解　説

# ニューラルネットモデルのバックプロパゲーション学習について

田口　功

## 1．はじめに

中国語ではコンピュータのことを電脳と呼んでいる。コンピュータとは、電気的に構成された脳の働きを持つ装置と解釈すべきであろうか。情報処理の基礎研究はとどまるところをしらない。従来のノイマン型コンピュータは、人間の考えたプログラムに忠実に従って入力に対する出力処理を行なっていた。その処理において思考要素は全く含まれていない。それに対して、考えるコンピュータと言われる現在のニューロコンピュータの現状は、まだ初期の段階と考えられ、ノイマン型コンピュータを使って、プログラム的にニューロ構成を実現していると考えられ人間の脳に関する研究はまだ始まったばかりである。しかし、文字や図形などのパターン処理、音声分析、連続的なアナログデータの処理、ロボットの制御などに応用されている。本稿では、初期段階のニューロコンピュータを考えるうえで非常に基本的な概念となるバックプロパゲーションアルゴリズムについて述べる。

## 2．ニューロコンピュータを考えるうえでの脳の神経細胞構造とニューロ素子

人間の脳の中には、およそ140億個の神経細胞、すなわちニューロン（Neuron）が最小の機能単位として働いている。容積は1.4ℓ、さらに重さ1.4kgと言われている。この脳の働き、すなわち、目や耳からの情報によって筋肉を制御してさまざまな運動を行なったり、数学の問題を解いたり、常に脳は働き続けている。簡単に考えれば脳の働きとして、

(1) 計　算
(2) 推　論
(3) 制　御
(4) 認　知
(5) 創　造

などの5種類の機能に分類できると考えられる（参考文献(1)）。ここでニューロンの形態を簡単に考えることとする。普通の細胞と同様に細胞体の中心に細胞核がある。細胞体から延びる信号の入力部、出力部に特徴があり、入力部にあたる樹状突起さらに、自らの信号を次のニューロンに減衰することなく伝達するための伝送部にあたる軸索から構成されている。また、

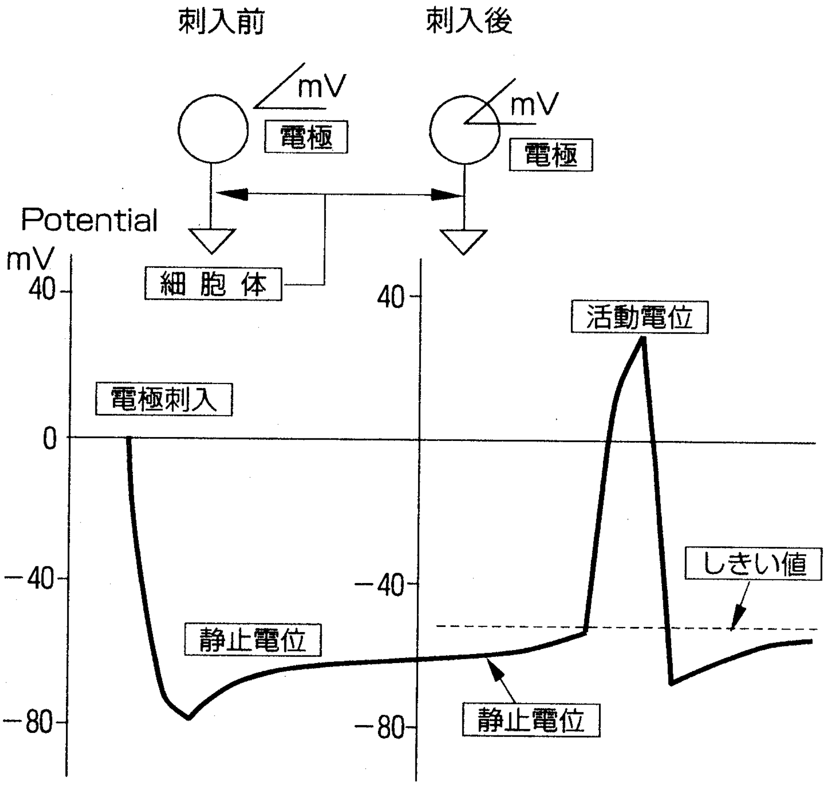


図１　ニューロンの細胞内電位および活動電位

軸索はその先端でシナプスと呼ばれる信号出力部になり、信号を受けるための樹状突起とつながっている。樹状突起を入力部、軸索を信号伝達ケーブル、シナプスを信号出力部と考えるとうまく対応がつく。

