

階層型ニューラルネットワークにおける 学習法の一検討

西 正明* · 降矢順治** · 林川基治***

An Examination of Learning Methods for Multi-Layer Neural Networks

Masaaki NISHI* , Junji Furuya** , and Motoharu HAYASHIKAWA***

(Received November 8, 2001)

1. まえがき

生物の神経系の仕組みを工学的に実現しようとするのがニューラルネットワークである。ニューラルネットワークを構成することによって、従来からある構造のコンピューターでは苦手としていたアナログ的な処理を実現させ、究極的には人の脳と同じような処理ができるようにしようと考えられている。人の大脳の場合は100億を越える神経細胞のひとつひとつが互いに1万もの結合をしている。ニューラルネットワークでは神経細胞自体を簡単にユニット化し、数十ないし数百個程度のユニットを用いてネットワークにモデル化して用いている¹⁾。ネットワークにはユニット間で全てが結合する相互結合型とユニットを層状に配置して層間でのみ結合させる階層型の2種類に大別される。また、ニューラルネットワークでは、所望の処理をするように学習データを与えて学習させる必要がある。学習することによってユニット間の結合の強さが適切に調整されていく。最もよく用いられている学習アルゴリズムに誤差逆伝播 (Error Back Propagation : BP) 法がある²⁾。

本論文では階層型ニューラルネットワークをパターン認識に適用し、学習方法としてBP法を基本に用いた場合について扱う。ユニット間の結合の強さの調整の仕方について、各学習パタンのひとつひとつの学習のたびに結合荷重を修正する逐次修正法と幾つかの学習パタンの学習についてまとめて結合荷重を修正する一括修正法、および過去の修正量を一部利用するモーメント法をアルファベット文字の認識に適用して検討したので報告する。

2. ネットワーク構成と学習アルゴリズム

本論文では、階層型ニューラルネットワークを扱う。3層のネットワーク構成を図1に示す。BP法の学習アルゴリズムを以下に簡単に述べる。図1に示した3層の階層型ニューラ

* 信州大学教育学部

** 城西国際大学経営情報学部

*** 山口大学教育学部

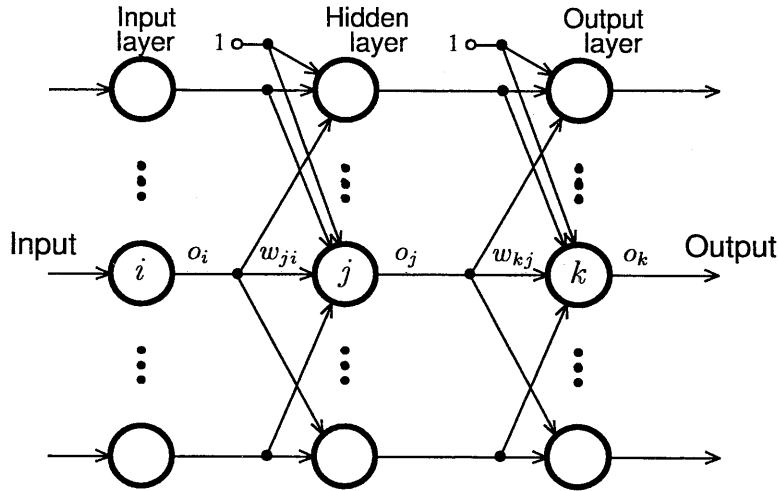


図1 階層型ニューラルネットワークの構成

ルネットワークで出力層の k 番目ニューロンの内部状態 s_k と出力 o_k は次式で示される。ここで、 $f[\]$ は応答関数であり、 u_0 は応答関数の傾きを規定するパラメータである。

$$s_k = \sum_j w_{kj} o_j \quad (1)$$

$$o_k = f[s_k] = \frac{1}{2} \{1 + \tanh(\frac{s_k}{u_0})\} \quad (2)$$

結合荷重の修正量は教師信号を d_k として、平均自乗誤差を次式 (3) とし、学習係数を α とすれば、次式 (4) で計算することができる。全てのパターンにおける誤差の和 E は次式 (5) で示される。最終的にはこの誤差の和 E を最小にするように結合荷重値を修正する。

