

デジタル型BP法における教師信号の制限

蔵淵, 義光 / LI, Lei / ZOBUCHI, Yoshimitsu / 李, 磊

(出版者 / Publisher)

電子情報通信学会

(雑誌名 / Journal or Publication Title)

電子情報通信学会技術研究報告. KBSE, 知能ソフトウェア工学 / 電子情報通信学会技術研究報告. KBSE, 知能ソフトウェア工学

(号 / Number)

383

(開始ページ / Start Page)

21

(終了ページ / End Page)

26

(発行年 / Year)

2006-11-18

デジタル型BP法における教師信号の制限

蔵 淵 義光[†] 李 磊[†]

[†] NTT 東日本, 東京都

[†] 法政大学工学部 〒184-8584 東京都小金井市梶野町3-7-2

E-mail: ††lilei@hosei.ac.jp

あらまし 現在, 誤差逆伝搬法 (BP 法) を用いた階層型ニューラルネットワークの研究が盛んに行われており, その中でも不可微分な 2 値出力素子を含む三層ニューラルネットワークに対する新しい誤差逆伝搬法であるデジタル型誤差逆伝搬法 (DBP 法) が提案されている. 従来の単純パーセプトロン学習が出力にしか教師信号を与えないのに対して, DBP 法では中間層にも出力誤差を減少させるような教師信号を与える. したがって, DBP 法では上層結合係数ばかりではなく下層結合係数の修正も可能となり, 高速化やハードウェア化の可能な 2 値出力素子のみからなるネットワークが容易に得られる. 本論文では, この DBP 法の教師信号の修正に条件式を導入し, 制限を加えた方法を提案する. 提案手法を, XOR 問題, 半加算器問題, 全加算器問題, パリティ問題を対象に実験を行った結果, 従来の方法と比較して学習成功率, 平均学習回数の向上が得られたことを示す.

キーワード デジタル型 BP 法, 教師信号

Restriction of the Teacher Signal in the Digital Type BP Learning Algorithm

Yoshimitsu ZOBUCHI[†] and Lei LI[†]

[†]

[†] Faculty of Engineering, Hosei University Kajino-cho 3-7-2, Koganei, Tokyo 184-8584 Japan

E-mail: ††lilei@hosei.ac.jp

Abstract A study of various neural networks has been performed till now. For the simple perceptron learning algorithm that gives only the output layer a teacher signal, DBP learning algorithm to give the middle class a teacher signal is suggested in that. I suggest the learning algorithm that introduced a condition type into update of a teacher signal of DBP learning algorithm. As a result of having experimented on the learning algorithm that I suggested for the issue of XOR, the issue of half adder, the issue of full adder, the issue of parity, a good result was provided than conventional learning algorithm.

Key words Layered Neural Network, BP Learning Algorithm, Simple Perceptrons Learning Algorithm, Nondifferentiable Unit

1. はじめに

よく知られているように, ニューラルネットワーク (NN) は階層型 (パーセプトロンなど) と相互結合型 (アソシアトロン, ホップフィールドネットワーク, ボルツマンマシンなど) がある. 階層型 NN の特徴として, 入力を与えられる入力層, 外部へ出力する出力層とそれ以外の隠れ層 (中間層) に分けることができる. 出力層を構成するニューロンモデルは望ましい出力を教師信号として与えることで教師あり学習を行うことができるが, 隠れ層を構成するニューロンモデルに対しては望ましい

出力を定義することができないため教師あり学習をすることができない. そこで, 教師あり学習としてアナログ型の誤差逆伝搬法 (ABP 法) が提案されている^[1]. その他にも, 双対学習^[2] やデジタル型 BP 法 (DBP 法)^[3] などが提案されている. 本論文では, DBP 法に対し, 教師信号の更新に条件式を導入し, 制限を加えた方法を提案する. 提案手法を, XOR 問題, 半加算器問題, 全加算器問題, パリティ問題を対象に実験を行った結果, 従来の方法と比較して学習成功率, 平均学習回数の向上が得られたことを示す.