次に、アメフラシによるニューロンの動作について述べる。

ニューロンの内と外の間には、内外のイオン組成の違いにより電位差が発生する。これを電気化学的には分極という。ニューロンの外の電位を基準（０ボルト）にとり、内部（通常は負の電位であり静止電位と呼ばれ−40mv〜−60mv）の電位が他のニューロンからの出力および自らの変化によって、零ボルトのほうに変化したと仮定する。細胞内電位が静止電位から徐々に上昇し、ニューロンの持つ一定値（しきい値と呼ばれる）を超えると電位は急上昇しスパイク状の電位特性を生じる。ピーク最大値は20mv程度で再び細胞内の電位は静止電位に戻る。このスパイク状の電位変化は、いつも一定であり、ピーク最大値が変化したり、特性が変化してしまうことはない。常に一定の入力を超えると一定のスパイク状の電位を発生する。活動電位の発生は発火または興奮と呼ばれている（参考文献(1)）。電極を入れた場合の細胞内電位が変化して静止電位に落ち着く様子の電位波形およびその静止電位に対して、何かの理由で静止電位がプラス方向に変化し、ある値を超えた場合に突然発生する細胞体内の活動電位の波形（スパイク状の波形）を図１に示した。

細胞体で生じた活動電位は、軸索を伝搬する間減衰することはなくシナプスまで伝達される。シナプス部まで電気信号のままで伝わってきた活動電位はここで化学的信号に変換されてポスト側のニューロンに伝達される。シナプスに到達した活動電位によって、ポストニューロンの樹状突起に化学物質を放出する。この時、化学物質は、ポストニューロンの細胞内電位をプラスまたはマイナスに変化させる効果を持つ。活動電位を発生させるための基準電位（しきい値）は静止電位より高い電位となっているから電位をプラス方向に変化させる化学物質は、ポストニューロンを興奮しやすい状態に近づける。ポストニューロンを興奮しやすい状態にするシナプスを興奮性シナプスという。逆に、電位を興奮の方向から遠ざけるように働くシナプスを抑制性シナプスという。以上の説明を簡単にまとめるとニューロンは、しきい値を

持った非線形多入力1出力素子であることになる。これを、図で表現すると、

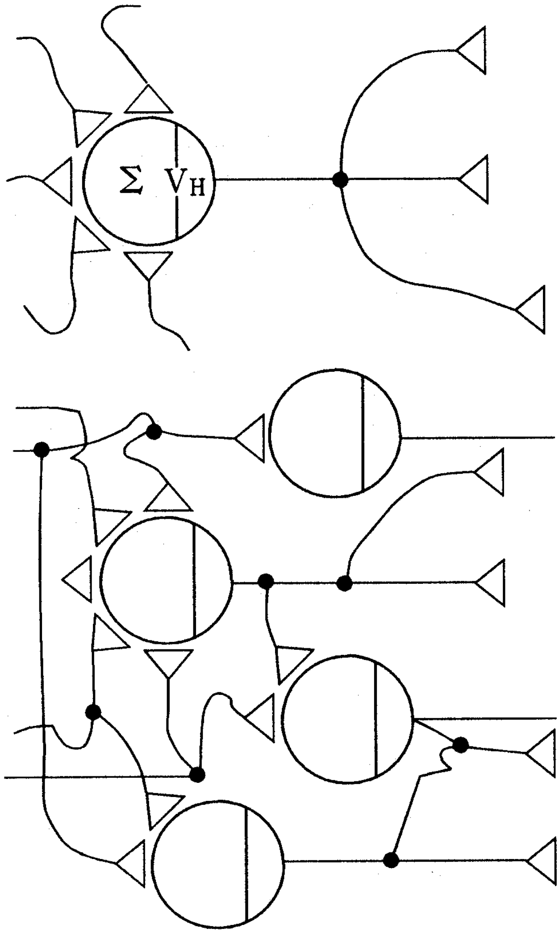


図2　多入力1出力素子としてのニューロンおよびニューロネットワーク

となる。この素子の集まり（種々のネットワークの考え方がある）を考えて一つのシステムを構成することになる。ここで円は細胞体、三角形はシナプス、その間の直線および曲線は軸索を表わすものとする。

