

デジタル型誤差逆伝播法における学習改良の研究

杉本, 和也 / SUGIMOTO, Kazuya

(発行年 / Year)

2013-03-24

(学位授与年月日 / Date of Granted)

2013-03-24

(学位名 / Degree Name)

修士(工学)

(学位授与機関 / Degree Grantor)

法政大学 (Hosei University)

P377.5

M34

2012-3

2012 年度 修士論文

論文題名

デジタル型誤差逆伝播法における学習改良の研究

指導教員

李 磊 教授

法政大学大学院工学研究科情報電子工学専攻

学籍番号 10R4129

氏名 杉本 和也

文部省 1913 年

書影文録

京都府立総合資料館蔵書目録



京都府立総合資料館蔵書目録

PC11A01 言語学

言語学 本館 蔵書

Abstract

A learning based error back propagation algorithm, a proposed non-differential digital back propagation by using two-valued output element is a multilayer neural network (DBP). Traditional simple Perceptron learning without teacher signal only to the output layer, DBP learning algorithm gives the teacher signal to reduce output error to the middle layer. In this paper, Non-linear separable issues to solve that Using the concept of tolerance for giving teacher signal in DBP algorithm, For Improving learning according to effect of an appropriate learning parameter settings, and the high speed calculation method by batch learning error function gradient Resilient back-propagation algorithm (Rprop algorithm), confirmed that fluctuations in the success rate and the number of learning.

目次

1 序論	
1.1 研究背景	1
1.2 研究目的	2
1.3 本文構成	2
2 ニューラルネットワーク	3
2.1 神経細胞(ニューロン)	3
2.2 ニューロンモデル	4
2.3 階層型ニューラルネットワーク	6
3 線形分離可能不可能問題	7
3.1 線形分離可能	7
3.2 線形分離不可能	8
4 誤差逆伝播法(BP法)	9
4.1 パーセプトロン	9
4.2 単純パーセプトロン(SP法)	10
4.3 アナログ型誤差逆伝播法(ABP法)	12
4.4 弾性バックプロパゲーション(Rprop法)	14
5 デジタル型誤差逆伝播法(DBP法)	15
5.1 出力素子が1個の場合	16
5.2 出力素子が複数個の場合	19
5.3 許容範囲設定	20
6 提案手法	21
6.1 Rprop法に許容範囲概念の適用	21

6.2	通常 DBP 法,改良 DBP 法,改良 Rprop 法の組み合わせ(確率的)	21
7	計算機実験	23
7.1	実験環境	23
7.2	実験設定	23
7.3	実験結果	28
8	考察	30
9	結論	31
10	謝辞	32
	参考文献	33

1 序論

1.1 研究背景

現在,ニューラルネットワーク(Neural Network)には様々な種類があり,階層型(パーセプトロン,RBF)等が用いられている.これらによって人によるパターン認識をコンピュータが擬似的に行う構成や,教師的に信号を与えて学習を行い,与えられた問題に最適化されていく教師あり学習などができる.

階層型ニューラルネットワーク学習に可能性を持たせる土台となった方法には Rosenblatt によって提案された入力層と2値出力素子をもつ出力層の2層からなるパーセプトロン(単層パーセプトロン)というものがある.この方法でパターン認識を行う場合,入力のパターン集合が線形分離可能(AND や OR の出力)であるならば有限回学習においてカテゴリ間の識別が可能である.しかし入力パターンの中に一つでも線形分離不能なカテゴリ(XOR などの出力)が存在するならば学習不能という欠点がある.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)という方法で解決できる.しかしこの方法も中間層に十分多くの素子を用いて入力-中間層間でのランダム変換によるものであり³⁾,工学的に実現可能なネットワークにするには中間層素子数をある程度抑制する必要があるという点と下層(入力-中間)の結合係数を修正できない点から理想的とはされなかった.

そこで不可微分な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誤差逆伝搬法であるデジタル型誤差逆伝播法(DBP 法)というものが提案され.これは従来の SP 法で出力にしか教師信号を与えないのに対して,DBP 法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のみではなく下層結合係数の修正も行うことができ,伝播学習概念が構成された.

本稿では,DBP 法において中間層の教師信号を求める時に従来では制限を行っており¹⁾,この概念を許容範囲として用いる事でどのような改善があるのかを考察すること.

またバッチ学習である結合係数更新を一括に行う Resilient バックプロパゲーション(Rprop 法)には従来 DBP 法と同じように許容範囲の概念を取り入れ,従来 DBP 法と組み合わせをすることで学習の向上を図る.この提案した手法を用いて線形分離不能問題を対象とした実験を行い,従来手法との比較を学習成功率,平均学習回数から確認していく.

1.2 研究目的

本稿では,許容範囲を取り入れた DBP 法を用いた手法 1) と Rprop 法にも許容範囲の概念を取り入れ(改良 Rprop とする),2 つの方法を合わせたものを提案手法とする.
性能評価を行うにあたって従来手法である DBP 法,許容範囲を取り入れた改良 DBP 法,提案手法を用いて線形分離不能問題を対象に実験を行い,学習成功率,平均学習回数の結果から提案手法の有効性を検証する.

1.3 本文構成

- 1 章では研究背景と目的についての記述.
 - 2 章ではニューラルネットワークの根本的な概念と学習,特徴等についての記述.
 - 3 章では線形分離の可能不可能な問題についての例を用いて記述.
 - 4 章では誤差逆伝播法を用いた学習についての方法や概念の例についての記述.
 - 5 章では従来手法の概念とどのように改良されてきたのかについての記述.
 - 6 章では提案手法についての方法やアルゴリズムについての記述.
 - 7 章では実験環境,実験設定,実験結果を記述.
 - 8 章では考察.
 - 9 章では結論.
 - 10 章では謝辞.
- 最後に参考文献の一覧を記述.

2 ニューラルネットワーク

以下にニューラルネットワークについて説明をする.

2.1 神経細胞(ニューロン)

ニューロンとは Fig.2.1 にあるように細胞体,樹状突起,軸索の 3 つの部より成り立っていて,細胞体がニューロン本体,樹状突起が入力部分,軸索が出力部分というようになっている.人の脳にはこのニューロンが 100 億から 140 億ともいわれる程の膨大な数の神経細胞として存在している.単一で見るとニューロンというものは入力(樹状突起)と出力(軸索)の関係を持つ情報処理素子であると見なすことができる.そしてニューロンの出力は他のニューロンと結合しており,結合されたニューロンがまた別のニューロンに結合しているというように人の脳にはニューロンが無数に結合した神経回路網(ニューラルネット)が構成されている.つまり出力機能を担っている軸索が結合途中で何本も枝分かれをしていて,多数の他ニューロンの樹状突起に繋がっている.この結合部はシナプスと呼ばれ,ニューロンの情報は 1 つの細胞の軸索から他の細胞の樹状突起へとシナプスを介して伝播されるという機能を持つ.

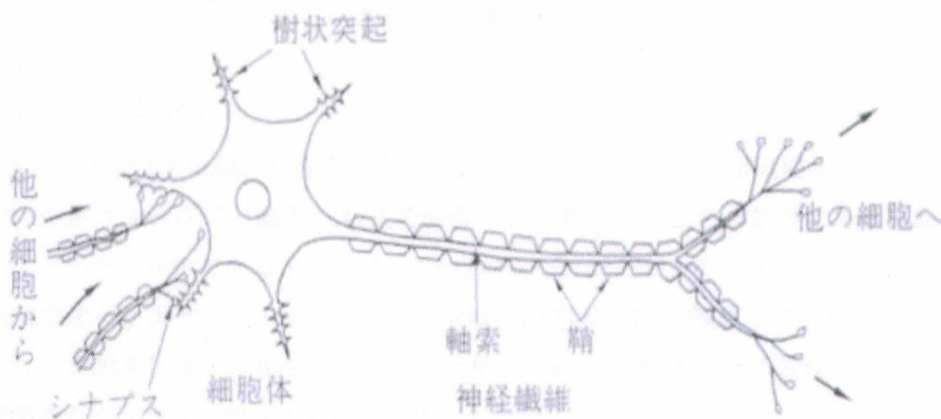


Fig. 2.1 神経細胞 (ニューロン) の概念

ニューロンの基礎機能としては,神経細胞に入力となる刺激が入ってきた場合に,活動電位というFig.2.2に示すような他のそれぞれの組織において異なる性質を持ち,発生させることで他の各細胞に情報を伝達する.活動電位がおきる条件では閾値を変化させる,他の多数細胞から入力することで情報の修飾が行われ,それぞれの出力に変化をつける.