2. 単層及び単純パーセプトロン

単純パーセプトロンは、Rosenblatt により提案されたパターン認識機械の一種であり、入力パターンがどのカテゴリー（同じものと判定すべきパターン集合）に所属するかを判別する働きを持つ。単層パーセプトロンは、入力素子が線形で出力素子が線形しきい値の二層ネットワークであり Fig. 1 のように構成される。

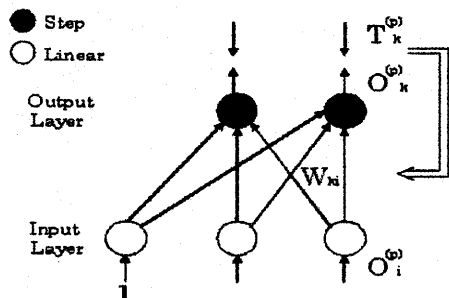


図1 単層パーセプトロン

出力層には識別すべきカテゴリーの数の素子を配し、ある入力パターンが与えられたとき、次式に従い 0 か 1 を出力する。

$$U_k^{(p)} = \sum_i W_{ki} O_i^{(p)}$$

$$O_k^{(p)} = f(U_k^{(p)}) = \begin{cases} 0 & (U_k < 0) \\ 1 & (U_k \geq 0) \end{cases}$$

ここで、 W_{ki} は第 i 入力素子から第 k 出力素子への結合係数を表す。 $O_k^{(p)}$ 、 $U_k^{(p)}$ はそれぞれ、第 p パターン入力時の第 k 出力素子の出力と活性値、 $O_i^{(p)}$ は第 i 入力素子の出力を表す。

単層パーセプトロンの学習では、学習パターン p に対する出力 $O_k^{(p)}$ と教師信号 $T_k^{(p)}$ の誤差をもとに、 W_{ki} を次式より更新する。

$$W_{ki} = W_{ki} - \alpha (O_k^{(p)} - T_k^{(p)}) O_i^{(p)}$$

ここで、 α は学習係数と呼ばれる正の定数である。学習はすべてのパターンについてパターンごとに繰り返し行われ、出力と教師信号の誤差が 0 になるまで続けられる。単層パーセプトロンは、各カテゴリーが他の全てのカテゴリーと線形分離可能であれば、有限回の反復学習で任意のパターンを識別することができる。しかし、入力パターン集合の中で一つでも線形分離不能なカテゴリーが存在すれば、学習不能に陥り入力パターンを識別できなくなるという欠点があった。

単層パーセプトロンのこの問題点は、入力層と出力層の間に 2 値出力素子からなる中間層を加えた三層パーセプトロン（単純パーセプトロンまたは SP 法）により解決することができる（Fig.2）。

SP 法において中間層に十分多くの素子を用いるならば、入力層-中間層間のランダム変換により、任意の線形分離不能なパターン集合が分離可能になることが証明されている [4]。

しかし、膨大な処理を必要とする情報処理方式では、工学的にあまり魅力のないものである。そこで、ある程度現実的な数の中間素子に積極的に学習させ、それぞれの問題に必要な変換ネットワークを中間層に形成することが必要となる。

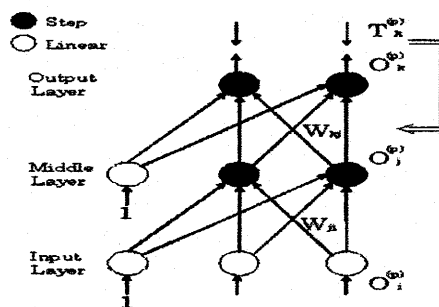


図2 SP 法の学習

3. アナログ型 BP 法

中間層素子を不可微分な 2 値出力素子の代わりに連続で可微分なアナログ素子で構成することにより、出力誤差が上層から下層へ逆伝搬し下層の結合係数の修正が可能となるアナログ型三層パーセプトロンに対する誤差逆伝搬学習（ABP 法）を提案した [1] (Fig3)。