## 3．バックプロパゲーション（逆伝搬学習則）アルゴリズム

1989年ルメルハルト（Rumelhart）らによって、提案された逆伝搬学習則は学習機能を持つことで一躍注目をあびることになった。本節においては、入力層、中間層（一層以上)、出力層からなるモデルを考え、多層パーセプトロン重み学習アルゴリズムについて考察する。特に、図3に示すようなM層多層パーセプトロンモデルについて考察し、学習によって決定される重み係数$W_{ij}^{(m)}$は、評価関数を最小にする極小値探索手法であることを述べる。また、重み係数$W_{ij}^{(m)}$の決定は、M層多層パーセプトロンが、入力に対するパターンを学習したということを強調したい。

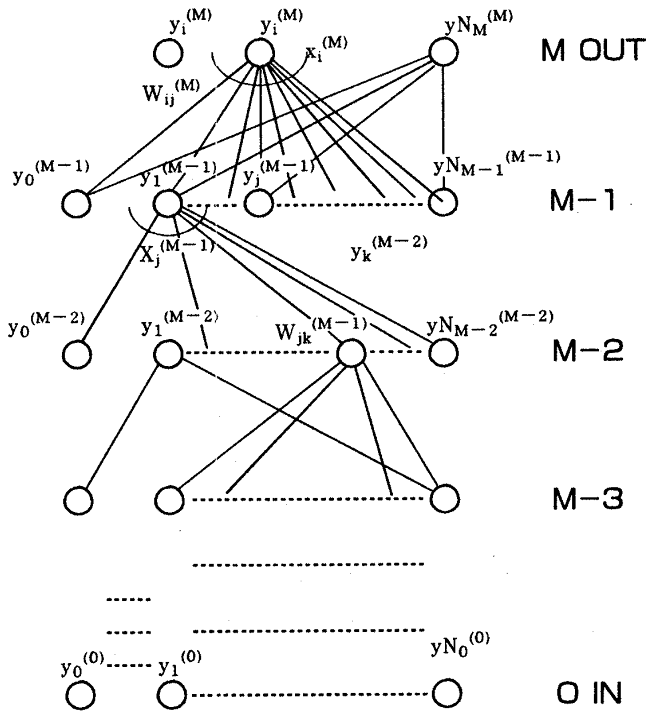


図3　M層ニューラルネットワーク

図3において、$y_i^{(m)}$は、第m層（m＝1，…，M）i番目のニューロユニットの出力を示し、$x_i^{(m)}$は、m層i番目のニューロユニットの入力の総和を表わすものとする。すなわち、

$$x_i^{(m)} = \sum_j W_{ij}^{(m)} y_j^{(m-1)} \qquad (3\text{-}1)$$

となる。ただし、$W_{ij}^{(m)}$はm－1層のj番目のユニットからm層のi番目のユニットへの重み

係数、$y_j^{(m-1)}$ はm－１層のｊ番目のユニットの出力を表わすものとする。ただし、ｊに関する総和はｍ－１層のすべてのユニット出力に対して取るものとする。また、各層mに常に１を出力する特殊なユニット $y_0^{(m-1)}$ を設け、$-Wi_o^{(m)}$によって第m層ｉ番目のユニットのしきい値 $\theta_i^{(m)}$を表現する。しきい値が異なる場合には当然 $Wi_o^{(m)}$の値は異なる。なお、ここでは、表現を簡単化するために、時間パラメータを省略した。$x_i^{(m)}$ と $y_i^{(m)}$ は非線形変換Ｆによって