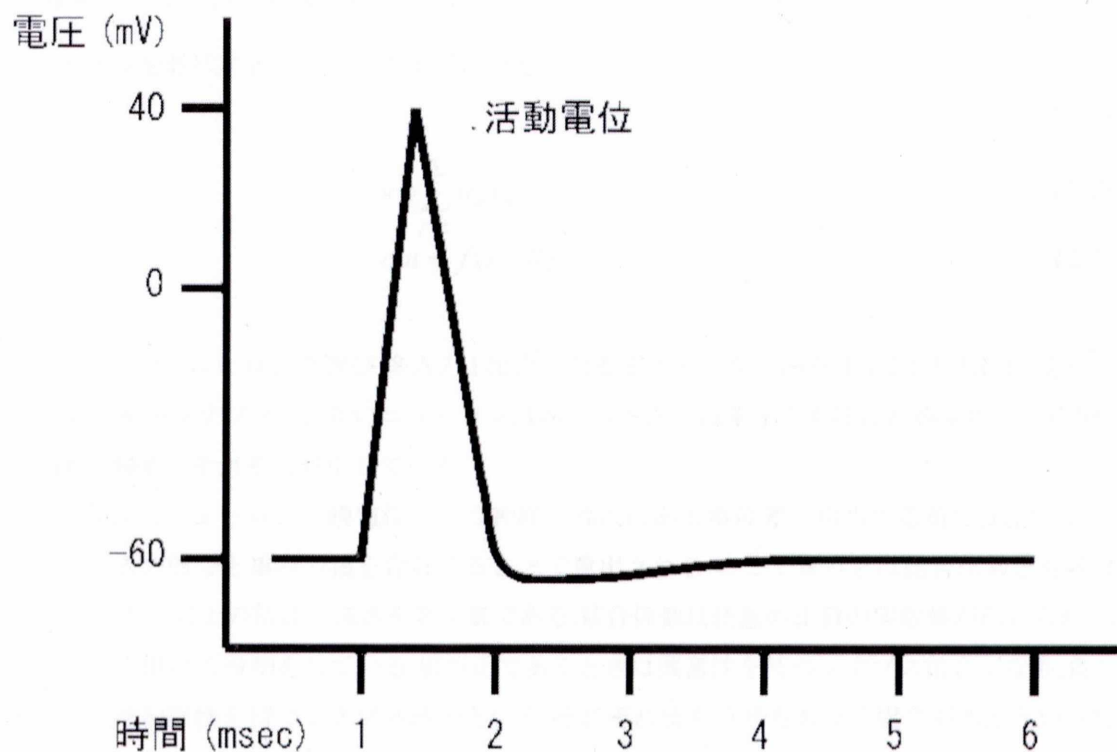


Fig. 2.2 活動電位

2.2 ニューロンモデル

2.1で示した機能は人の脳に存在するものであり,それを工学的に情報処理システムとして活用するには生体ニューロンの働きを,数式で表現する必要がある.この自然界にある事象を数式表現することがモデル化である.以下の Fig.2.3 はニューロンモデルの一部を表したものである.しかし前項で述べたように人の持つニューロンは実際膨大な数と,複雑な構成をしていて,全てのニューロンを忠実にモデル化することは不可能とされている.そのため表現に用いる数式もある程度簡単かつ人ニューロンの動作的本質を失わないことが必要条件である.

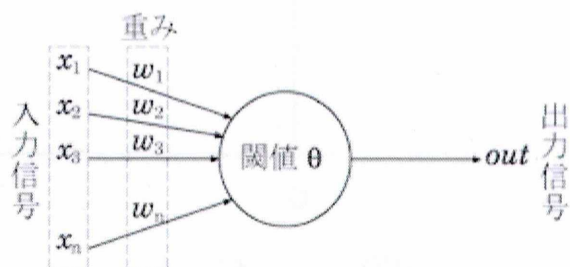


Fig. 2.3 ニューロンモデル

Fig.2.3 の x_1, x_2, \dots, x_n は,他ニューロンからの入力を表し, out は出力を表す.

このモデルを数式に表すと以下のようなになる.

$$s = \sum_{n=1}^N w_n x_n \quad (2.1)$$

$$out = f(s - \theta) \quad (2.2)$$

まず1つのニューロンを選び,多入力1出力となる素子モデルである. Fig.2.1 と Fig.2.3 を比べると重み w がシナプスに,丸いニューロンはユニットまたは素子とも呼ばれ細胞体に,出力信号 out は軸索にそれぞれ対応している.

まず s は人ニューロンの膜電位という神経の外内にある電位差に相当する量で,式(2.1)に示すように入力信号と重みの積を合計することで算出される.ここで重みとは結合係数とも呼ばれ,ニューロン同士の結合の強さを表す量である.結合係数は任意の正負の実数値が用いられ,入力信号 x に掛ける役割をしている.値が正であるときは興奮性を持つシナプス結合になり,負であるときは抑制性を持つシナプス結合という.それぞれ値も符号も異なる場合があるという特徴を持つ.この結合荷重の値を変化させていくことで,出力を変化させて学習を行うという概念をコンピュータに情報処理として行わせるのである.

またユニットの中に閾値 θ というのがあり,ユニット自体が興奮して活動電位を発生させるか興奮しないかの閾値として存在し, s が閾値を超えるかどうかの判定する.よって式(2.2)のように $s - \theta$ を活性化関数 $f(x)$ に代入して得られた値を出力とすることが一通りの流れである.

McCulloch-Pitts-typeにより最初に提案された活性化関数はFig.2.4に示すようなステップ関数が用いられており,0か1を出力する問題においては現在も使用されている.

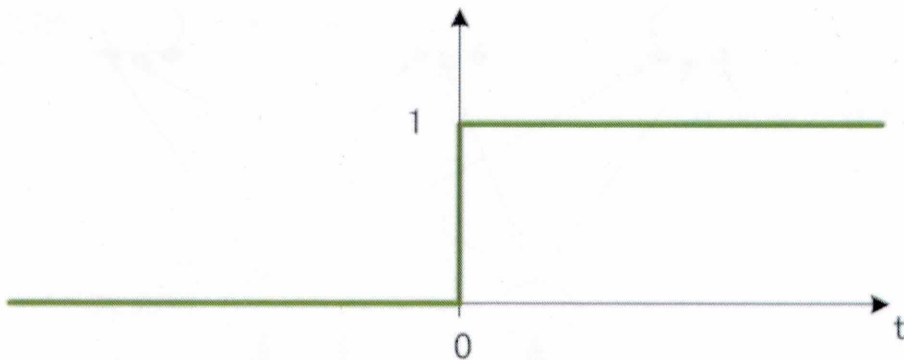


Fig. 2.4 ステップ関数

2.3 階層型ニューラルネットワーク

脳の神経系は神経細胞を単位として構成されており、それらが互いに結びついて広大なネットワークを作り出している。そしてニューロンを代わりに用いて工学的ネットワークとして構築するのにFig. 2.5のような概念図を階層型ニューラルネットワークという。

まず入力信号を受けて他ニューロンへ分配するための入力用ニューロン、これを有する層は入力層と呼ばれる。次に外部へ出力する信号を有する最終的な層は出力層と呼ばれ、ここには目的の結果を出力するという機能のニューロンで構成されている。さらに入力層と出力層の間には入力を調整し、幅広い出力を行うための層があり、中間層と呼ばれる。

この階層型ニューラルネットワークによる学習の流れはまず入力層ニューロンに入力値が入り、中間層ニューロンの重み量によって中間層ニューロンの出力調整を行い、0か1などを中間層ニューロンから出力して出力層の入力となり、出力層の重みによってさらに出力層ニューロンの出力を変化させ様々な目的の結果を出力させるという一定方向の前方伝播になる。

Fig. 2.5の教師信号とはある目的の値を出力させようとするときに予めその値を用意しておき、学習機器に与えるべき答えとなるものである。これを与える学習を教師あり学習と呼ばれ、現在でもよく使用されている。