$$E_p = \frac{1}{2} \sum_k (d_k - o_k)^2 \quad (3)$$

$$\Delta w_{kj} = -\alpha \frac{\partial E_p}{\partial w_{kj}} \quad (4)$$

$$E = \sum_p E_p \quad (5)$$

本論文では、学習パターンをひとつひとつ学習するたびに式 (4) にしたがって結合荷重を修正していく逐次修正法と、ひとまとまりの学習パターンの学習後にそれまでの修正量の蓄積を一括して結合荷重に加えて修正する一括修正法を検討する。一括修正法では結合荷重を修正する回数を減らせるため、学習処理時間の短縮が期待できる。また、一括修正法については、前回の修正量の一部を反映させることによって、学習の向上をはかろうとするモーメント法を併用する場合も検討する。モーメント法での結合荷重の修正量は m をモーメント係数、 $\Delta w_{kj}^{\text{old}}$ を前回の結合荷重修正量とすれば、次式 (6) で示される。なお、学習回数は学習

パターンを一巡して学習するのを1回と数えることにする。

$$\Delta w_{kj} = -\alpha \frac{\partial E_p}{\partial w_{kj}} + m \Delta w_{kj}^{\text{old}} \quad (6)$$

3. シミュレーション

図1に示した3層構造のニューラルネットワークを用いる。各層のユニット数が入力層で I 、中間層で J 、出力層で K の場合、ネットワーク構成を I - J - K で示すことにする。

学習における結合荷重修正方法として、逐次修正法と一括修正法、および一括修正法にモーメント法を併用する一括修正法+モーメント法の3種類について比較する。パターン学習・認識には“ A ”～“ Z ”のアルファベット26文字を、 9×7 と 7×5 の2種類の格子上に“1”と“0”の値で描いたパターンを用いる。学習時にはこのパターンをそのまま用いる。認識時にはランダムな位置で格子1個分の値を“1”と“0”を入れ替えたノイズ入りパターンを1文字あたり50個作成して用いる。

シミュレーションでは学習・認識文字を“ A ”～“ J ”の10種または“ A ”～“ Z ”の26種とする。出力層のユニット数は扱う文字種に対応させて10個または26個とする。中間層のユニット数は10個と20個を試みる。応答関数の傾きは $u_0 = 0.2, 0.4, 0.8, 1.2, 1.6, 3.0$ 、学習係数は $\alpha = 0.1, 0.5, 1.0$ をそれぞれ試みる。モーメント係数は $m = 0.01 \sim 4.0$ を試みる。以下にシミュレーション結果を述べる。

表1 9×7 格子パターンの学習終了時 ($E < 1.0$) の学習回数と認識率

結合荷重修正法	学習係数 α	モーメント係数 m	学習回数	認識率 (%)
逐次修正法	1.0	0.0	875	97.8
	0.5	0.0	1816	86.4
	0.1	0.0	3451	77.2
一括修正法	1.0	0.0	513	96.0
	0.5	0.0	1563	87.4
	0.1	0.0	2956	82.3
一括修正法+ モーメント法	1.0	0.1	244	95.2
		0.2	443	93.2
		0.6	401	96.4
		1.0	161	98.2
	0.5	0.1	1205	91.6
		0.2	805	93.2
		0.6	506	90.6
		1.0	380	92.0
	0.1	0.1	4562	87.2
		0.2	3621	88.6
		0.6	1968	85.6
		1.0	2312	85.8

3.1 学習終了時の学習回数と認識率

逐次修正法，一括修正法，一括修正法+モーメント法の各結合荷重修正法で，誤差が $E < 1.0$ になった時点で学習を終了したときの，学習回数と認識率を比較した。対象文字は“A”～“J”の10種，応答関数の傾きは $u_0 = 0.2$ とした。 9×7 格子のパタンで 63-10-10 の構成のニューラルネットワークで学習させた場合を表1に， 7×5 格子のパタンで 35-10-10 の構成のニューラルネットワークで学習させた場合を表2にそれぞれ示す。