$$y_i^{(m)} = F〔x_i^{(m)}〕 \tag{3-2}$$

のように表現する（一般的にはプログラムの簡単化を考えてシグモナイド関数を使用する場合が多い）。

次に、学習パターンとして一連の入力に対して一連の期待出力のペアが用意されているとする。そして、任意の入力に対して、ネットワークの出力が期待出力に近づくようにネットの重み係数を逐次学習するアルゴリズムを考える。最初に学習パターンとして、１組の入力 $y_i^{(0)}$（$i=1, \cdots, N_0$）と期待出力 $d_i$（$i=1, \cdots, N_M$）が与えられる場合について考える。ここで、$N_m$（$m=0, \cdots, M$）はm層のニューロユニット数を表わす。すべての重み係数がなんらかの値を初期値として持つものとすれば（一般的には乱数を用いてランダムに与えられる）学習入力データが入力層に与えられると、出力 $y_i^{(M)}$（$i=1, \cdots, N_M$）は、（３－１）式および（３－２）式に従って計算される。出力は、期待出力とはかなり異なった値となってしまうと考えられる（この段階では評価関数を最小にするような重み係数の更新は行なわれていない）。そこで期待出力とネットの実際の出力の２乗誤差の総和を評価関数として

$$E=(1/2)\sum_{i=1}^{N_M}(y_i^{(M)}-d_i)^2 \tag{3-3}$$

を定義する。この評価関数の値を最小にするため重み係数をどのように調整するかの極小値探索法について以下に述べる。評価関数Ｅのネットに含まれるすべての重み係数に関する偏微分 $\partial E/\partial W_{ij}^{(m)}$($m=1, \cdots, M; i=1, \cdots, N_m; j=0, \cdots, N_{m-1}$) を求める。偏微分値が正の値をとるとした場合、重み $W_{ij}^{(m)}$ を少し増加すると誤差はＥは増加することになる。重みを減少すれば誤差Ｅは減少するような重み計算が必要となる。逆に、負の場合には$W_{ij}^{(m)}$を増加した場合誤差Ｅは減少することになる。従って、正負両方の場合を考慮しＥを減少させるためにパラメーター $\eta$を用いて、重み係数計算

$$W_{ij}^{(m)}〔n+1〕=W_{ij}^{(m)}〔n〕-\eta(\partial E/\partial W_{ij}^{(m)}) \tag{3-4}$$

を行なう。この方法を最急降下法という。また、(3-4) 式の右辺に一学習サイクル前の重み係数変化に比例した項

$$\alpha(W_{ij}^{(m)}〔n〕-W_{ij}^{(m)}〔n-1〕) \quad ;0<\alpha<1 \tag{3-5}$$

を (3-4) 式に加えると

$$W_{ij}^{(m)}〔n+1〕=W_{ij}^{(m)}〔n〕-\eta(\partial E/\partial W_{ij}^{(m)})+\alpha(W_{ij}^{(m)}〔n〕-W_{ij}^{(m)}〔n-1〕) \tag{3-6}$$

となる。これは、収束速度を速めるのに役立つ。

次に、一般的によく使用されているシグモイド関数を使用して（3-6）式を具体的に評価することを考える。出力層ユニットに対しては直接（3-3）式を使用して、

$$\partial E/\partial y_i^{(M)} = (1/2)\,2\,(y_i^{(M)} - d_i) = y_i^{(M)} - d_i\ ;\ i=1, \cdots, N_M \quad (3\text{-}7)$$

となる。また、$\partial E/\partial x_i^{(M)}$を計算すると、

$$\partial E/\partial x_i^{(M)} = (\partial E/\partial y_i^{(M)}).\,(\partial y_i^{(M)}/\partial x_i^{(M)}) \quad (3\text{-}8)$$

の関係が成立するから、シグモイド関数

$$F[x] = 1/(1+e^{(-x)}) \quad (3\text{-}9)$$

を適用すると、微分$F'[x]$は、

$$\begin{aligned} F'[x] &= \{0-(-1)e^{(-x)}\} / \{1+e^{(-x)}\}^2 \\ &= \{1/(1+e^{(-x)})\} \cdot \{(1+e^{(-x)}-1)/(1+e^{(-x)})\} \end{aligned}$$

と変形することができるから、

$$F'[x] = F[x] \cdot (1-F[x]) \quad (3\text{-}10)$$

の関係があるので、

（3-7）式、および（3-10）式の関係を（3-8）式に代入すると、

$$\partial E/\partial x_i^{(M)} = (y_i^{(M)} - d_i)\, y_i^{(M)} (1-y_i^{(M)}) \quad (3\text{-}11)$$

