $\varepsilon$  以下である。

上記以外の場合は、次のSTEP4に進む。なお、本論文では、学習回数が制限回数に達した場合、または誤差が設定値以下になった場合にも学習を終了するようにしている。

## STEP4 [探索の方向]

新しい方向ベクトル  $D^{(k+1)}$  を求める。

$p$  を全体の結合荷重の数とゲインの数の和として、以下の場合分けをする。

- ・  $k+1 = 0 \pmod{p}$  の場合  
 $D^{(k+1)} = -g(W^{(k+1)})$
- ・  $k+1 \neq 0 \pmod{p}$  の場合  
 $D^{(k+1)} = -g(W^{(k+1)}) + \beta_k D^{(k)}$

ここで、 $\beta_k$  は以下のPolak-Ribiereの式(7)を用いて求める。

$$\beta_k = \frac{g_{k+1}^T \cdot (g_{k+1} - g_k)}{g_k^T g_k} = \frac{g_{k+1} \cdot (g_{k+1} - g_k)}{g_k^2} \quad (7)$$

## STEP5

$k+1 = k$  として、STEP1に戻る。

## [2] 囲い込み

関数  $f(x)$  において、任意に初期値が与えられたとき、最小点を含むような区間を求める方法として囲い込みを用いる。 $f(x_1) \geq f(x_2) \leq f(x_3)$  (等号は片方のみ) で、 $x_1 < x_2 < x_3$  を満たす3点  $x_1, x_2, x_3$  を求める。囲い込みのアルゴリズムを以下に示す。

## STEP0

$x_0 \in R_1, h \in R_1$  とする。

## STEP1

$x_1 = x_0 + h$  において  $f(x_1)$  を計算し、以下の場合分けに従う。

- ・  $f(x_1) \leq f(x_0)$  ならば、 $k = 1$  においてSTEP2に進む。
- ・  $f(x_1) > f(x_0)$  ならば、 $h = -h, x_{-1} = x_1, x_1 = x_0 + h$  において  $f(x_1)$  を計算する。 $k = 0$  において、STEP3へ進む。

## STEP2

$h = 2h, x_{k+1} = x_k + h$  において  $f(x_{k+1})$  を計算する。

## STEP3

以下の場合分けに従う。

- ・  $f(x_{k+1}) \leq f(x_k)$  ならば  $k = k+1$  においてSTEP2に進む。
- ・  $f(x_{k+1}) > f(x_k)$  ならば  $(x_{k-1}, x_k, x_{k+1})$  を小さい順に並べて  $(x_1, x_2, x_3)$  として終了する。

### [3] 黄金分割法のアルゴリズム

黄金分割法のアルゴリズムを以下に示す。

#### STEP0

最初の区間  $[a^0, b^0]$  を囲い込みによって求める。  $\varepsilon > 0$ ,  $k = 0$  とする。

#### STEP1

$x_1^0 = a^0 + F_1(b^0 - a^0)$ ,  $x_2^0 = a^0 + F_2(b^0 - a^0)$  を計算する。

ただし,  $F_1 = \frac{3-5^{1/2}}{2}$ ,  $F_2 = \frac{5^{1/2}-1}{2}$  とする。

#### STEP2

$|a^k - b^k| \leq \varepsilon$  が成り立つならば終了する。

#### STEP3

$f(x_1^k)$  と  $f(x_2^k)$  の比較により以下のように場合分けをする。

- ・  $f(x_1^k) < f(x_2^k)$  の場合  
 $a^{k+1} = a^k$ ,  $b^{k+1} = x_2^k$ ,  $x_2^{k+1} = x_1^k$ ,  $x_1^{k+1} = a^{k+1} + F_1(b^{k+1} - a^{k+1})$  とおいて,  $f(x_1^{k+1})$  を計算し,  $k = k + 1$  として STEP2 に進む。
- ・  $f(x_1^k) \geq f(x_2^k)$  の場合  
 $a^{k+1} = x_1^k$ ,  $b^{k+1} = b^k$ ,  $x_1^{k+1} = x_2^k$ ,  $x_2^{k+1} = a^{k+1} + F_2(b^{k+1} - a^{k+1})$  とおいて,  $f(x_2^{k+1})$  を計算し,  $k = k + 1$  として STEP2 に進む。