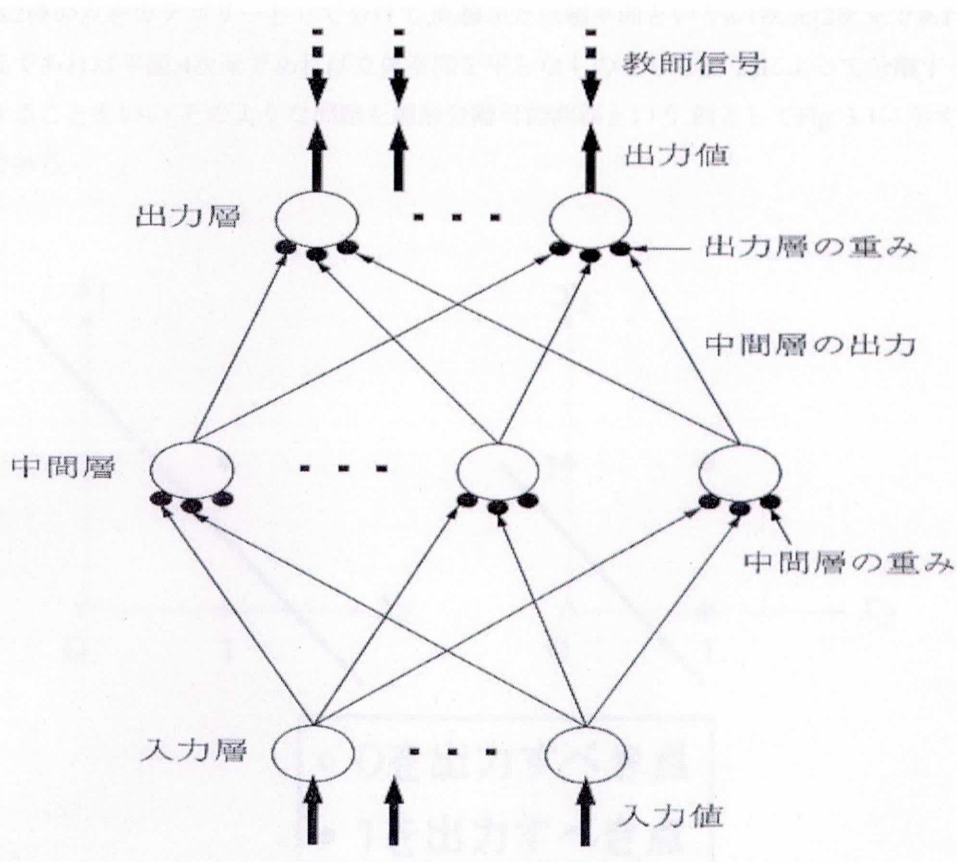


Fig. 2.5 階層型ニューラルネットワーク

特徴として、任意関数近似に優れている、また誤差逆伝播法による学習方式を用いることで学習の改良を行うという点は、誤差を逆から伝播させるという点を人が行う行動で見ると、悪い評価となる製品のものは質を上げるようにし、良い評価となる製品のものはさらなる質の向上か継続を行うといったように、まさに工学的な学習というものを階層型ニューラルネットワークでも行っているのである。

理論として、中間層のニューロン数を増やすと、情報の増加による学習精度向上を行うことができる。しかし本来ニューラルネットワークの欠点として大規模な対象問題における高精度な制御や学習しすぎると今度は過学習という過度の計算による入力データに存在する誤差(ノイズ)によって精度が落ちてしまうといった問題がある。

このようにならないためどのくらいのニューロン数や式を使用すればいいのかという完全な答えはまだ決まっていない。そのため人の脳を完全再現することは不可能とされている。

3 線形分離可能不可能問題

線形分離可能とは、ANDやORといった、0を出力する点と1を出力する点を式(2.1)で表し、それぞれ2種の点をカテゴリーとして分けて、直線または超平面という $n-1$ 次元(2次元であれば直線、3次元であれば平面、4次元であれば立体空間を平らなものとして扱う)によって分離することができることをいい、そのような問題を線形分離可能問題という。例としてFig. 3.1に示すようなものである。

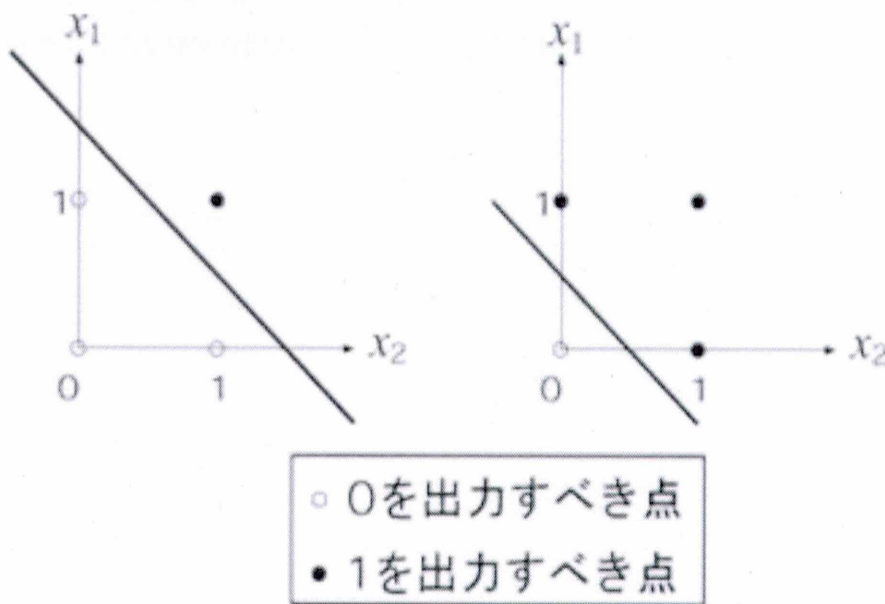


Fig. 3.1 線形分離可能 (左がAND, 右がOR)

線形分離不可能とは, 上記のようなFig3.1に示されるように直線1本で,0出力と1出力のカテゴリ分類ができない状態をいう.例として以下のようなFig3.2の排他的論理輪(XOR)が該当する.

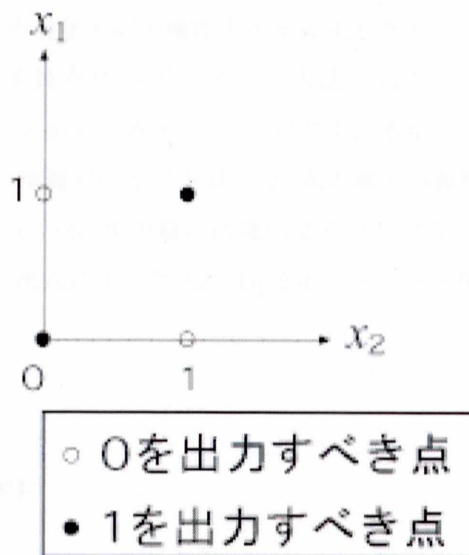


Fig. 3.2 線形分離不可能 (XOR)

この問題は1つのニューロンでは実現不可能な論理関数とされている.
この問題を解くためには2本の直線が必要であり,中間層と最低でも2つのニューロンが必要とされている.この問題を解くために階層型ニューラルネットワークというものをいられ,その他の分離不可能な様々な問題にも使用されるようになった.

4 誤差逆伝播法(BP 法)

4.1 パーセプトロン

パーセプトロンは単層パーセプトロンとも呼ばれ、Fig.4.1 に示す。Rosenblatt により提案されたニューラルネットにおける視覚と脳の機能をモデル化したもので、ニューラルネットワークに始めて学習という可能性を持たせ、土台となった方法である。これはパターン認識機械の一種である。入力されたパターンがどのカテゴリー(同じであるかどうか判断する集合)に所属するかを認識する機能を持つ。単層パーセプトロンは、入力素子が線形、出力素子が線形閾値の二層ネットワークで構成されている。出力層には識別するカテゴリー数の素子を配し、ある入力パターンが与えられたとき、次式に示した式が Fig.2.4 のステップ関数となり、0 か 1 を出力する。

