

# 最近の技術から

## 学習神経回路

平井 有三

筑波大学電子・情報工学系 〒305 つくば市天王台 1-1-1

### 1. はじめに

最近米国で、ニューラルネットに関する研究開発が爆発的に盛んになっている。その原因は、バックプロパゲーションと呼ばれる強力な学習アルゴリズム<sup>1)</sup>が提案されたことと、組合せ問題のように現代の計算機でも膨大な時間のかかる「計算」を、「ニューラルネット」で解くことができることが示された<sup>2,3)</sup>ことにある。本稿ではバックプロパゲーションに焦点を当て、その基本的な考え方を解説する。

### 2. バックプロパゲーション型学習神経回路

学習神経回路の歴史は、1950年代に提案されたパーセプトロン<sup>4)</sup>にその起源を求めることができる。パーセプトロンは入力された複数のパターンを、教師の指示に従って二つのクラスに分類することを学習することができる。しかしながら、入力パターンが線形分離でなければならぬことが大きなネックとなった。しかし、最近カリフォルニア大学の Rumelhart (ラメルハート、現在はスタンフォード大学)を中心とするグループが、パーセプトロンを改良したバックプロパゲーションと呼ばれる学習アルゴリズムを提案し、きわめて強力なアルゴリズムであることが認識された<sup>1)</sup>。

図1にバックプロパゲーション型学習回路例を示した。パーセプトロンがA層とR層の間の結合しか学習させることができなかったのに対して、彼らはR層で発生した誤差信号を入力側に伝播させるアルゴリズム(バックプロパゲーション)を提案することによって、S層とA層の間の結合をも学習させることが可能であることを示した。

パーセプトロンの場合、 $p$ 番目の入力パターンに対する素子  $A_j$  から素子  $R_k$  への結合係数  $W_{kj}$  の変更量は、

$$\Delta_p w_{kj} = c(T_{pk} - R_{pk})A_{pj} \quad (1)$$

のように定義される。ここで、 $c$ は正の比例定数、 $T_{pk}$ は  $p$ 番目の入力時に  $R_k$ に与えられる教師信号、 $R_{pk}$ は出力、 $A_{pj}$ は結合係数  $w_{kj}$ を介した  $R_k$ への入力である。

パーセプトロン型学習アルゴリズムは、次のように定義された誤差関数

$$E_p = (1/2) \sum_k (T_{pk} - R_{pk})^2 \quad (2)$$

を最小にする方向に、最急降下法によって計算していることに対応する。すなわち、

$$-\frac{\partial E_p}{\partial w_{kj}} = -\left(\frac{\partial E_p}{\partial R_{pk}}\right)\left(\frac{\partial R_{pk}}{\partial w_{kj}}\right) = -(T_{pk} - R_{pk})A_{pj} = -\eta_{pk}A_{pj} \quad (3)$$

となるからである。パーセプトロンの場合は、 $T_{pk} - R_{pk}$ そのものが誤差信号  $\eta_{pk}$ になる。バックプロパゲーションの場合の誤差信号は、式(3)を変形していくことによって得られる。そのためには、各素子の内部電位を

$$\text{net}_{pk} = \sum_j w_{kj}A_{pj} \quad (4)$$

のように、その出力を、

$$R_{pk} = f[\text{net}_{pk}] \quad (5)$$

のように表現する。ここで、関数  $f$ は微分可能でかつ非減少関数とする。シグモイド関数がよく使用されている。

さて、式(3)の左辺は、微分の連鎖則を用いると

$$\begin{aligned} \frac{\partial E_p}{\partial w_{kj}} &= \left(\frac{\partial E_p}{\partial \text{net}_{pk}}\right)\left(\frac{\partial \text{net}_{pk}}{\partial w_{kj}}\right) \\ &= \left(\frac{\partial E_p}{\partial \text{net}_{pk}}\right)A_{pj} \end{aligned} \quad (6)$$

となる。式(3)と(6)を比較すれば、

$$\eta_{pk} = -\frac{\partial E_p}{\partial \text{net}_{pk}} \quad (7)$$

が成り立たねばならない。さらに、式(7)の右辺は

$$= -\left(\frac{\partial E_p}{\partial R_{pk}}\right)\left(\frac{\partial R_{pk}}{\partial \text{net}_{pk}}\right)$$

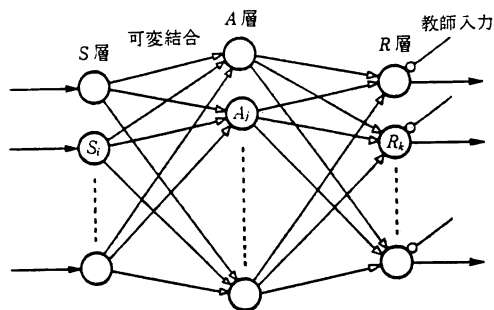


図 1 バックプロパゲーション型学習回路の構造

$$= - \left( \frac{\partial E_p}{\partial R_{pk}} \right) f'[\text{net}_{pk}] \quad (8)$$