## 4 シミュレーション結果

ノイズ除去特性を評価するシミュレーションには, AM 波, FSK 波, 矩形波の 3 種類の波形を用いる。図 3 にこれらの学習原波形を示す。学習の処理には, これらにノイズをランダムに  $\pm 10\%$  付加して用いる。これらのノイズ付加学習波形を図 4 に示す。学習は誤差が所定の許容値 (0.006 とした) を達成するまで行い, 5000 回で打ち切ることとする。学習終了後のニューラルネットワークを用いた認識処理では, 学習に用いたのとは異なる波形を用いた。AM 波は学習原波形の振幅の大きさを反転させた波形を, FSK 波と矩形波は学習原波形の左右を反転した波形を用いた。認識原波形を図 5 に示し, 認識源波形に  $\pm 10\%$  のノイズをランダムに付加した認識処理用のノイズ付加認識波形を図 6 に示す。シミュレーションは以下のパラメータ設定で行った。

- ・ 最急降下法

シグモイド関数の傾き ( $u_0$ ): 0.75

FIR シナプスの結合荷重修正用学習定数 ( $\alpha$ ): 0.4

オフセットの結合荷重修正用学習定数 ( $\gamma$ ): 0.11

出力フィードバックの結合荷重修正用学習定数 ( $\beta$ ): 0.4

- ・ 共役勾配法

シグモイド関数の傾き ( $u_0$ ): 0.75

ニューラルネットワークの構成は, 時系列信号をシリアルに入出力するので, 入力層と出力層のユニット数は 1 個とする。中間層のユニット数は 5, 10, 15 個の 3 通り, 出力フィードバックのディレイ段数と FIR シナプスのディレイ段数は, それぞれ中間層と出力層の組み合わせで, 5-5, 5-10, 5-15, 10-5, 10-10, 10-15, 15-5, 15-10, 15-15 の 9 通りについて, 全ての組み合わせを学習させた。誤差が 0.006 に達したときの学習回数を, AM 波については図 7, FSK 波については図 8, 矩形波については図 9 にそれぞれ示す。これらの図において, 横の欄は FIR シナプスのディレイ段数 (中間層-出力層) を, 縦の欄は, 出力フィードバックのディレイ段数 (中間層-出