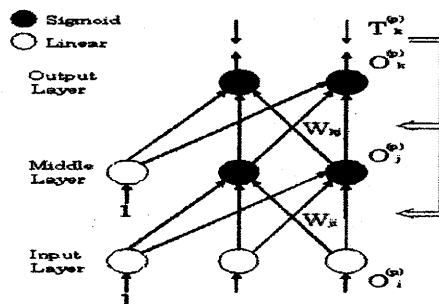


図3 ABP 法の学習

中間、出力層の出力値は、次式で定義される。

$$O_j^{(p)} = f(U_j^{(p)}) = f\left(\sum_i W_{ji} O_i^{(p)}\right)$$

$$O_k^{(p)} = f(U_k^{(p)}) = f\left(\sum_j W_{kj} O_j^{(p)}\right)$$

ここで、 f は中間層、出力層素子の出力関数であり、次式のシグモイド関数を用いる。

$$f(x) = \frac{1}{1 + e^{-x}}$$

誤差関数は、次式で定義される。

$$E = \frac{1}{2} \sum_p \sum_k (O_k^{(p)} - T_k^{(p)})^2$$

上層結合係数 W_{kj} は、誤差関数に対する最急降下則に基づき

次式で更新される。

$$W_{kj} = W_{kj} - \alpha (O_k^{(p)} - T_k^{(p)}) f'(U_k^{(p)}) O_j^{(p)}$$

同様に、下層結合係数 W_{ji} は次式で更新される。

$$W_{ji} = W_{ji} - \alpha \left\{ \sum_k (O_k^{(p)} - T_k^{(p)}) f'(U_k^{(p)}) W_{kj} \right\} f'(U_j^{(p)}) O_i^{(p)}$$

ABP 法は汎用性が高く、その有用性は多くの応用を通して検証されてきたが、最急降下法による誤差関数の最小点探索にかかわるいくつかの問題点が指摘されている。その第一は、線形分離不能課題に対しては、初期値によっては極小解に陥り学習不能となる場合が少なくないことである。最小解に収束する場合でも膨大な学習回数を要することが報告されている。第二の問題点は、処理速度の向上やハードウェア化の容易さからすべての素子出力を 2 値に制限する問題や、オートマトンの内部状態推定やベクトル量子化のように 2 値であることが本質的であるバイナリー型問題に対しては、直接的に解くことができないことである。出力関数としてアナログ型のシグモイド関数を用いた場合でも、誤差関数に 2 値化のためのペナルティ項を付加したり、出力関数の傾斜を急しゅん化するなどの手法が用いられてきたが、いずれも収束に難点があった。

4. デジタル型 BP 法

この方法は、全ての層に不可微分な 2 値出力素子を用いたデジタル型三層パーセプトロンに対する誤差逆伝搬法 (DBP 法) である [3]。デジタル型パーセプトロンでは、中間層が不可微分な素子で構成されているので、上層からの誤差を中間層を通して下層へ逆伝搬することができないという欠点があった。しかし、DBP 法では、出力の誤差を減少させるように、中間層にも教師信号を与えるので、下層の結合係数の修正も可能となる。以下に出力層素子が 1 個と複数の場合に分けて、中間層に与える教師信号の求め方を中心に DBP 法の説明を示す。

4.1 出力層素子が 1 個の場合

Fig.1 に示す 2 値出力素子からなるデジタル型三層パーセプトロンにおいて、まず各入力パターン p に対して、入力層から出力層へ順方向に計算し、出力層出力及び教師信号との誤差を求めらる。