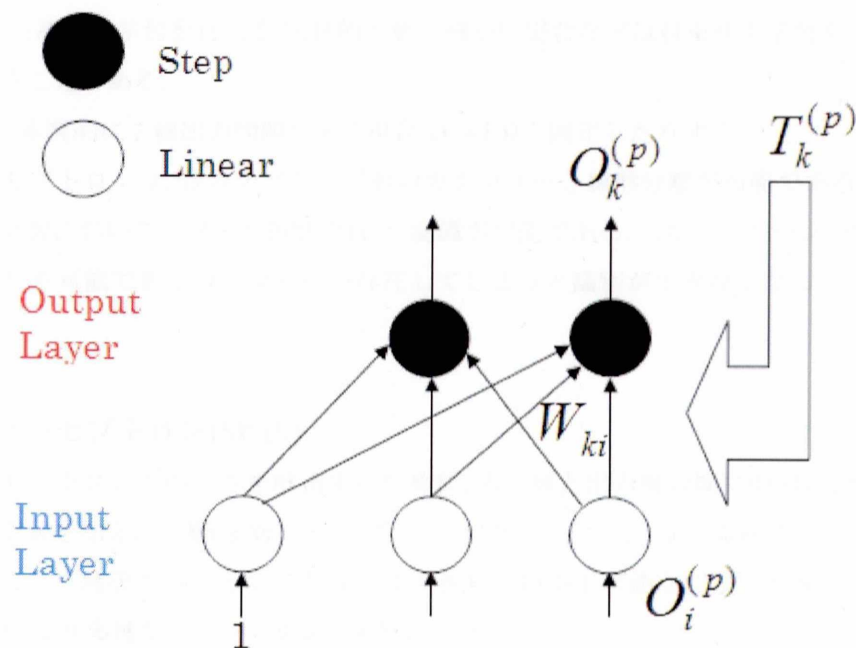


Fig. 4.1 単層パーセプトロン

$$U_k^{(p)} = \sum_i W_{ki} O_i^{(p)} \quad (4.1)$$

$$O_k^{(p)} = f(U_k^{(p)}) = \begin{cases} 0 & (U_k < 0) \\ 1 & (U_k \geq 0) \end{cases} \quad (4.2)$$

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し、式(4.1)はそれぞれ第 p パターン入力時の第 k 出力素子の活性化値、式(4.2)はそれぞれ第 p パターン入力時の第 k 出力素子の出力を表し、 $T_k^{(p)}$ は問題パターンの教師信号、 $O_i^{(p)}$ は第 i 入力素子の出力を表す。ここでパターンとは教師的信号となる組み合わせを(例としてAND問題を見ると入力が2つ、出力が1つであり、その組み合わせが4つなので4パターンとなる)を指す。学習目的は問題の全パターンについてパターン毎に繰り返し学習が行われ、出力と教師信号の誤差を減らしていくことでパターン集合の分離を可能にすることである。学習方法は学習パターン p に対する出力 $O_k^{(p)}$ と教師信号 $T_k^{(p)}$ の誤差を基にして W_{ki} は次式(デルタ則)により更新する。

$$W_{ki} = W_{ki} - \alpha(O_k^{(p)} - T_k^{(p)})O_i^{(p)} \quad (4.3)$$

ここで、 α は学習係数と呼ばれる正の定数である。この値が小さいほど学習は細かく行われるが、時間を要してしまい、また学習が進まなくなってしまう。逆に値が大きいほど学習は大部分を変動させて、高速に学習を行えるが、目的の値が細かい場合などは収束せず学習が失敗して終了するということがある。

0 か 1 を扱う本質的に 2 値出力問題である場合、 $\alpha = 1.0$ と固定したりする。

単層パーセプトロンは、各カテゴリーが他のカテゴリーと線形分離が可能であるならば、有限回の反復学習においてパターン識別を行え、認識が可能である。しかしパターンの中に1つでも線形分離が不可能であるカテゴリーが存在してしまうと識別ができなくなってしまう欠点がある。

4.2 単純パーセプトロン(SP法)

単層パーセプトロンでの欠点を解消するために、入力層と出力層の間に0か1の2値出力素子からなる中間層を加えた三層(多層)パーセプトロン(SP法, Fig.4.2), つまり階層型ニューラルネットワークによって解決することができる。つまり線形分離が不可能な問題でも解くことができるようになり、より多様なパターン認識を可能にした。

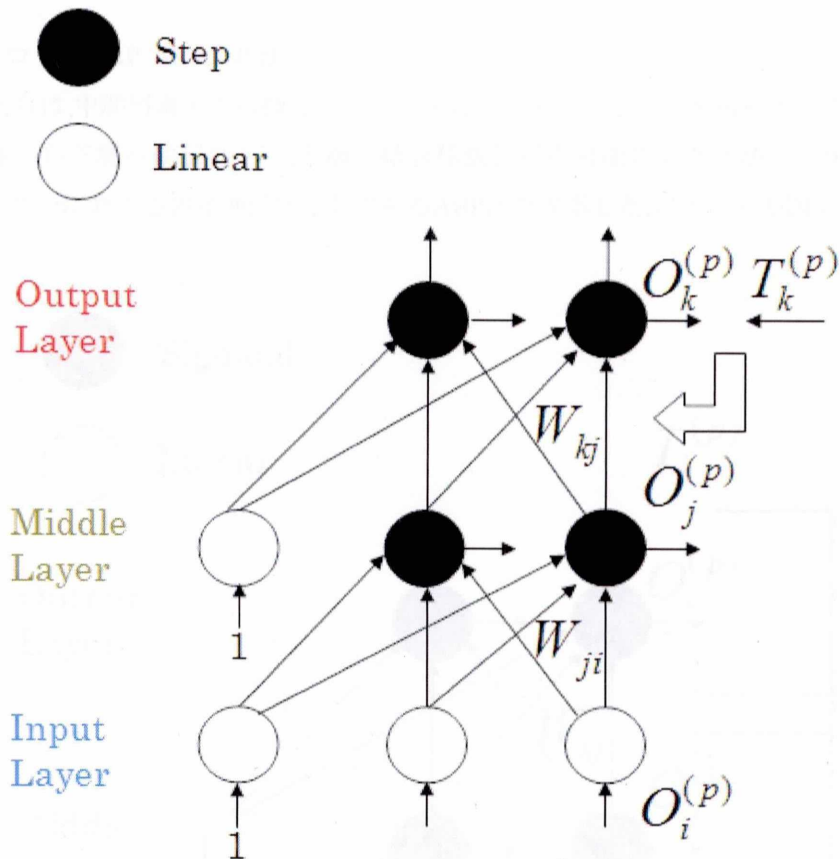


Fig. 4.2 単純パーセプトロン (SP法)

Fig. 4.2の入力層と中間層の左に示される入力1のニューロンは、ネットワーク簡略化のために閾値を結合係数として扱うような方法もよく用いられる。

$$W_0 = -\theta, \quad O_0^{(p)} = 1 \quad *i, j \text{のみ} \quad (4.4)$$

SP法では中間層に十分に多くの素子を用いる事によって入力層-中間層の間によるランダムな変換、つまり結合係数を乱数で扱うのでそれに任せてしまう形になる。これにより線形分離不可能なパターン集合でも分離が可能になるとされ、パーセプトロンの欠点が克服された。

しかし、膨大な処理を行い、必要とする情報処理工学においてはいかに単純なネットワーク処理にするかが重要であり、魅力が薄いとされた。つまり人の脳では簡単な問題や単純な内容では刺激が薄い、ネットワークとしては小さいものということためである。

さらにSPでは中間層-出力層間である上層のみの結合係数修正しか行えず、ネットワーク全体から見ると学習という点においての流れが欠けてしまっているため学習性能の向上を期待す

ることはできなかった.

4.3 アナログ型 BP 法(ABP 法)

Rumelhartらは,中間層素子を可微分なアナログ素子でネットワークを構成することより,出力誤差が上層から下層へ逆伝播して,下層の結合係数修正を可能にした方法をアナログ型多層パーセプトロンに対する誤差逆伝播学習であるABP法を提案した.以下にその図を示す.