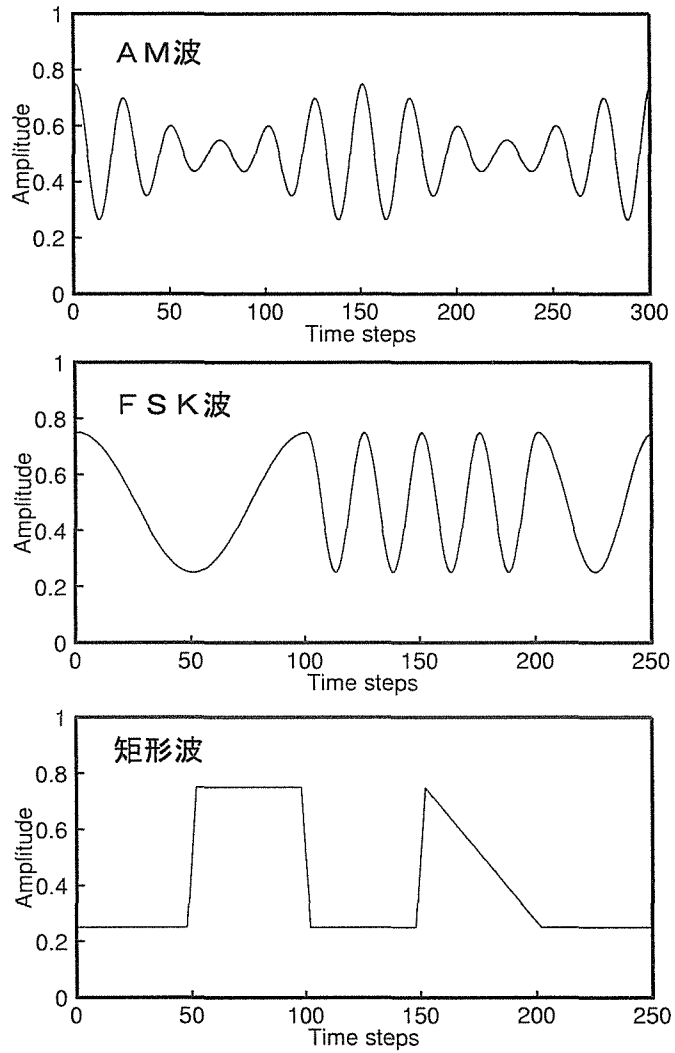


图3 学习原波形

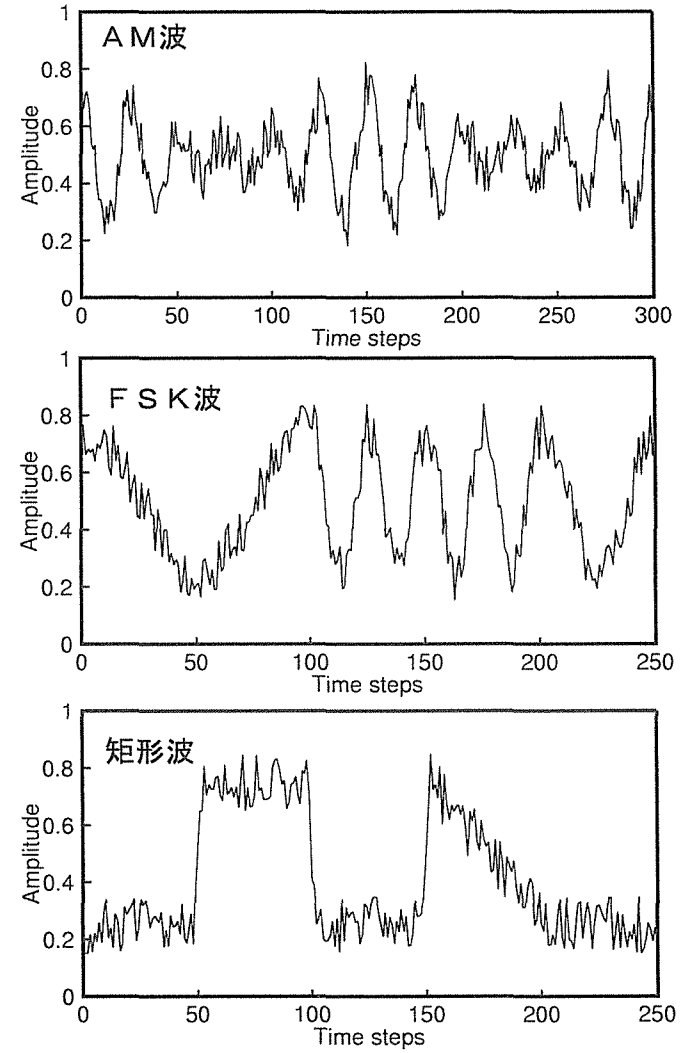


图4 ノイズ付加学習波形

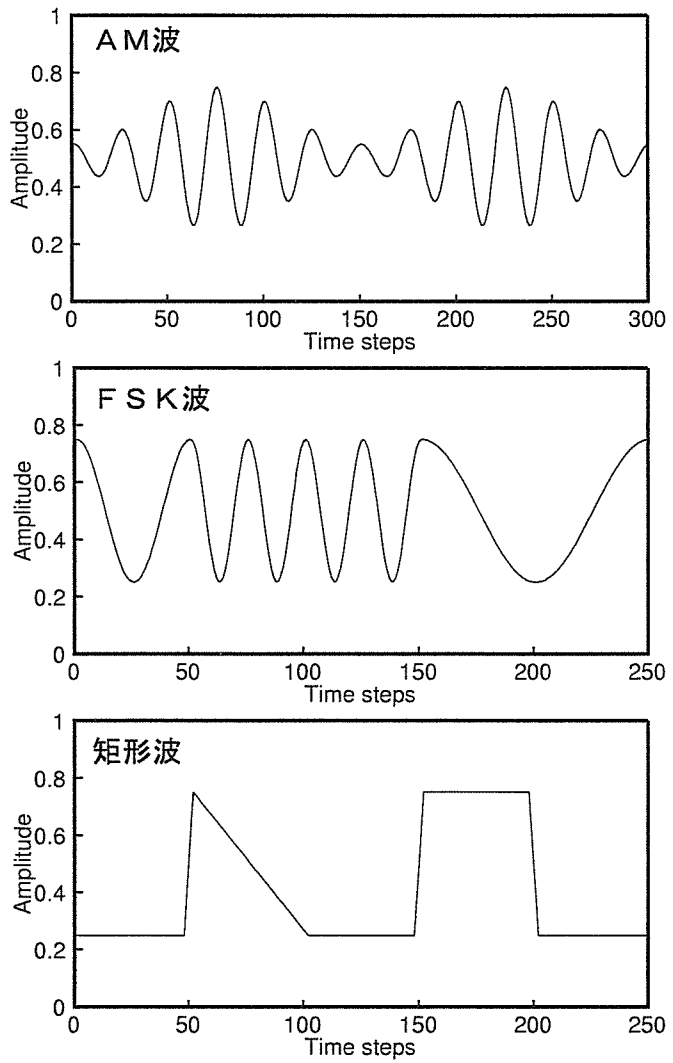


図5 認識原波形

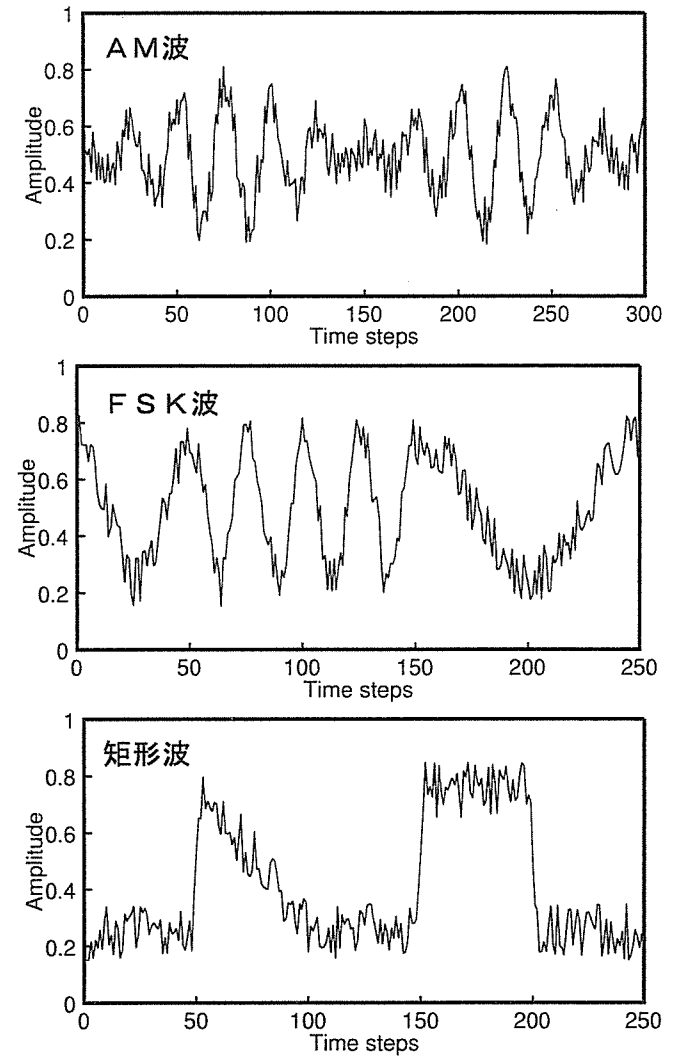


図6 ノイズ付加認識波形

FIR FB	5-5	5-10	5-15	10-5	10-10	10-15	15-5	15-10	15-15
5-5	69	38							2
5-10		156							
5-15	2863 383	494 73	771 194	316 42	3001 612	3914 1007		801 232	85
10-5									
10-10									
10-15	175 57	251 179	3092 1372	167 290		4427 961	626 28	751 104	3053 38
15-5									
15-10							1831	3664	
15-15	120 17	377 26		83	992		407 74	12	

(1) 中間層ユニット数5の場合のAM波学習回数

FIR FB	5-5	5-10	5-15	10-5	10-10	10-15	15-5	15-10	15-15
5-5	38	51							
5-10				2140					
5-15	3171 2190	3036 1958	131	2508 497		396	3260 1644	416 201	2966 668
10-5									
10-10									
10-15	3617 2208	712		401			528 46	394 27	1391 338
15-5						4461 381	2270 274	3849 729	3749 1201
15-10					1541 527		3109 491	193	2254
15-15	3809	1133 890		1149	436 329	216	427 105	1666 318	876 11