次に、単純パーセプトロンと同様に、上層結合係数を次式のデルタ則に従い修正する。

$$W_{kj} = W_{kj} - \alpha (O_k^{(p)} - T_k^{(p)}) O_j^{(p)}$$

一方、もし中間層に教師信号 $T_j^{(p)}$ を与えることができるならば、下層結合係数は、上層と同様に次式のデルタ則に従い修正することができる。

$$W_{ji} = W_{ji} - \alpha (O_j^{(p)} - T_j^{(p)}) O_i^{(p)}$$

次に、中間層素子の教師信号 $T_j^{(p)}$ の求め方を示す。上層の結合係数 W_{kj} の修正後、 W_{kj} を一時的に固定して考えるならば、

中間層の出力を出力層誤差が減少する値に設定するのが望ましい。すなわち、第 k 出力素子の活性値は、

$$U_k^{(p)} = \sum_j W_{kj} O_j^{(p)}$$

であるので、 W_{kj} が正のときは、 $O_j^{(p)}$ が 0 から 1 に変化すると活性値は増加、1 から 0 に変化すると活性値は減少する。 W_{kj} が負のときは、 $O_j^{(p)}$ の変化とは逆に活性値が増減する。したがって、出力 $O_k^{(p)} = 0$ で教師 $T_k^{(p)} = 1$ のときは ($O_k^{(p)} - T_k^{(p)} = -1$)、出力素子をより活性化するために、中間層の教師信号 $T_j^{(p)}$ を

$$\text{if } W_{kj} > 0 \text{ then } T_j^{(p)} = 1$$

$$\text{if } W_{kj} < 0 \text{ then } T_j^{(p)} = 0$$

とする。一方、出力 $O_k^{(p)} = 1$ で教師 $T_k^{(p)} = 0$ のときは ($O_k^{(p)} - T_k^{(p)} = 1$)、出力素子を非活性にするために、 $T_j^{(p)}$ を

$$\text{if } W_{kj} > 0 \text{ then } T_j^{(p)} = 0$$

$$\text{if } W_{kj} < 0 \text{ then } T_j^{(p)} = 1$$

とする。以上をまとめると

$$\text{if } (O_k^{(p)} - T_k^{(p)}) W_{kj} > 0 \text{ then } T_j^{(p)} = 0$$

$$\text{if } (O_k^{(p)} - T_k^{(p)}) W_{kj} < 0 \text{ then } T_j^{(p)} = 1$$

となる。また、 $O_k^{(p)} = T_k^{(p)}$ のときは上層、下層とも修正する必要はなく、 $W_{kj} = 0$ のときは修正の効果はない。したがって、下層結合係数が修正されないように、形式的に教師信号を現在の中間層出力値に設定する。すなわち、

$$\text{if } (O_k^{(p)} - T_k^{(p)}) W_{kj} = 0 \text{ then } T_j^{(p)} = O_j^{(p)}$$

とする。

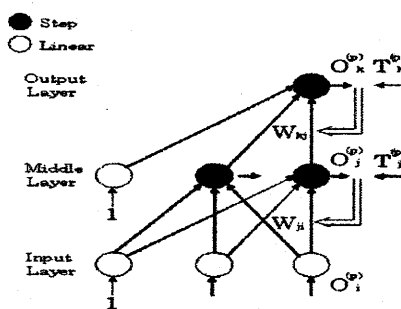


図 4 DBP 法の学習 (出力層素子が 1 個の場合)

以上のことから、出力層素子が 1 個の場合に中間層素子に与える教師信号は Table.1 のようにまとめることができる。ここで、 $S_j^{(p)} = S_{jk}^{(p)} = (O_k^{(p)} - T_k^{(p)}) W_{kj}$ は、中間層素子の教師信号を求めるための根拠となる値であるので、教師決断因子と呼ばれている。

4.2 出力素子数が複数個あるとき

出力層素子が複数個ある場合も (Fig.5), 上層結合係数は 4.1 及び SP 法と同様に, 次式のデルタ則に従い修正する.

$$W_{kj} = W_{kj} - \alpha (O_k^{(p)} - T_k^{(p)}) O_j^{(p)}$$

また, 下層結合係数の修正も, 4.1 と同様に教師決定因子

$$S_{jk}^{(p)} = (O_k^{(p)} - T_k^{(p)}) W_{kj}$$

の符号に基づき中間層の教師信号 $T_j^{(p)}$ を設定し, 次式のデルタ則に従い修正する.