となる。従って、目的とする偏導関数$\partial E/\partial W_{ij}^{(M)}$は、

$$\partial E/\partial W_{ij}^{(M)} = (\partial E/\partial x_i^{(M)}) \cdot (\partial x_i^{(M)}/\partial W_{ij}^{(M)}) \quad (3\text{-}12)$$

となる。従って、（3-1）式より

$$\partial x_i^{(M)}/\partial W_{ij}^{(M)} = y_j^{(M-1)} \quad (3\text{-}13)$$

の関係が容易に求まるから、（3-12）式に（3-11）式および（3-13）式の関係を代入すると、

$$\partial E/\partial W_{ij}^{(M)} = (y_i^{(M)} - d_i)\, y_i^{(M)} (1-y_i^{(M)})\, y_j^{(M-1)} \quad (3\text{-}14)$$

となる。（3-14）式は期待出力$d_i$、および出力$y_i^{(M)}$および$y_j^{(M-1)}$からなっているので計算可能である。

次に、出力層以外の中間層ユニットへの重み係数に関する評価関数Eの偏微分を求める。最初にM－2層のユニットから出力層の一段下のM-1層のユニットへの重みに関する偏微分$\partial E/W_{jk}^{(M-1)}$ $(j=1, \cdots, N_{M-1}\ ;\ k=0, \cdots, N_{M-2})$について考ええることとする。ここで、出力層（M層）、M-1層、M-2層のユニットを示す添え字をｉ、ｊ、ｋとおく。評価関数Eの出力層への重みM-1層による偏微分を求めた時と同様にEのM-1層のｊ番目のユニット出力$y_j^{(M-1)}$に関する偏微分を求める。ここでは、期待出力が使用できないことに注意する。従って、

$$\begin{aligned} \partial E/\partial y_j^{(M-1)} &= \sum_{i=1}^{N_M} (\partial E/\partial x_i^{(M)}) \cdot (\partial x_i^{(M)}/\partial y_j^{(M-1)}) \\ &= \sum (\partial E/\partial x_i^{(M)})(W_{ij}^{(M)}) \\ &= \sum \{(y_i^{(M)} - d_i) y_i^{(M)} (1-y_i^{(M)})\} \cdot (W_{ij}^{(M)}) \end{aligned} \quad (3\text{-}15)$$

となる。（3-15）式における$\partial E/\partial x_i^{(M)}$の値は、（3-11）式の値を用いることが出来る（$i=1, \cdots, N_M$）。さらに、シグナイド関数を使用する

と$\partial E/\partial x_j^{(M-1)}$は、

$$\partial E/\partial x_j^{(M-1)} = (\partial E/\partial y_j^{(M-1)})(\partial y_j^{(M-1)}/\partial x_j^{(M-1)}) = (\partial E/\partial y_j^{(M-1)})(y_j^{(M-1)})(1-y_j^{(M-1)}) \qquad (3\text{-}16)$$

となる。目的とするＥの重み係数$W_{jk}^{(M-1)}$に関する偏微分($j=1,\cdots,N_{M-1}$ ; $k=0,\cdots,N_{M-2}$)は、(3-17) 式となる。

$$\partial E/\partial W_{jk}^{(M-1)} = (\partial E/\partial x_j^{(M-1)})(\partial x_j^{(M-1)}/\partial W_{jk}^{(M-1)}) = (\partial E/\partial x_j^{(M-1)})y_k^{(M-2)} \qquad (3\text{-}17)$$

従って、(3-17) 式に (3-16) 式を代入し、さらに (3-15) 式の関係を代入すると (3-17) 式の関係は、

$$\begin{aligned}\partial E/\partial W_{jk}^{(M-1)} &= (\partial E/\partial y_j^{(M-1)}).(y_j^{(M-1)})(1-y_j^{(M-1)})(y_k^{(M-2)}) \\ &= \sum_{i=1}^{N_M}\{(y_i^{(M)}-d_i)y_i^{(M)}(1-y_i^{(M)})\}(W_{ij}^{(M)})(y_j^{(M-1)})(1-y_j^{(M-1)})(y_k^{(M-2)}) \qquad (3\text{-}18)\end{aligned}$$