(2) 中間層ユニット数10の場合のAM波学習回数

FIR FB	5-5	5-10	5-15	10-5	10-10	10-15	15-5	15-10	15-15
5-5	31	25	97	4077		3844	1701	1772 691	3894 2188
5-10		2780 4002					209	747 79	752 81
5-15	2424 919	818 128	284	1057 1201	572		462 174	253 58	2192 911
10-5			3829 2186		2240 1563	3902	1375 319	1965 832	2420 1007
10-10	4153 2891						1446 409	723 101	720 96
10-15	53	678	366	295	619	907	386 63	300 61	1134 372
15-5	4102	2346			526 3009	3443	1512 197	1848 757	1551 339
15-10	3313	916 171	4056	3376			999 62	729 30	2847 509
15-15	3393 1609	1462 890	7	1404 431	1198	622 86	588 27	241 152	1589 442

(3) 中間層ユニット数15の場合のAM波学習回数

図7 AM波の学習回数



FIR FB	5-5	5-10	5-15	10-5	10-10	10-15	15-5	15-10	15-15
5-5	449	83					3009		
5-10	564								
5-15					729	685			
10-5									
10-10									
10-15		939			796	109		251	127
15-5						1861	2705		
15-10								3107	108
15-15			2803		893	276		463	72

(1) 中間層ユニット数5の場合のFSK波学習回数

FIR FB	5-5	5-10	5-15	10-5	10-10	10-15	15-5	15-10	15-15
5-5	78	97	14			923		4434 1901	
5-10	52						3973 3839	4235 2379	2219 1803
5-15	693	3216	4824 982			1788 1093	2640	3327 1706	2238 262
10-5		3209		3829			2974	1829 891	1187 723
10-10							2193	3643 2028	1829 828
10-15			3729 938		2561 1210	2424 901	1383 571	94	932 601
15-5	3619			1325	2057 829		3614 2761	3229 1003	1248 739
15-10					419		2707 938	2938 622	1878 82
15-15	2153 918			1769 912	2307 1786	1066 517	1465 117	1358 378	909

(2) 中間層ユニット数10の場合のFSK波学習回数

FIR FB	5-5	5-10	5-15	10-5	10-10	10-15	15-5	15-10	15-15
5-5	154					3195 2763	3201 881	1453 548	717 88
5-10		2009	2251		4960 2095	1746	1870 497	1975 821	424 93
5-15	2928	2382 1949	1999 3850	2015 873	1048 285	2044 913	2629 182	592 296	708 176
10-5	3987	2961		2853	3025 2103	2810 1729	1614 528	1557 612	695 298
10-10					3629 2951	3175 1412	1068 398	1160 696	517 39
10-15	2812 1429	1874 834	1336 187	2218 192	872 210	1263 929	1667 823	562 211	708 332
15-5	3896		2550 1169	2370 907	3248 1291	4757 912	1830 293	748 364	1328 1218
15-10		4637 825			1253 899	3607 1218	1439 631	473 283	495 102
15-15		1107 251	1549 374	2876 1989	646 92	925 89	1123 937	568 192	957 427

(3) 中間層ユニット数15の場合のFSK波学習回数

図8 FSK波の学習回数

FIR FB	5-5	5-10	5-15	10-5	10-10	10-15	15-5	15-10	15-15
5-5		38		982					
5-10	821								
5-15			659	2195	3075	3752		631	728
10-5									
10-10									
10-15				2017	1741		929	198	836
15-5				2964			2482		
15-10							1249	772	217
15-15				1332	627	725		52	396

(1) 中間層ユニット数5の場合の矩形波学習回数

FIR FB	5-5	5-10	5-15	10-5	10-10	10-15	15-5	15-10	15-15
5-5	254	512			3487 2008	4529 3189	1285		2455 1090
5-10						4297 2668			4106 1434
5-15		4231	2541		2918 1219	2408 2109		1811	2873
10-5	4108	2859	4593 2101	3291	2844 1289	2273 1755	4527 1211		2479 924
10-10					2270 1346	3813 989	4821 1019	2518	2998 1623
10-15			381		628	2432 481	3045 1008	3698 1258	2284 1091
15-5		3817 1138	3021 1224	4781 2196		2327 926	2620 1482	1548 492	1939 982
15-10		2854	3003 1439	4741 2754	2512 1663	1825 419	2452 257	1900 911	1874 1232
15-15	2178	2378 1092	2685 1311	2415 1434	2838 1725	1568 1281	2444 1721	1655 653	4067 38