表2 7×5 格子パタンの学習終了時 ($E < 1.0$) の学習回数と認識率

結合荷重修正法	学習係数 α	モーメント係数 m	学習回数	認識率 (%)
逐次修正法	1.0	0.0	538	83.2
	0.5	0.0	1359	88.4
	0.1	0.0	2351	74.4
一括修正法	1.0	0.0	575	87.8
	0.5	0.0	1612	85.8
	0.1	0.0	3211	72.6
一括修正法+ モーメント法	1.0	0.1	384	90.6
		0.2	364	90.6
		0.6	108	94.8
		1.0	113	86.8
	0.5	0.1	1525	85.4
		0.2	920	90.4
		0.6	491	88.8
		1.0	379	87.4
	0.1	0.1	3250	78.9
		0.2	3579	78.1
		0.6	4061	82.6
		1.0	2261	78.0

表1と表2より，学習回数は α の値が大きいと少なくなっている。同じ α の値で逐次修正法と一括修正法を比べると， 9×7 格子パタンでは一括修正法， 7×5 格子パタンでは逐次修正法でそれぞれ学習回数が幾分少なくなっている。したがって，この場合，学習回数における逐次修正法と一括修正法による有意差はほとんど認められない。モーメント法を併用すると $\alpha = 0.5, 1.0$ の場合には学習回数が少なくなる傾向があることがわかる。しかし， $\alpha = 0.1$ の場合のように α の値が小さいとモーメント法の効果は小さくなることがわかる。一方，認識率は $\alpha = 0.5, 1.0$ の場合に高く，逐次修正法と一括修正法による有意差はほとんど認められない。一括修正法にモーメント法を併用すると，モーメント係数が $m = 0.1 \sim 1.0$ のほぼ全域で認識率が向上する傾向のあることがわかる。

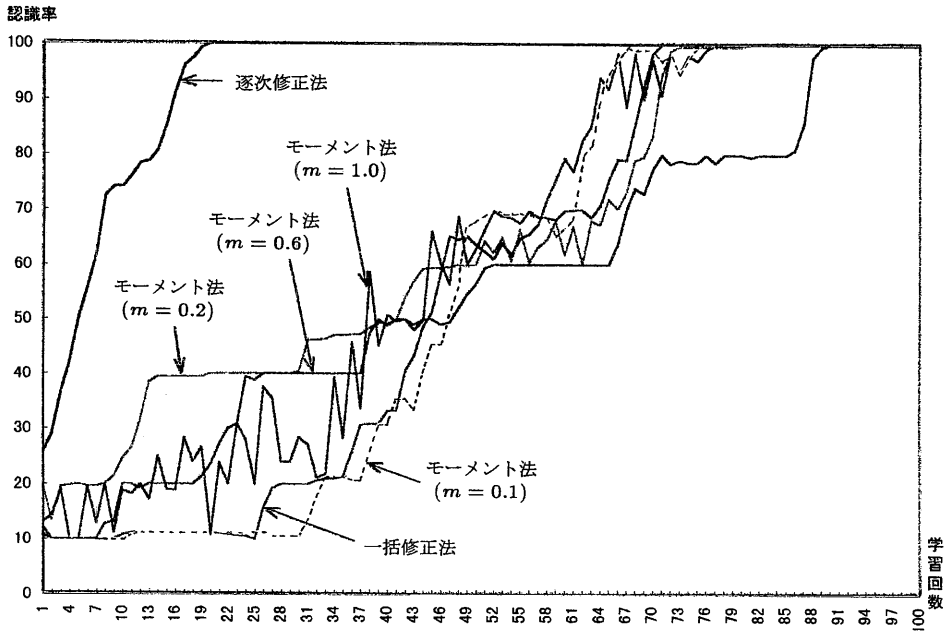


図2 結合荷重修正法による認識率の推移 (35-10-10)