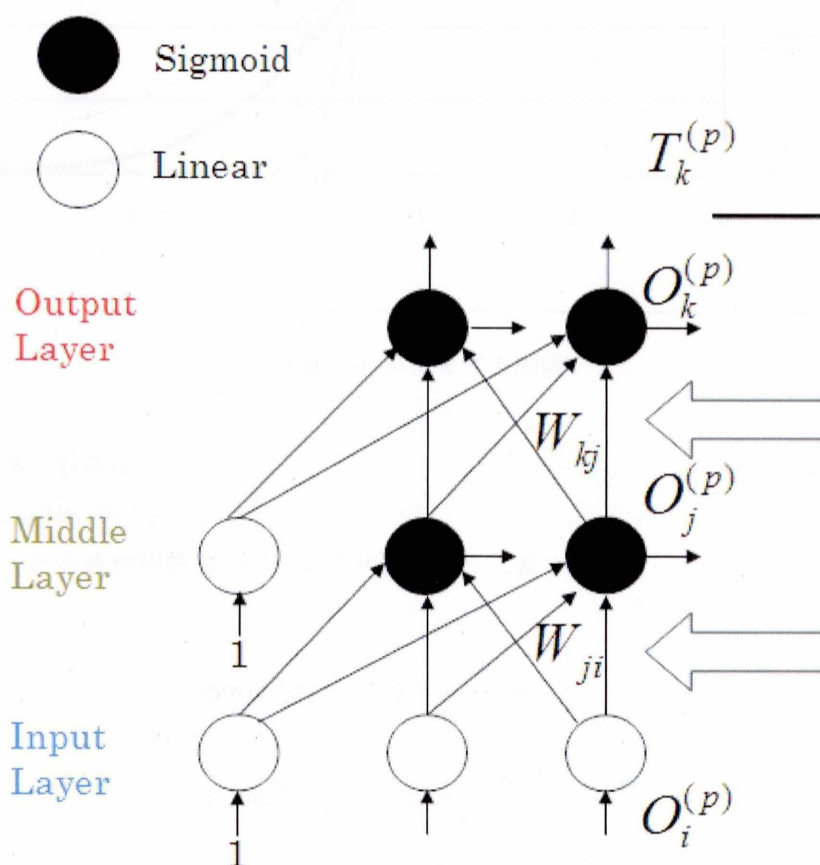


Fig. 4.3 アナログ型誤差逆伝播法 (ABP法)

中間層,出力層の出力値は,次式で定義する.

$$O_j^{(p)} = f(U_j^{(p)}) = f\left(\sum_i W_{ji} O_i^{(p)}\right) \quad (4.5)$$

$$O_k^{(p)} = f(U_k^{(p)}) = f\left(\sum_j W_{kj} O_j^{(p)}\right) \quad (4.6)$$

これらの出力素子関数は次のようなシグモイド関数,式(4.7),Fig. 4.4を用いる.

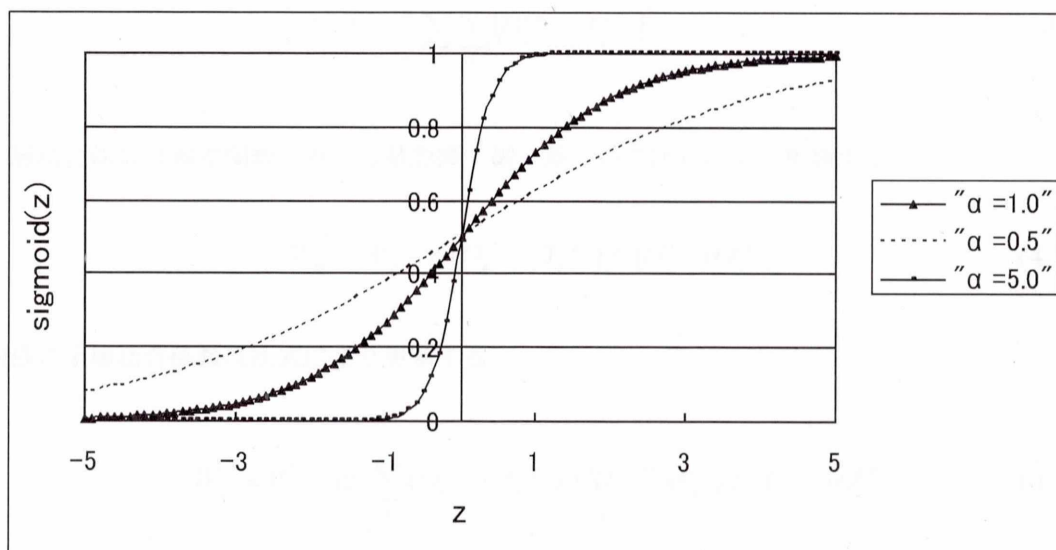


Fig. 4.4 シグモイド関数

この出力関数の特徴は,

: 出力範囲は(0.0,1.0)である.

: 微分可能である連続関数.微分式は以下の(4.8)ように表せる.

$$\text{sigmoid}(z) = f(z) = \frac{1}{1+e^{-\alpha z}} \quad (4.7)$$

$$\begin{aligned} \text{sigmoid}'(z) = f'(z) &= \frac{\alpha e^{-\alpha z}}{(1+e^{-\alpha z})^2} \\ &= \alpha \frac{1}{1+e^{-\alpha z}} \left(1 - \frac{1}{1+e^{-\alpha z}} \right) \\ &= \alpha \cdot \text{sigmoid}(1 - \text{sigmoid}) \end{aligned} \quad (4.8)$$

: $z=0$ のとき,微分値は最大となり,0.25 となる.

以上の特徴を用いることで細かい実数値を出力できるので,ニューラルネットにおける学習によく用いられる.

次に誤差関数を示す.

$$E = \frac{1}{2} \sum_p \sum_k (O_k^{(p)} - T_k^{(p)})^2 \quad (4.9)$$

上層結合係数は誤差関数に対する最急降下則に基づいて次式により更新する.

$$W_{kj} = W_{kj} - \alpha (O_k^{(p)} - T_k^{(p)}) f'(U_k^{(p)}) O_j^{(p)} \quad (4.10)$$

同様に下層結合係数は次式により更新する.

$$W_{ji} = W_{ji} - \alpha \left\{ \sum_k (O_k^{(p)} - T_k^{(p)}) f'(U_k^{(p)}) W_{kj} \right\} f'(U_j^{(p)}) O_i^{(p)} \quad (4.11)$$

ABP法は汎用性の高さから多くの学習に応用されてきたが,最急降下法による誤差関数の最小点探索を行う部分に関して,以下のような問題点がある

- 1:線形分離不可能問題に対して,初期値によっては極小解に陥り学習不能になる場合が少なくない.最小解に収束する場合でも学習に時間が掛かり,学習回数を膨大に要してしまう.
- 2:処理速度の向上や,ハードウェア化の容易性に対して,0と1の2値からなる本質的な問題とされるバイナリ型問題を直接的に解くことはできない.よってなんらかの手を加えるか,バイナリ型問題で近似を行う場合などに用いる必要が出てくることもある.

4.4 弾性パーセプトロン(Rprop 法: Resilient backpropagation)

Rprop 法は, Riedmiller によって提案されたフィードフォワードニューラルネットワークの一括で行うバッチ学習アルゴリズムであり,勾配法の一つである.バッチ学習の処理は制限が無いなどで様々な学習において活用できる.学習方法は,誤差関数勾配の符号変化に基づいて,結合係数の修正量を算出するので,一括学習において高速な学習が可能とされている.

ある学習ステップ t において下層での誤差勾配関数は以下の式のようになる.

$$\frac{\partial E(t)}{\partial W_{ji}(t)} = g_{ji}(t) \quad (4.5)$$

$$f(x) = \frac{1}{1+e^{-x}} \quad (4.6)$$

$$g_{ji}(t+1) = (1 - f(x))f(x) \sum_i g_{ji}(t)W_{ji}(t) \quad (4.7)$$

学習式は以下のようになる。

$$W_{ji}(t+1) = W_{ji}(t) - \frac{g_{ji}(t)}{|g_{ji}(t)|} \Delta_{ji}(t) \quad (4.8)$$

ここで $\Delta_{ji}(t)$ は $g_{ji}(t)$ の符号変化に基づいて次式のように更新される。これは上層も同様である。

$$\Delta_{ji}(t) = \begin{cases} \eta^+ \Delta_{ji}(t-1) & g_{ji}(t)g_{ji}(t-1) > 0 \\ \eta^- \Delta_{ji}(t-1) & g_{ji}(t)g_{ji}(t-1) < 0 \\ \Delta_{ji}(t-1) & g_{ji}(t)g_{ji}(t-1) = 0 \end{cases} \quad (4.9)$$

ここで η^- と η^+ は学習係数であり、 $0 < \eta^- < 1 < \eta^+$ とする。

式(4.8)により1度に全ての結合係数を勾配の大きさに関係なく一定の量だけ変化させ、修正量は式(4.9)より算出する。流れとしては学習ごとのステップにおいて $g_{ji}(t)$ の符号が変化するたびにその修正量を小さく変化させて最終的には収束に導く。

しかしこの方法は細かい修正を行うことができないので結合係数の値が多い。つまりネットワークが大規模である程学習が進まない場合や複雑な値の出力、初期値の影響など大きく受けるのでそのままではオンライン学習として向かないとされている。

5 デジタル型誤差逆伝播法(DBP 法)

ネットワークにおいて中間層が連続的でない不可微分な 2 値出力素子で構成されており、上層からの誤差を中間層へ通して下層へ逆伝搬することができないので、ネットワーク全体は整っているとは言えなかった(中間層によるランダム要素が影響してしまうため)。そこで出力誤差を減少させるために中間層にも別に教師信号を与えることで、下層の結合係数修正を可能にする。ここで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として DBP 法の説明を示す。