(2) 中間層ユニット数10の場合の矩形波学習回数

FIR FB	5-5	5-10	5-15	10-5	10-10	10-15	15-5	15-10	15-15
5-5		3134 2718	763 274	2711	1913 18	614 217		1536 635	569 102
5-10	31	2543 1905	1848 819		1325 552	1219 538	4986 1678	1080 173	4130 2369
5-15		3239 3142	3126 1676		1402 689	1941 1097	2911	1648 17	828 298
10-5	4290 2861	1777 1652	1007 537	1464 928	2089 1802	641 293	1845 962	1005 569	745 209
10-10	1893	1940 871	1138 496	1389 689	1876 737	900 281	1939 679	758 297	744 262
10-15	2996 2108	2031 1194	1700 700	2803 1002	1304 439	1816 83	1573 691	1165 598	901 652
15-5	1342 720	1661 797	630 83	876 208	1001 289	780 91	915 261	669 172	869 391
15-10	1110 411	1638 769	1082 727	910 503	1308 392	797 172	796 189	804 305	613 87
15-15	841 340	950 389	1799 908	607 247	783 302	1115 619	906 216	845 492	743 281

(3) 中間層ユニット数15の場合の矩形波学習回数

図9 矩形波の学習回数

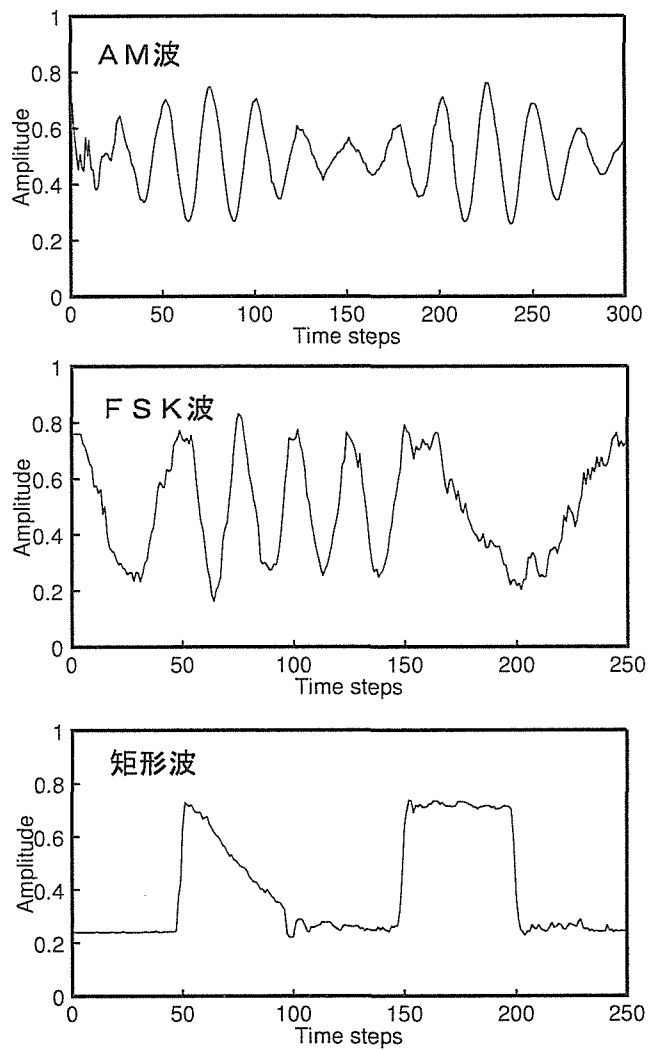


図 10 最急降下法での認識出力波形

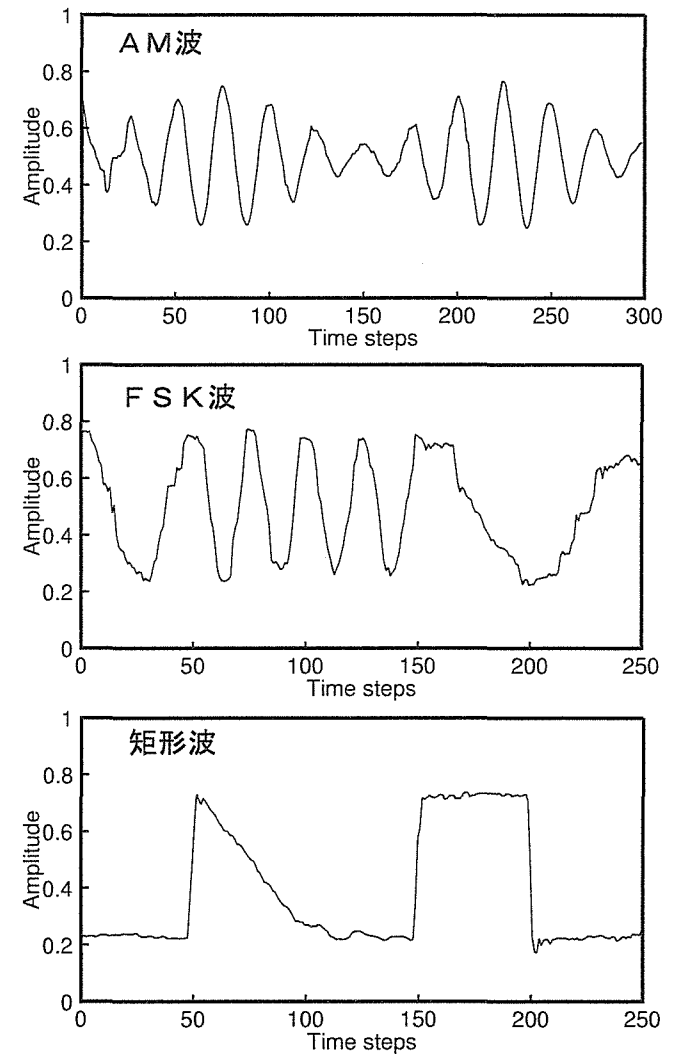


図 11 共役勾配法での認識出力波形

力層)を表している。図中の数値は上段が最急降下法の場合の学習回数、上段が共役勾配法の場合の学習回数を示し、空欄は学習回数が5000回でも許容誤差が達成されず収束しなかったことを示す。