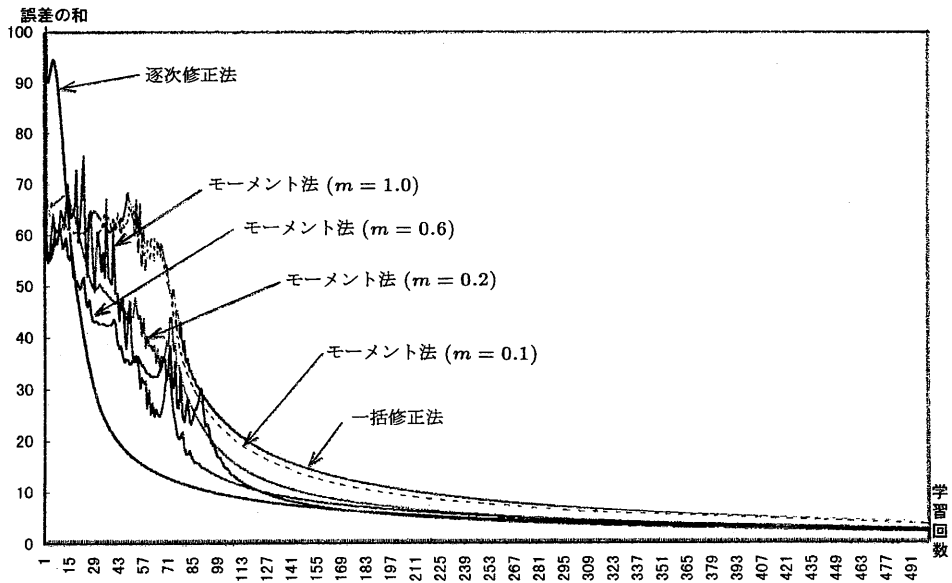


図3 結合荷重修正法による誤差の推移 (35-10-10)

3.2 結合荷重修正法による認識率と誤差の推移

学習・認識の対象文字は“A”～“J”の10種で7×5格子のパタンとし、35-10-10の構成のニューラルネットワークを用いる。学習係数は $\alpha = 1.0$ ，応答関数の傾きは $u_0 = 0.8$ とする。逐次修正法，一括修正法，一括修正法+モーメント法の各結合荷重修正法における認識率と誤差の推移を図2と図3にそれぞれ示す。図2から，逐次修正法が最も少ない学習回数の段階で高い認識率を達成し，一括修正法は多くの学習回数が必要になっていることがわかる。モーメント法を併用すると，逐次修正法には至らないが改善されることがわかる。誤差の推移についても同様のことが言え，図3から，逐次修正法が最も少ない学習回数で誤差が少なくなっており，一括修正法は多くの学習回数を必要とすることがわかる。モーメント法を併用すると，学習初期の段階で誤差値に揺れが生じるが，逐次修正法にかなり近づき改善されることがわかる。

3.3 学習条件による認識率と学習回数

逐次修正法，一括修正法，一括修正法+モーメント法の各結合荷重修正法において，さまざまな応答関数の傾き値とモーメント係数の値での認識率と学習回数を調べた。認識率の結果を9×7格子の26文字の場合を図4，9×7格子の10文字の場合を図5と図6，7×5格子の10文字の場合を図7にそれぞれ示す。いずれの場合も，学習係数は $\alpha = 1.0$ ，学習終了条件は $E < 0.2$ で，学習打ち切り回数は5000回とする。使用するニューラルネットワークの構成は，図4では9×7格子の26文字に合わせて入力層ユニット数63個，出力層ユニット数26個にして，63-10-26とする。同様に，図5では9×7格子の10文字に合わせて63-10-10とし，図6では中間層のユニット数を20個に増やしてニューラルネットワーク構成を63-20-10として評価する。図7では7×5格子の10文字に合わせて，入力層ユニット数35個，出力層ユニット数10個にして，35-10-10の構成のニューラルネットワークとする。この場合の学習回数を図8に示す。

図4，図5，図6，図7より，さまざまな応答関数の傾きと学習法やモーメント係数の値の場合の認識率に関して，まず u_0 の値が大きい，即ち応答関数の傾きが小さい方が認識率は高くなることがわかる。また，認識文字が10種の場合には，図5と図6と図7より，応答関数の傾きが小さい領域ではノイズを含むパタンを全て認識していることがわかる。これに対して，認識文字が26種の場合には図4に示すように，応答関数の傾きが大きくても，認識率は60%程度であった。これは中間層のユニット数が少な過ぎて処理能力が十分でないためと思われる。この場合，モーメント法を併用しても認識率はほとんど改善されず，モーメント係数を0.5以上の値にするとむしろ認識率が下がる傾向があった。