$$W_{ji} = W_{ji} - \alpha (O_j^{(p)} - T_j^{(p)}) O_i^{(p)}$$

しかし, 出力層素子 k が複数個あるので, $S_{jk}^{(p)}$ の符号が異なるとき互いに競合する. そこで, $S_{jk}^{(p)}$ の k に関する総和を改めて教師信号決定因子 $S_j^{(p)}$ とおき競合を解消する. すなわち

$$S_j^{(p)} = \sum_k (O_k^{(p)} - T_k^{(p)}) W_{kj}$$

とおき, 次式により教師信号を設定する.

$$\text{if } S_j^{(p)} > 0 \text{ then } T_j^{(p)} = 0$$

$$\text{if } S_j^{(p)} < 0 \text{ then } T_j^{(p)} = 1$$

$S_j^{(p)} = 0$ のときは, 因子 $S_{jk}^{(p)}$ の競合によりバランスし, 修正する必要がないので

$$\text{if } S_j^{(p)} = 0 \text{ then } T_j^{(p)} = O_j^{(p)}$$

とおく.

表 1 DBP 法の間層教師信号の設定 (出力層素子が 1 個の場合)

$O_k^{(p)}$	$T_k^{(p)}$	W_{kj}	$S_j^{(p)} = S_{jk}^{(p)} = (O_k^{(p)} - T_k^{(p)}) W_{kj}$	$T_j^{(p)}$
0	1	正	負	1
		負	正	0
1	0	正	正	0
		負	負	1
0	0	-	0	$O_j^{(p)}$
1	1			

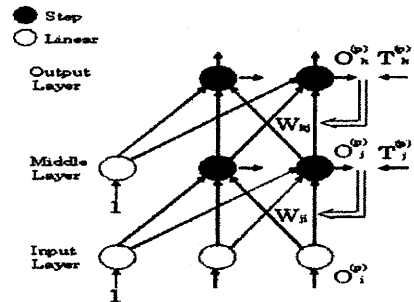


図 5 DBP 法の学習 (出力層素子が複数個の場合)

5. 提案手法

DBP 法の教師信号の更新式に条件を加えることで教師信号に制限を与え, 適切なネットワークを作ることを目的とする方法を提案する. 具体的なアルゴリズムは, 出力素子数が複数個あるときの DBP 法の教師信号を更新する式に以下のような条件式を確率的に加える.

$$\text{if } \sum_k (O_k^{(p)} - T_k^{(p)}) W_{kj} > 1 \text{ then } T_j^{(p)} = 0$$

$$\text{if } \sum_k (O_k^{(p)} - T_k^{(p)})W_{kj} < -1 \text{ then } T_j^{(p)} = 1$$

$$\text{if } -1 \leq \sum_k (O_k^{(p)} - T_k^{(p)})W_{kj} \leq 1 \text{ then } T_j^{(p)} = O_j^{(p)}$$

この条件式を 10% の確率で DBP 法に導入する。この確率は 7 で示すが、実験から経験的に定めた数字である。確率で条件に入らないときは通常通りの処理を行う。これより、教師決定因子である $\sum_k (O_k^{(p)} - T_k^{(p)})W_{kj}$ の値にあまり変化のない場合は教師信号の変化を制限する方法である。

アルゴリズムを Fig.6 に示す。以下に示すように、中間層の教師信号を求めるときに提案手法を導入する。

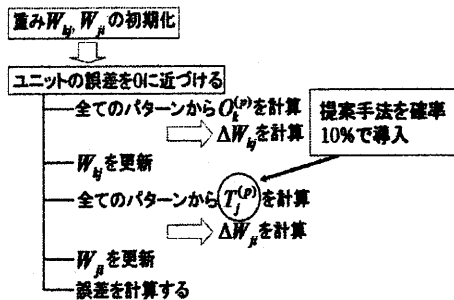


図 6 提案手法のアルゴリズム

6. 実験

本研究で計算機実験に用いた対象問題は、XOR 問題、半加算器問題、全加算器問題、パリティ問題の 4 つである。これらは線形分離不能な問題である。つまり、1 本の線でクラスを 2 つに分けられないという問題である。本研究においての、それぞれの問題の特徴を Table.2 に示す。