と変形できる。

式(8)の  $\partial E_p / \partial R_{pk}$  は、 $R_{pk}$  が出力素子、すなわち教師信号を直接受け取っている素子の場合には、式(3)と同様に、

$$\eta_{pk} = -(T_{pk} - R_{pk}) f'[\text{net}_{pk}] \quad (9)$$

となる。さて、中間層の素子  $A_j$  の誤差をどう計算するかであるが、式(7)と同様に、

$$\eta_{pj} = - \frac{\partial E_p}{\partial \text{net}_{pj}} \quad (10)$$

とおく。式(8)と同様にして、

$$= - \left( \frac{\partial E_p}{\partial A_{pj}} \right) f'[\text{net}_{pj}] \quad (11)$$

が得られる。 $\partial E_p / \partial A_{pj}$  については、次のように計算する。

$$\begin{aligned} & \sum_k \left( \frac{\partial E_p}{\partial \text{net}_{pk}} \right) \left( \frac{\partial \text{net}_{pk}}{\partial A_{pj}} \right) \\ &= \sum_k \left( \frac{\partial E_p}{\partial \text{net}_{pk}} \right) \left( \frac{\partial \sum_i w_{ki} A_{pi}}{\partial A_{pj}} \right) \\ &= - \sum_k \eta_{pk} w_{kj} \quad (12) \end{aligned}$$

この式は、中間層の素子  $A_j$  が結合している出力層の各素子の誤差  $\eta_{pk}$  に結合係数を掛けて加えたものを、 $A_j$  の誤差信号とすることを意味している。中間層がいくつもある場合にも、各層で計算した誤差を前段に同様な方法で伝播させることにより学習が可能である。

このバックプロパゲーションアルゴリズムを用いて、Sejnowski は NETtalk というテキスト音声合成システムを提案した<sup>5)</sup>。このシステムは、最初は赤ん坊のようにしか発音できないが、学習が進むにつれて子供のように発音することができるようになる。そのほかに、曲面

検出や、ソナー音認識システムの構成にバックプロパゲーションを用いている。最近では音声認識<sup>6)</sup>や運動制御<sup>7)</sup>などにも盛んに応用されている。

### 3. バックプロパゲーションの問題点

しかしながら、バックプロパゲーションにも以下のような問題がある<sup>8)</sup>。

(1) いつもうまく学習できるとは限らない。

これは、誤差の2乗和で定義されたエネルギー関数の極小値が、一般には複数あることが原因となっている。隠れユニットの層が増えると深刻な問題となる。

(2) 追加ができない。

なぜならば、一つのことを覚えると、すでに覚えたことを壊してしまう破壊的な学習アルゴリズムだからである。したがって、学習にも時間がかかる。

(3) 位置ずれを吸収できない。

バックプロパゲーションでも、パターン認識で重要な位置の正規化はできない。

(4) わけのわからない問題がある。

隠れユニットの数や層をいくつにしたらよいのか、学習パターンの数をいくつにすればよいのか、上記(1)と(2)が複合して解析できていない問題がある。

### 文 献

- 1) D. Rumelhart, et al.: *Parallel Distributed Processing*, Vol. 1 & 2 (MIT Press, 1986).
- 2) J. J. Hopfield: "Neurons with graded response have collective computational properties like those of two-state neurons," *Proc. Natl. Acad. Sci. USA* (1984) 3088-3092.
- 3) J. J. Hopfield and D. W. Tank: "'Neural' computation of decisions in optimization problems," *Biol. Cybern.*, **52** (1985) 141-152.
- 4) F. Rosenblatt: *Principles of Neurodynamics* (Academic Press, 1962).
- 5) T. J. Sejnowski and C. R. Rosenberg: "NETtalk: a parallel network that learns to read aloud," *Neurocomputing*, eds. J. A. Anderson and E. Rosenfeld (MIT Press, 1988).
- 6) 住永洋子, 稲葉則夫: "ニューラルネットで音声認識は変わる", *日経エレクトロニクス*, 2月22日号, No. 441 (1988) 175-179.
- 7) H. Miyamoto, M. Kawato, T. Setoyama and R. Suzuki: "Feedback-error-learning neural network for trajectory control of a robotic manipulator," *Neural Networks*, **1** (1988) 251-265.
- 8) 平井有三: "ニューラルネットワーク", *電子情報通信学会技術研究報告*, ICD 88-124 (1988).

(1989年2月25日受理)