図5と図6からは，中間層ユニット数が10個の場合よりは，20個に増やした方が，認識率が高まる傾向のあることが認められる。

また，図7と図8から， u_0 の値が大きい，即ち応答関数の傾きが小さいと，学習回数は増えるが認識率は高まることがわかる。一括修正法にモーメント法を併用すると，同程度の認識率でありながら，逐次修正法および一括修正法よりも学習回数が少なくて済む傾向のあることがわかる。

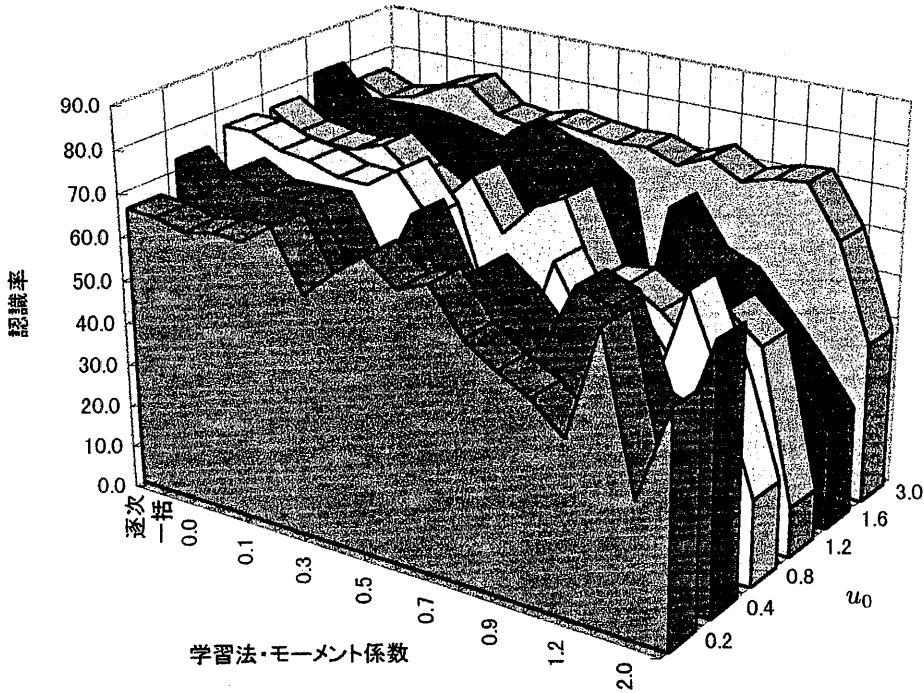


図4 9×7格子の26文字の認識率(63-10-26)

4. むすび

階層型ニューラルネットワークにおいて、最急降下法に基づいた誤差逆伝播法で学習を進めていくうえで、結合荷重値を修正する方法として、逐次修正法と一括修正法、および一括修正法にモーメント法を併用する方法を比較検討した。学習データとしてアルファベット文字10種または26種を9×7格子または7×5格子上に“1”と“0”の値でパターンを描いて用いた。認識データは、学習データにランダムな位置の1格子分の値を“1”と“0”を入れ替えてノイズを含んだ文字とし、これを1文字当たり50個を作成して用いた。学習係数と応答関数の傾きとモーメント係数からさまざまな条件を設定して、認識率と学習回数を調べた。

その結果、応答関数の傾きが小さくなると学習回数は増えるが認識率は高まること、逐次修正法と一括修正法による有意差はほとんど認められないこと、一括修正法にモーメント法を併用するとモーメント法を用いない一括修正法および逐次修正法よりも学習回数が少なくて済むことがわかった。一括修正法にモーメント法を併用する場合は、最も学習回数が少なく、かつ最も認識率が高かった。このことは、今後の情報教育に応用していくうえで有益な知見である。