表 2 対象問題の特徴

対象問題	特徴
XOR 問題	2 つの入力と 1 つの出力
半加算器問題	2 つの入力と 2 つの出力
全加算器問題	3 つの入力と 2 つの出力
パリティ問題	3 つの入力と 1 つの出力

これらの問題を用いた理由としては、ネットワークを構成するニューロン数を変えたときに解がどのように変化するかを見るためである。

XOR 問題、半加算器問題、全加算器問題、パリティ問題を対象に実験を行った。そこで SP 法、ABP 法、DBP 法、提案手法の各手法における学習成功率と平均学習回数を比較した。実験で用いたパラメータとしては、SP 法、DBP 法、提案手法では学習係数 α を 0.002、ABP 法では学習係数 α を 2 と設定した。最大学習回数は 3 万回と少し大きめの範囲をとった。ここでの学習係数は予備実験を行った際に比較的安定して、良好の解が得られた値である。最大学習回数は収束させるという意味で、大きめにとることとした。また、中間層素子数は、2 ビット入力課題では 2、3 ビット入力課題では 3 とした。中間層の数も計算時間(学習回数)を考えうえでの妥当な数を設定した。計算機環境

表 3 計算機環境

OS	Windows XP
CPU	Athlon XP 2600+ 2.13GHz
Memory	768MB
Compiler	Visual C++6.0

は Table.3 に示すとおりである。

今回、提案手法で用いる条件式を確率 10% で導入する理由を予備実験を元に説明する。まず実際に半加算器問題を例に実験を行い、確率を 0~40% まで変動させ、その結果を Fig7 に示す。

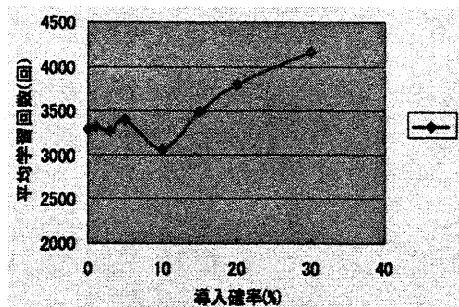


図 7 半加算器問題における確率を変動させた学習回数

この結果からもわかるように、10% のときにもっともよい結果が得られたので、その数値を引用することにした。

Table.4 より、どの課題においても SP 法の学習成功率が極端に低い。これは線形分離不能なパターン集合に対して、中間層出力で線形分離可能となる確率が極めて低いことを意味する。また、提案手法はほとんどの実験において比較手法よりも成功率が高かった。提案手法の学習性能は個々の問題にかなり依存するが、半加算器、全加算器などの複数出力課題において、良好の解が求められると考えられる。また、パリティ問題においては ABP 法の方が学習成功率は良かったが、Table.5 からわかるように、平均学習回数は大幅に多いことから、提案手法の方が実用的であると考えられる。

Table.5 から、提案手法は比較手法よりも平均学習回数が少なく、パリティ問題を除いて学習成功率が高いほど平均学習回数が減少する傾向にあることがわかった。

表 4 学習成功率の比較 (%)

		SP	ABP	DBP	提案
入力 2 ビット	XOR	1	77	82	84
	半加算	0	84	100	100
入力 3 ビット	パリティ	0	100	90	93
	全加算	0	79	100	100

表 5 平均学習回数の比較

		SP	ABP	DBP	提案
入力 2 ビット	XOR	190	3277	2001	1829
	半加算	-	5269	3292	3058
入力 3 ビット	パリティ	-	4576	3885	3413
	全加算	-	7113	4727	4482