5.1 出力素子が1個の場合

Fig. 5.1は出力層素子が1個かつ2値出力素子からなるデジタル型三層パーセプトロンの構成を示す。

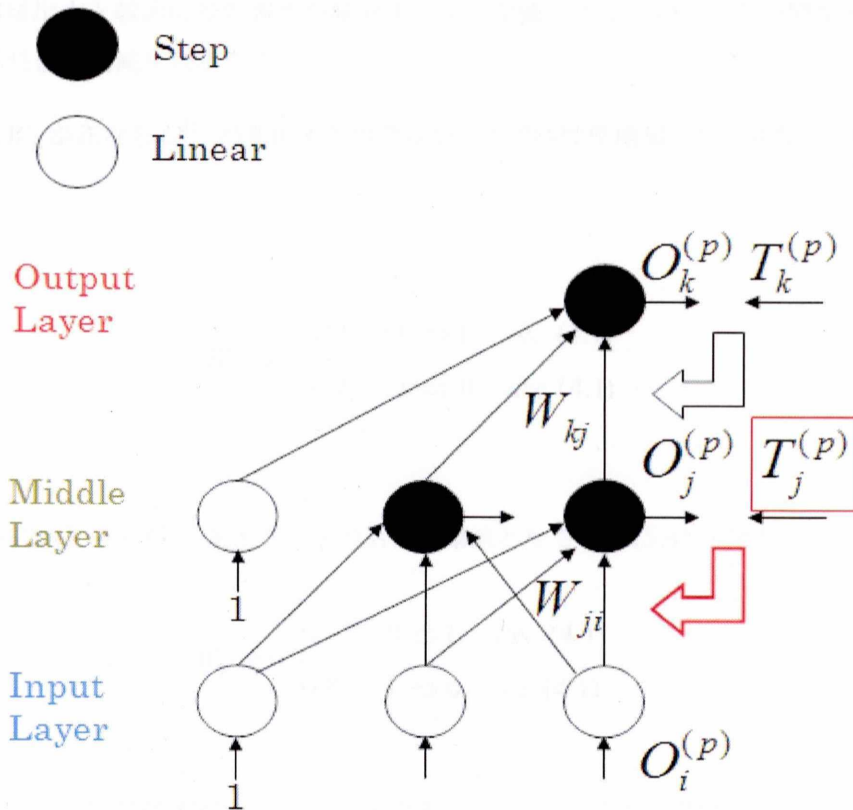


Fig. 5.1 DBP法の学習 (出力素子1個)

まず各入力パターン p (XORなら4)に対して,入力層から出力層へ順方向に出力 $O_j^{(p)}$ $O_k^{(p)}$ を計算し,出力層出力と教師信号の誤差を求める.次に上層結合係数をデルタ則に従い次式によって更新する.

$$W_{kj} = W_{kj} - \alpha(O_k^{(p)} - T_k^{(p)})O_j^{(p)} \quad (5.1)$$

ここでDBP法によって中間層に教師信号を与えることができるなら上層と同じ方法で結合係数更新をすることができるので次式のようになる.

$$W_{ji} = W_{ji} - \alpha(O_j^{(p)} - T_j^{(p)})O_j^{(p)} \quad (5.2)$$

そして目的の中間層素子教師信号 $T_j^{(p)}$ の求めるにおいて, 上層結合係数 W_{kj} の修正後, 中間層出力を出力層出力と教師信号の誤差が減少するような値にすることが望ましい. そのため式(1)の活性値が増減を利用する.

以下の式は W_{kj} が正なら $O_j^{(p)}$ が変化する場合 0 から 1 で活性値増加, 1 から 0 なら活性値減少を意味する.

$$W_{kj} > 0 \begin{cases} O_j^{(p)} : 0 \Rightarrow 1 & \text{inc (4.1)} \\ O_j^{(p)} : 1 \Rightarrow 0 & \text{dec (4.1)} \end{cases} \quad (5.3)$$

W_{kj} が負ならば (5.3) の $O_j^{(p)}$ 変化とは逆の活性値増減となるので (5.4) となる.

$$W_{kj} < 0 \begin{cases} O_j^{(p)} : 0 \Rightarrow 1 & \text{dec (4.1)} \\ O_j^{(p)} : 1 \Rightarrow 0 & \text{inc (4.1)} \end{cases} \quad (5.4)$$

また, 出力 $O_k^{(p)} = 0$, 教師 $T_k^{(p)} = 1$ ならば, 出力素子を活性化(活性値増加)させ, 1 になるように中間層教師信号 $T_j^{(p)}$ を (5.3) の条件から

$$\text{if } W_{kj} > 0 \text{ then } T_j^{(p)} = 1 \quad (5.5)$$

$$\text{if } W_{kj} < 0 \text{ then } T_j^{(p)} = 0 \quad (5.6)$$

として, 出力 $O_k^{(p)} = 1$, 教師 $T_k^{(p)} = 0$ ならば, 出力素子を非活性化(活性値減少)させ, 0 になるように中間層教師信号 $T_j^{(p)}$ を (5.4) の条件から

$$\text{if } W_{kj} > 0 \text{ then } T_j^{(p)} = 0 \quad (5.7)$$

$$\text{if } W_{kj} < 0 \text{ then } T_j^{(p)} = 1 \quad (5.8)$$

として、 $O_k^{(p)} = T_k^{(p)}$ ならば上層と下層の修正する必要はなく式 (5.1) はそのまま、 $W_{kj} = 0$ ならば活性の影響がないので修正効果はない。以上より (5.3) から (5.8) をまとめる

と、中間層教師信号 $T_j^{(p)}$ は

$$\text{if } (O_k^{(p)} - T_k^{(p)})W_{kj} > 0 \text{ then } T_j^{(p)} = 0 \quad (5.9)$$

$$\text{if } (O_k^{(p)} - T_k^{(p)})W_{kj} < 0 \text{ then } T_j^{(p)} = 1 \quad (5.10)$$

$$\text{if } (O_k^{(p)} - T_k^{(p)})W_{kj} = 0 \text{ then } T_j^{(p)} = O_j^{(p)} \quad (5.11)$$

となり、図 4 に示すようになる。

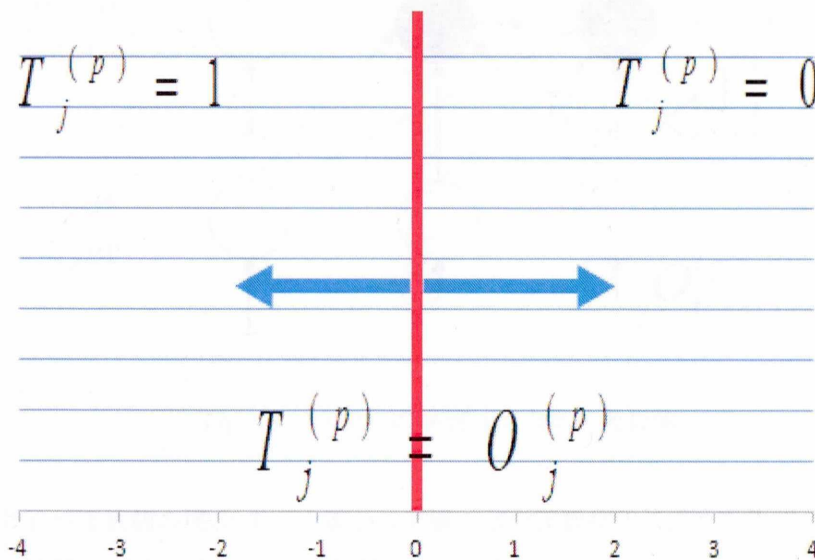


Fig. 5.2 DBP 法の教師決定因子による $T_j^{(p)}$

$(O_k^{(p)} - T_k^{(p)})W_{kj}$ は中間層教師信号を求める要因なので以降, 教師決定因子 $S_j^{(p)}$ (出力層素子 1) とする.

5.2 出力素子が複数個の場合

Fig.5.3 は出力層素子が複数個かつ2値出力素子からなるデジタル型三層パーセプトロンの構成を示す.