となる。同様に、Ｅの各層ｍ（$m=M-1,\cdots,1$）のユニットへの重みに関する偏微分を入力層方向に計算することによって順次計算を行なう。さらに (3-6) 式を使用し重みを更新し、ネットワーク係数をＥが最小にするように計算を繰り返すことによってネットワーク係数を決定することが出来る。

次に学習データが複数ある一般的な場合について考える。学習データとしてＬ組の入力$y_{i,\ell}^{(0)}$（$i=1,\cdots,No$ ; $\ell=1,\cdots,L$）と期待出力$d_{i,\ell}$（$i=1,\cdots,L$）が与えられたとする。期待出力とネットの実際の出力の２乗誤差の総和は、

$$E = 1/2\sum_{\ell=1}^{L}\sum_{i=1}^{N_M}(y_{i,\ell}^{(M)}-d_{i,\ell})^2 = 1/2\sum_{\ell=1}^{L}E_L \qquad (3\text{-}19)$$

となる。この時、Ｅの各層の重み係数に関する偏微分$\partial E/\partial W_{ij}^{(m)}$（$m=1,\cdots,M$）は次式の各学習データに対する偏微分の和として

$$\partial E/\partial W_{ij}^{(m)} = \sum_{\ell=1}^{L}\partial E_\ell/\partial W_{ij}^{(m)} \qquad (3\text{-}20)$$

となる。$\partial E/\partial W_{ij}^{(m)}$の値は今まで述べた方法によって計算できるので、各学習データに対してこれを計算し、全学習データにわたって総和を求め (3-4) 式を適用し重み係数を更新することによってＥを減少させることができる。もう一つの方法として (3-3 ) 式を用い順次学習データをネットに提示し、重み係数を更新する。さらに、各学習データに対するＥがすべて十分小さくなるまで繰り返すことによって最小の重み係数学習を行なうことができる。

## 5．ニューロコンピュータ研究の現状について

ニューロコンピューティングモデルにはポップフィールドモデルやバックプロパゲーションモデルがある。本稿では学習機能を取り入れたバックプロパゲーションモデルの原理を述べた。このバックプロパゲーションモデルはニューロンの数を人間が決め、シナプス強度は教師信号

を与えることによって、自らエネルギー最小になるように決定される。このことについては第4節で詳しく述べた。このバックプロパゲーションモデルは手書き文字認識にも応用されている。

ここで、ニューロコンピューティングの現状についてここで考える。本稿においては最も基本となるバックプロパゲーションアルゴリズムを取り上げたがこのアルゴリズム自体においても評価関数Eが大域的な最小値に収束しない可能性があり、この問題を解決するための決定的な手法は現在では知られていない。

ニューロコンピュータは、幅広い研究分野の人々を巻き込んで研究がおこなわれ人間の脳神経系を直接研究する分野もある。なめくじに似たアメフラシの研究から始まり人間の脳の構造を明らかにし、脳神経系の情報処理メカニズムを研究しようというわけである。さらに、神経情報処理原理の数理的解明やニューロコンピューティングモデルの構築を研究する分野もある。

工学的には、デバイス的なニューロチップの開発問題、ニューロチップを応用した情報処理マシーンの開発と応用、ロボット制御、システム同定問題などに応用されつつある。

## 6．おわりに

本稿では、主に文献(1)、(2)を参考にし、ニューロコンピュータを考える上で基礎となる学習機能を持つバックプロパゲーションアルゴリズムについて述べた。このアルゴリズムを使用した応用としては、手書き文字認識などがある。しかし、このバックプロパゲーションアルゴリズム自体にも問題があり、エネルギー最小値（大域的な最小値）に結合係数をセットするための確実な、しかも簡単な方法はまだ確立されていない。現在、中間層のユニット数を増やしたり、確率的に時々エネルギーを増大する方向に意図的に重み修正を行なう、などの方法がとられている。この問題を解決できれば、バックプロパゲーションアルゴリズムはもっと効率よく多方面に応用されるものと考えられる。

## 7．参考文献

(1) 八名和夫・鈴木義武共著：ニューロ情報処理技術、海文堂、pp.42-49、1992年
(2) 塩野　悟著：ニューロコンピュータへの挑戦（アメフラシから電脳を学ぶ）、アグネ承風社、pp.1-2、pp.10-18、1990年
(3) 菊池豊彦著：入門ニューロコンピュータ、オーム社、1990年