図7, 図8, 図9から, 中間層のユニット数は5個よりは10個あるいは15個と多い方が収束し易い傾向のあることがわかる。出力フィードバックのディレー段数は特に出力層で15段の場合に収束し易いように思われる。FIR シナプスのディレー段数は特に中間層で15段の場合に収束し易いように思われる。また, 組み合わせのほとんどの場合で, 最急降下法の場合よりも共役勾配法の場合の方が収束することが多く, 学習回数は少ない(1/2 ~ 1/100) ことがわかる。

出力フィードバックにディレー素子をもつFIRニューラルネットワークを最急降下法と共役勾配法で学習させた場合のそれぞれの認識出力波形を図10と図11に示す。両者を比較すると, 最急降下法で学習させた場合よりも共役勾配法で学習させた場合の方が認識原波形に近く, 最急降下法よりも共役勾配法の方が優れているといえる。

## 5 むすび

出力フィードバックにディレー素子を含むFIRニューラルネットワークにおいて学習に共役勾配法を用いることの効果を検討した。AM波とFSK波および矩形波におけるノイズ除去問題を取りあげて, ノイズを付加した波形の認識波形と学習過程を通常の最急降下法の場合と比較検討した。共役勾配法は, 最急降下法の処理に比べ, 学習アルゴリズムが複雑であるが, 学習回数が少なく済み, 学習が収束する確率が高い。また, 波形の認識結果も良好であることが確認された。

以上の結果から, 出力フィードバックにディレー素子を含むFIRニューラルネットワークにおいて共役勾配法は最急降下法に比べ, 学習アルゴリズムが複雑であるが, 学習効果の高い学習方法であることがわかった。このことは, ニューラルネットワークを情報教育に応用していくうえで有益な知見であり, 大いに役立てていきたい。また, 今後は入力層にもディレー素子を設けた場合を検討していきたい。

## 参考文献

- 1) M.Nishi, J.Furuya, and T.Nakamura, "Backpropagation Networks Including Time Delay elements(BPD)," Proc. of IJCNN'93, IEEE, pp.1681-1684, 1993.
- 2) 西正明, 降矢順治, 林川基治, "ディレー素子を含むニューラルネットワークによるノイズ除去特性," 山口大学教育学部附属教育実践総合センター研究紀要, 第12号, pp.195-208, 2001.
- 3) 今野浩, 山下浩, "非線形計画法," 日科技連, 1978.

(2001年9月25日 受理)