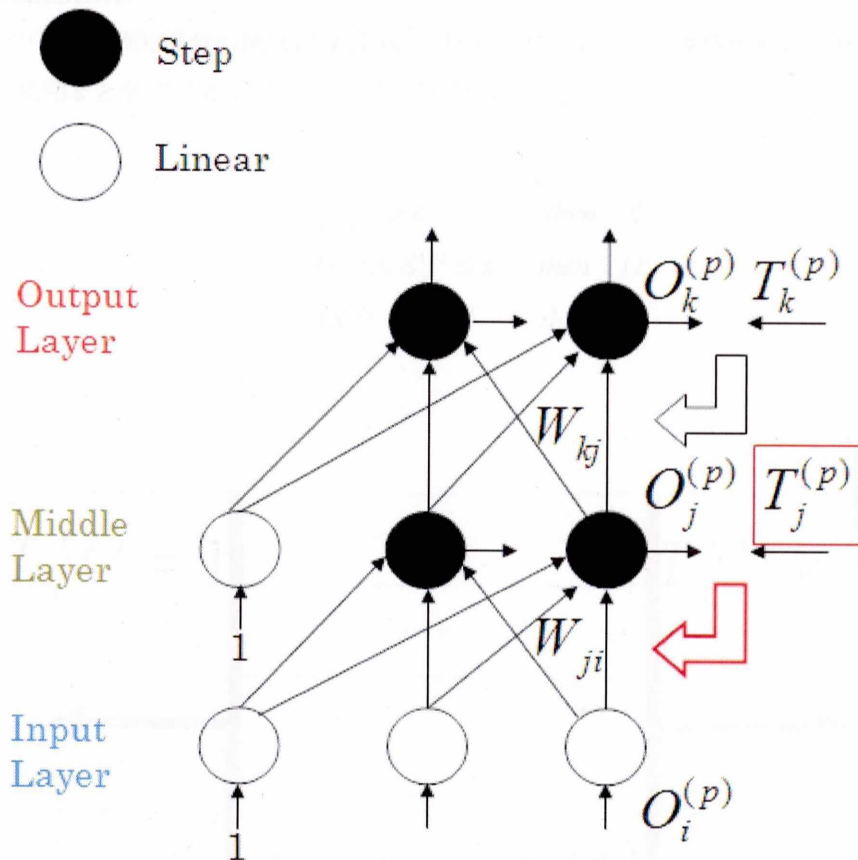


Fig. 5.3 DBP法の学習 (出力素子複数個)

出力層素子が複数個の場合, まず上層かつ下層の結合係数修正は, SP法及びDBP法の出力層素子1個と同様に(5.1), (5.2)のデルタ則に従い修正する. しかし, 出力層素子 k が複数ならば教師決定因子 $S_{jk}^{(p)}$ (出力層素子複数) も k 個になり, それぞれ符号が異なる場合は競合してしまう

ので, $S_{jk}^{(p)}$ の k に関する総和を求めたものを $S_j^{(p)}$ として符号競合解消をする.つまり

$$S_j^{(p)} = \sum_k (O_k^{(p)} - T_k^{(p)}) W_{kj} \quad (5.12)$$

となる.そして(5.12)式を用いて(5.9)から(5.11)式によって中間層教師信号 $T_j^{(p)}$ を求める.

5.3 許容範囲設定

DBP 法で中間層教師信号の決定に条件式が用いられており,この条件の範囲を広くしてその区間を許容範囲 ε と名づける.以下にその式と図を示す.

$$T_j^{(p)} = \begin{cases} S_j^{(p)} > \varepsilon & \text{then } 0 \\ -\varepsilon \leq S_j^{(p)} \leq \varepsilon & \text{then } O_j^{(p)} \\ S_j^{(p)} < -\varepsilon & \text{then } 1 \end{cases} \quad (5.13)$$

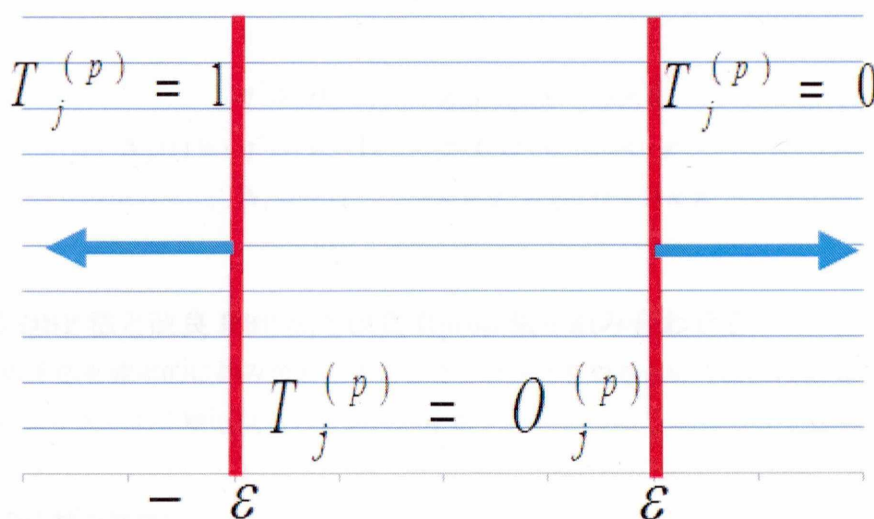


Fig. 5.4 許容範囲を取り入れた改良DBP法の教師決定因子による $T_j^{(p)}$

ここで $\varepsilon \geq 0$ とする.従来の DBP では初期値の影響がノイズとなり図 4 の赤い部分で中間層教師信号が振動してしまう事があり,理想的な学習を行えなくなり進まなくなる場合がある.

これは入力要素が多ければ多いほど発生確率も上がってしまう.

そこで許容範囲を設定することで図6の赤い区間なら下層の結合係数更新を行わないようにし、理想的出力付近の出力をある程度許容してしまうことで中間層教師信号の振動を防いで学習の停滞を解消し、学習成功の向上を図る。このDBP法を以降改良DBP法とする。また $\varepsilon = \infty$ であるならばSP法と同様で、入力層-中間層の下層結合係数が更新できないことを表し、 $\varepsilon = 0$ であるならばDBP法と同様で、許容範囲が無いものという意味を示す。この許容範囲は見てみると学習係数に近い概念であり、 ε の値を小さく設定すると0と1の出力が明確であるが、大規模ネットワークなど多数ニューロンによる結合係数などから教師信号の振動確率も高まってしまうので小規模な問題においては効果的である。また ε の値を大きく設定すると更新されない結合係数が増えるので学習は進みにくくなる可能性もあるが、更新される部分とされない部分の分類が大きく分かれていき、適切な教師信号の設定がしやすくなり理想的な出力ができやすくなる。

6 提案手法

6.1 弾性バックプロパゲーション(Rprop 法)に許容範囲を適用

改良として Rprop 法でも条件式(4.9)は DBP 法と同じように符号による条件式を用いているので、許容範囲概念を取り入れることで結合係数修正量を前回の更新値と同じ量のものを増やすことで偏りを減らす事ができると考えた。式は以下のようなになる

$$\Delta_{ji}(t) = \begin{cases} \eta^+ \Delta_{ji}(t-1) & g_{ji}(t)g_{ji}(t-1) > \varepsilon \\ \eta^- \Delta_{ji}(t-1) & g_{ji}(t)g_{ji}(t-1) < -\varepsilon \\ \Delta_{ji}(t-1) & -\varepsilon < g_{ji}(t)g_{ji}(t-1) < \varepsilon \end{cases} \quad (6.1)$$

6.2 通常 DBP 法と改良 DBP 法と改良 Rprop 法を組み合わせる

この3つの手法を確率的に組み合わせることでバランスを取るという概念に従って計算を行った。アルゴリズムの手順は以下のとおりになる。

1. 結合係数の初期化
2. 全パターンから出力 $O_k^{(p)}$ を求める
3. ΔW_{kj} を求め、式(3)より W_{kj} を計算

4-1 全パターンから $T_j^{(p)}$ を通常 DBP 法で計算

4-2 全パターンから $T_j^{(p)}$ を改良 DBP 法で計算

4-3 改良 Rprop 法で結合係数を一括修正

5. 4-1 か 4-2 で計算する場合, ΔW_{ji} を求め, 式(4)より W_{ji} を計算.

6. 出力層出力-教師信号間の誤差が無くなり, 正しい出力になるまで 2 から 5 を繰り返す.
もしくは設定した学習回数を越えた場合はその時点で学習終了となる, 学習失敗とする.
また 4-1, 4-2, 4-3 はどれか 1 つを行い, 5 に進む.

7 計算機実験

7.1 実験環境

シミュレーション実験に用いた計算機の環境を以下の表に示す.