線形分離不能である 4 つの問題を解くにあたって、ネットワークの形を変えても提案手法を導入した方が良好の解が得られた。これは本研究では提案した条件式を確率的に導入しているからであると考えられる。

実験で用いた 10% という確率は妥当な値だったと思われる。予備実験で示したように、確率を大きくすると、条件式の影響が大きくなってしまい学習に時間がかかるという傾向が見られた。また、確率を小さくし過ぎると条件式のよさが表れず、従来法と同じような値になってしまった。

このことからわかるように、妥当な確率で導入することで提案手法は有効であると思われる。

また、問題の特徴により解にばらつきがあることが確認できた。実験結果からも分かるように、DBP 法と同様であるが、出力が複数のネットワークに対して、比較的良い解が得られる傾向があることもわかった。

7. ま と め

本研究では、すべての層に不可微分な 2 値出力素子を用いた三層パーセプトロンに対して、DBP 法の教師信号の更新に条件式を確率的に導入し、制限を加えた方法を提案した。その結果、従来法と比較して良好の解を得ることができた。今後の課題として考えられるのは、ネットワークサイズを大きくしたときの解の安定性を図ることである。

文 献

- [1] D.E.Rumelhart, G.E.Hinton, and R.J.Williams, "Learning representation by back-propagation errors", Nature, Vol. 323, pp.533- 536, 1986.
- [2] 山田樹一, 黒柳奨, 岩田彰, "ニューロンモデルに対する双対性を利用したニューラルネットワークのための教師あり学習手法", 電子情報通信学会論文誌, Vol.J87-D-II, No.2, pp.399-406, 2004.
- [3] 長沼秀典, 大堀隆文, 渡辺一央, "微分不能素子をもつ階層型ニューラルネットに対する誤差逆伝搬法の提案", 電子情報通信学会論文誌, Vol.J88-D-II, No.10, pp.2125-2133, 2005.
- [4] 甘利俊一, 神経回路網の数理, 産業図書, 1978.
- [5] 佐藤紀章, 萩原将文, "3 次元物体認識のための並列階層型ニューラルネットワーク", 電子情報通信学会論文誌, Vol.J86-D-II, No.4, 2003.
- [6] 折川典生, 原田豊, "伝達関数の改良による誤差逆伝搬学習", 電子情報通信学会論文誌, Vol.J83-D-II, No.2, pp.852-854, 2000.
- [7] 山森一人, 堀口進, "並列計算機上の誤差逆伝搬学習法の並列学習モデル", 電子情報通信学会論文誌, Vol.J81-D-II, No.2, pp.370-377, 1998.
- [8] 永澤和行, 福村直博, 宇野洋二, "多層神経回路が逆モデルを学習するための Forward-propagation 則", 電子情報通信学会論文誌, Vol.J85-D-II, No.6, pp.1066-1074, 2002.
- [9] 大濱吉紘, 福村直博, 宇野洋二, "逐次最小 2 乗法による逆モデル神経回路学習のための Forward-propagation 則", 電子情報通信学会論文誌, Vol.J87-D-II, No.2, pp.726-734, 2004.
- [10] 山田浩太郎, 田口亮, 曾禰元隆, "逆伝搬誤差可変型学習法による学習停滞・停止回避", 信学論, Vol.J80-D-II, No.2, pp.626-633, 1997.
- [11] 渡辺栄治, "パターン認識問題に対する階層型ニューラルネットワークの汎化能力改善学習法", 信学論, Vol.J79-D-II, No.5, pp.917-923, 1996.
- [12] 殿村正延, 中山謙二, "多層パーセプトロンにおける内部情報最適化アルゴリズムと汎化能力の解析", 信学論, Vol.J84-D-II, No.5, pp.830-842, 2001.
- [13] 大堀隆文, 鎌田修平, 小林靖宏, 小西儀紀, 渡辺一央, "微分不能素子をもつ階層型ニューラルネットに対する誤差逆伝搬法の提案", 信学技報, NC99-85, Feb. 2000.