今回は一括修正法にモーメント法を併用する場合を検討したが、今後は逐次修正法にモーメント法を併用する場合も含めて、総合的に評価検討していかなければならない。

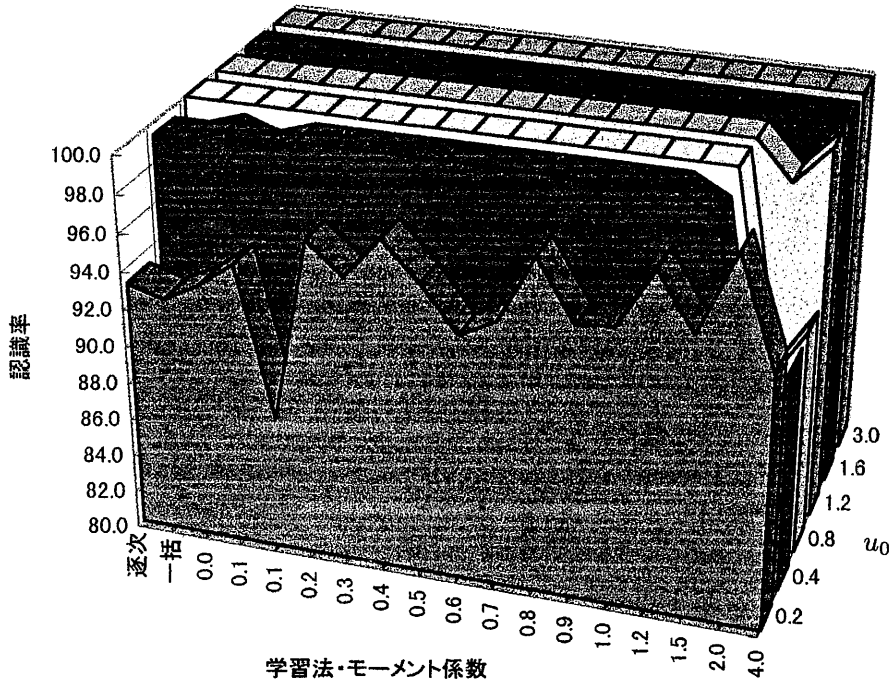


図5 9×7格子の10文字の認識率 (63-10-10)

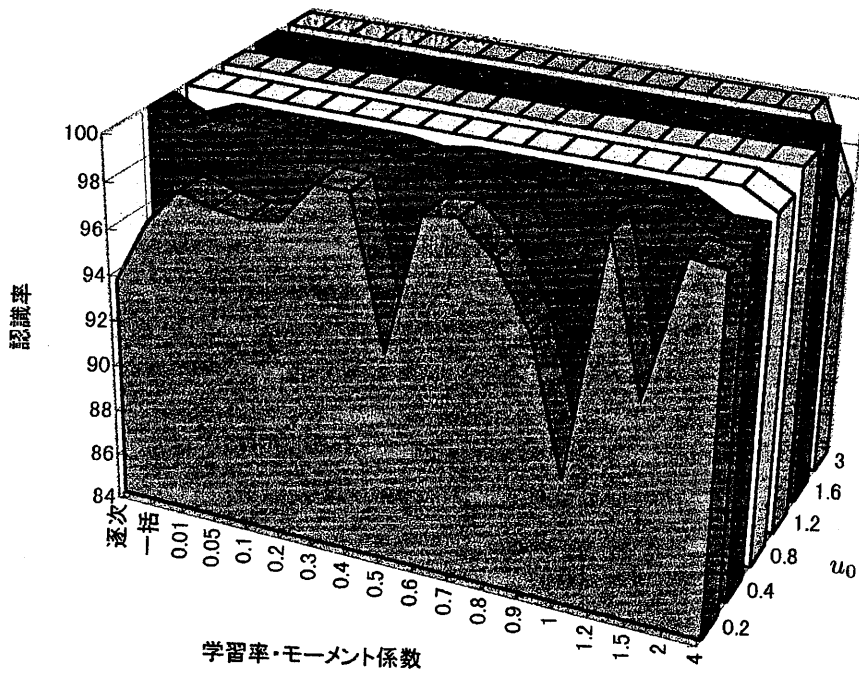


図6 9×7格子の10文字の認識率 (63-20-10)