CPU	Intel Core i7 2.80[GHz]
Memory	4.00[GB]
OS	Windows7 64bit Enterprise
Compiler	Visual C++ 2010 Express

7.2 実験環境

提案手法の有効性を検証するために計算機実験を行う.今回比較に用いた手法は従来手法であるデジタル型誤差逆伝播法(DBP 法),DBP 法に許容範囲の概念を取り入れ(改良 DBP 法),その発生確率を 10%に定めて残りは通常の DBP 法で計算する方法,提案手法としてさら Rprop 法に許容範囲概念を取り入れた方法を 10%の確率で発生させ,改良 DBP 法も 10%,残りは DBP 法で学習させるよう組み合わせた方法を用いた.

表 7.1 手法の発生確率

通常 DBP 法	80%
改良 DBP 法	10%
改良 Rprop 法	10%

以上の3つを線形分離不能問題である XOR 問題,半加算器,全加算器,4 入力加算器を対称に実験を行った.以下にこれらの特徴を示していく.

表 7.2 対象問題の性質

		中間層ニューロン数
XOR	2 入力 1 出力	2
半加算	2 入力 2 出力	2
全加算	3 入力 2 出力	3
4 入力加算	4 入力 2 出力	4

2 入力の線形分離不能問題について以下に示す.

表 7.3 XOR 問題の真理値表

入力 1	入力 2	出力&教師信号 1
0	0	0
0	1	1
1	0	1
1	1	0

表 7.4 半加算器問題の真理値表

入力 1	入力 2	出力&教師信号 1	出力&教師信号 2
0	0	0	0
0	1	0	1
1	0	0	1
1	1	1	0

3 入力の線形分離不能問題について以下に示す.

表 7.5 全加算器問題の真理値表

入力 1	入力 2	入力 3	出力&教師信号 1	出力&教師信号 2
0	0	0	0	0
0	0	1	0	1
0	1	0	0	1
0	1	1	1	0
1	0	0	0	1
1	0	1	1	0
1	1	0	1	0
1	1	1	1	1

4 入力の線形分離不能問題について以下に示す.

表 7.6 全加算器問題の真理値表

入力 1	入力 2	入力 3	入力 4	出力&教師信号 1	出力&教師信号 2
0	0	0	0	0	0
0	0	0	1	0	1
0	0	1	0	0	1
0	0	1	1	1	0
0	1	0	0	0	1
0	1	0	1	1	0
0	1	1	0	1	0
0	1	1	1	1	1
1	0	0	0	0	1
1	0	0	1	1	0
1	0	1	0	1	0
1	0	1	1	1	1
1	1	0	0	1	0
1	1	0	1	1	1
1	1	1	0	1	1
1	1	1	1	0	0

以下に各パラメータ設定値について示す.

表 7.7 各問題におけるパラメータ設定

α	0.002
ε	1
$\Delta_{j_i}(0), \Delta_{k_j}(0)$	0.1
η^+	1.2
η^-	0.5
最大学習回数	30000
改良 DBP 法導入確率	10 %
改良 Rprop 法導入確率	10 %

7.3 実験結果

これらの手法について,4 つの線形分離不能問題での学習成功率,及び,平均学習回数の結果を表 7.8, 7.9 に示す,

表 7.8 学習成功率の比較(%)

	DBP	改良 DBP	改良 DBP+改良 Rprop
XOR	80	83	82
半加算	99	100	100
全加算	97	100	100
4b 加算	83	84	89

表 7.9 平均学習回数の比較

	DBP	改良 DBP	改良 DBP+改良 Rprop
XOR	2106	1851	1138
半加算	3189	2996	2695
全加算	4813	4450	4208
4b 加算	7359	6983	6509

8 考察

DBP 法学習において許容範囲の設定や、一括学習における Rprop 法にも許容範囲の設定をして、これらを組み合わせることで、学習の成功率や回数の向上が見られた。

今回の改良 DBP で許容範囲を 10%の確率と、範囲を 1 に設定したのは参考文献 1) を引用した。これは離散的でも小規模なネットワークであるなら、少ない確率や許容範囲 を小さい値に設定しておけば、結合係数の更新を効率的に行えて、学習の停滞(学習回数増加)がしないことを表している。

逆に大規模ネットワークであるほど処理する量は増えるので外部入力誤差(ノイズ)も増加してしまい、教師決定因子の符号振動も増えてしまうことから、設定値を大きくする必要もある。Rprop 法は本来一括学習なので高速な学習は行えるが問題や初期値によって学習成功率がかなり変動する。また学習内で一括更新する時、ランダム要素が多いのと学習回数に対する効果も大きい初期が望ましいとは限らないので安定した学習を行うのは難しい。そのため今回の許容範囲は改良 DBP と同じ効果があると考えた。今回バッチ学習である Rprop 法を取り入れたのは、処理に制限が無い程度の問題に適応しやすいというのと DBP 法と同じように符号判定を行うという点では学習として同じ特徴を持つと考えたためである。学習を 1,2,3,4... と続ける場合と 1,2,(1,2,3,4...),3,4 のようにすれば 3 の学習時点で進んだ学習を行えるので高速化するという概念があった。万が一理想的な値から離れてしまっても許容範囲による設定で再学習を行い学習成功率にも影響が少ないようになるのではとも考えた。従来の Rprop 法は XOR や半加算等の小ネットワークを構成する場合であれば高速収束するが、要素が増えるとそれだけ学習に時間が掛かってしまい大規模に適用させるには難しいので他の方法と組み合わせる事でサポートできる傾向が強いだらう。

今回の確率的導入と許容範囲の設定は成功率を向上させるために行った。元々改良 Rprop 法を確率導入するので許容範囲まではそのまま確率的にする必要は少ない。

表 3 での改良 Rprop 法+改良 DBP 法の学習成功率は従来と大差はない。

しかし表 4 から学習回数を全般的に下げることになった。表 4 での入力数が少ない程 Rprop 法も効果が出るので組み合わせた手法でも XOR 問題が一番下がっている。

しかし許容範囲を広げすぎると学習は進まなくなり、Rprop 法も導入確率を増やすとそれだけ学習成功率が不安定になってしまう。

よって学習成功率と学習回数のバランスを保つために、提案手法での確率的に導入することは効果的である。

9 結論

本研究で, DBP 法において許容範囲を取り入れ, 一括学習である Rprop 法にも許容範囲の概念を取り入れて, 組み合わせた手法を提案した.

線形分離不能問題を対象とした計算機実験を行い, 従来手法と比較した結果, 全ての問題に対し平均学習回数の向上を確認したので提案手法の有効性を示すことができた. しかし学習成功率は問題によっては向上するが, 低下の可能性もあることが確認できた.

今後の展望としては DBP 法における設定値を最適に決定する方法や, 離散値を扱うシステム(ベクトル量子化など)に適用, 安定した収束を持つ条件を明確にすることやハードウェア化における実用性の確認等がある.

謝辞

本研究を進めるにおいて、法政大学工学部応用情報学科 李 磊 教授の指導のもとに行われたもので、同教授の丁寧なる御指導と御助言の結果、達成されたものであり、ここに深く感謝の意を表します。また本研究室諸氏の皆様にも様々な協力を頂き、深く感謝いたします。

最後に、常に温かく励まし応援をして下さった私の全ての友人と、肉体と精神を支えてくれた家族一同に対し、深い感謝の意をここに表したいと思います。

参考文献

- 1) 蔵淵 義光, 李 磊, “デジタル型 BP 法における教師信号の制限”, 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, No106(383), pp.21-26, 2006-11
- 2) 甘利俊一, 神経回路網の数理, 産業図書, 1978
- 3) 長沼 秀典, 大堀隆文, 渡辺一央, “微分不能素子を持つ階層型ニューラルネットに対する誤差逆伝播法の提案”, 電子情報通信学会論文誌, Vol, J88-D2, No10, pp,2125-2133, 2005
- 4) 生駒史, 丸山勇介, 長沼秀典, 大堀隆文, “デジタル型 BP によるベクトル量子化”, FIT(情報科学技術フォーラム一般講演論文集), 6(2), 507-508, 2007-8-22
- 5) Martin Riedmiller , Heinrich Braun, A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm CiteSeerX. [Online] (1993), <http://citeseerx.ist.psu.edu/viewdoc/versions?doi=10.1.1.21.1417>
- 6) 渡辺 雄二, “パーセプトロンにおける線形分離不可能問題に対する一解放” http://earth.iec.yamagata-cit.ac.jp/intro/watanabe/rr_2005.pdf