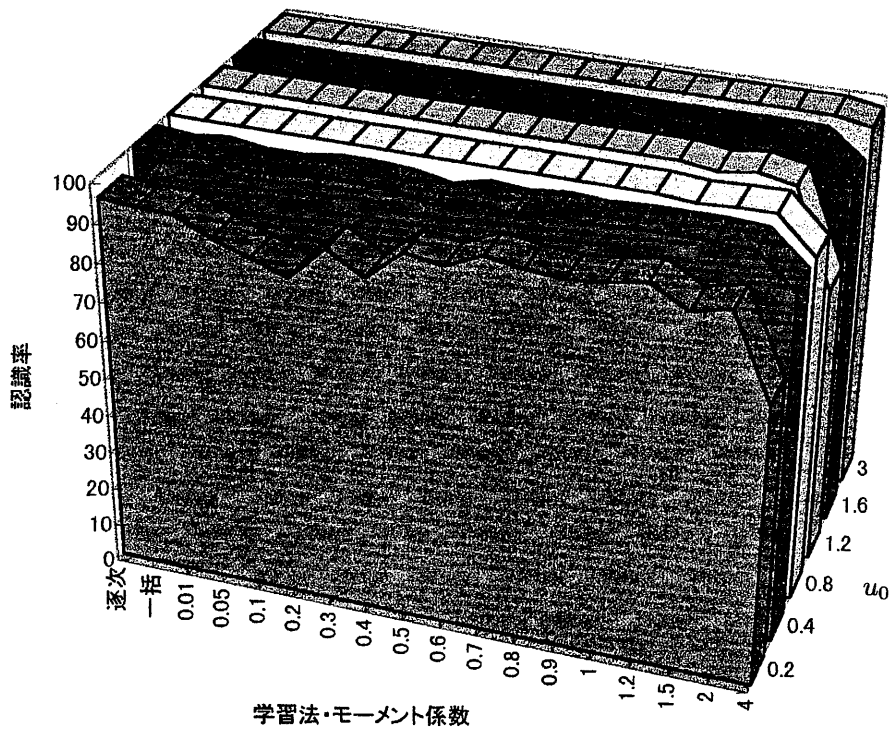


図7 7×5格子の10文字の認識率(35-10-10)

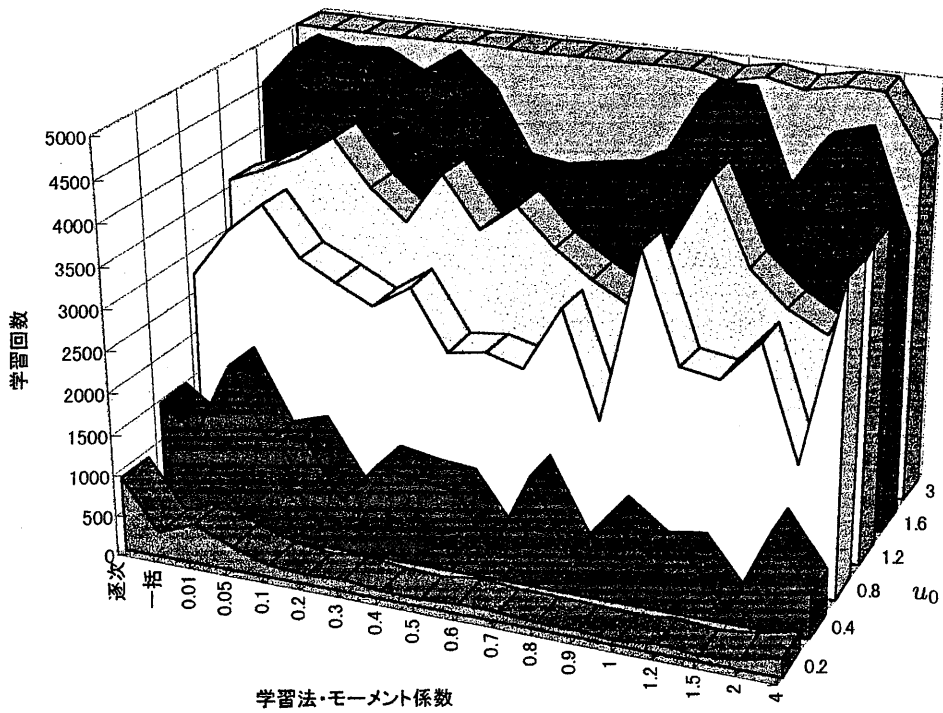


図8 7×5格子の10文字の学習回数(35-10-10)

参考文献

- 1) 中野馨, 飯沼一元, “入門と実習ニューロコンピュータ,” 技術評論社, 1989.
- 2) Robert Hecht-Nielsen, “Neurocomputing,” Addison-Wesley Publishing Company, 1990.