ニューラルネットの学習モデル

松岡 清利*

脳における学習のメカニズムは、生理学や心理学の分野のみならず、工学的にも興味深い研究対象である。ここでは、人工ニューラルネットの分野で考えられている、いくつかの基本的な学習モデルについて述べてみたい。

1. はじめに

生体の諸機能、特に脳のそれは、生理学者・生物学者・心理学者に限らず工学者にとっても極めて興味深い対象である。実際、半世紀近くにわたる人工知能の研究の歴史において、人の脳に似た(あわよくばそれを越える)機械の実現は常に目指すべき目標であった。

人工知能の研究には、2つの大きな流れがある。1つは、人が心に抱く概念や表に現れる行動などの個々の事象を記号で表し、人の知的行動をそれらの組み合わせや操作として説明するという方法論である。比較的最近まで、そのようなアプローチが人工知能学の主流であったと言え、実際「人工知能」という言葉は、狭義には、記号操作に基づく知能機械の実現を目指す研究に対して使われることが多い。

一方、これに対して、実際の生物の神経回路に多少なりとも似た人工的な神経回路を作り、それに何らかの情報処理をさせる試みが、10年ほど前から活発に行われるようになってきた。そのような考えのもとに作られた(あるいは計算機上で模擬された)神経回路を(人工)ニューラルネット、そこでなされる情報処理をニューロコンピューティングと呼んでいる。

このようなアプローチが注目されるようになった背景には、人の知的情報処理の中には記号操作としてはうまく記述しえない直感的な処理(例えば視覚系におけるパターン認識など)が多くあることが認

識され、それらに対する説明モデル(あるいは応用のためのモデル)として、ニューラルネットが適しているということがあろう。さらには、知能を、記号操作のような抽象的なレベルではなく、神経活動のレベルで説明してみたいという、いわば還元主義的な欲求が強まったことが挙げられようか。

脳を1つの情報処理機械と捉え数理的に探求しようという研究は、1940年代初めのマカロック(MuCullock, W. S.)とピッツ(Pitts, W.)の研究にまで遡ることができる。彼らは、ニューロンを線形閾値素子と呼ばれる一種の論理素子と見なし、それらを組み合わせると、どのような計算機械も実現できることを示した。もっとも、これはごく原理的な話で、単に、「ICを組み合わせればどんな計算機械を作ることもできる」という程度のことを言っているに過ぎないとも言える。与えられた機能を実現するのに、ICをどのように組み合わせればよいかを示されなければ、面白い話とは言えないであろう。しかし、マカロックとピッツの論文の歴史的意義は大きく、現在のニューラルネットの諸モデルも、基本的には、彼らの提案したモデルの延長線上にあると言える。

1950年代は脳、コンピュータを含めて、知的機能をどう解明したらよいかを模索した時代であったが、1960年頃にローゼンブラット(Rosenblatt, F.)が、パーセプトロンと呼ばれる学習モデルを提案した。ニューロン間の結合の強さ(シナプス効率)を、ある簡単な規則に従って変化させると、パターン分類器が自動的に形成されるというものである。このモデルは1960年代に多くの研究者の興味を引き、当時、これで生物の学習の根本的なカラクリは明らかになったと言う(今から考えれば馬鹿げた)人さえいたそうである。

しかし、1970年代になると、いくつかの理由から

*Kiyotoshi MATSUOKA, 九州工業大学工学部制御工学教室(〒804 北九州市戸畑区仙水町1-1)

図1 ニューロン・モデル

図2 出力関数。(a)シグモイド関数、(b)線形関数。

ニューラルネットの研究は下火になった。1つの理由は、当時の計算機の性能が低かったために、計算機上でニューラルネットを模擬しても実行に非常に時間がかかり、実用的な応用にまで手が届かなかったことが挙げられよう。また、パーセプトロンの限界が理論的に明らかにされ、その結果、研究者の興味がそがれたこともある。再び脳の工学的研究が盛んになってきたのは1980年代の半ばになってからで、新しい重要なモデルが数多く出現し、今日に至っている。

パーセプトロンと同様に現在のニューラルネットの研究においても、学習ないし自己組織化という生物に特徴的な機能が中心的なテーマとなっている。ここで、ニューラルネットの分野で使われる‘学習’という言葉について述べておきたい。一般に、ニューラルネットは外界から入力(刺激)を受け、それに対して何らかの演算をし、出力(反応)を出す。はじめは、ニューラルネットのネットワーク構造やニューロン間の結合の強さ(結合荷重)は勝手に与えられている。この状態では、与えられた入力に対する出力は、もちろん機能的に意味のないものである。そこで、何らかの規則に従ってネットワーク構造や結合荷重を変え、ニューラルネットに意味のある出力を出させるようにしたい。多くのモデルでは、構造や結合荷重を少しずつ変えていくスキームがとられ、そのプロセスを‘学習’と呼んでいるわけである。

学習に関して、2つのパラダイムが考えられている。1つは、管理学習(教師あり学習)と呼ばれるもので、内部をも含めてニューラルネットの動作の一

部始終を観察し、それに指示を与える‘教師’の存在を想定する。前述のパーセプトロンは、その最も単純な例である。学習規則以外に、教師によってニューラルネットの‘望ましい出力’が与えられ、実際の出力と望ましい出力の差が小さくなるように学習が進行する。これに対して、非管理学習(教師なし学習)と呼ばれるパラダイムがあり、そこでは教師が与える‘望ましい出力’なるものではなく、学習規則のみが与えられる。教師が存在しないのになぜ学習が可能なのかという疑問が生じるかもしれないが、加えられる入力に内包する特徴そのものが機能形成に寄与するわけで、いわば、入力が教師になるわけである。

本解説では、初めに管理学習について述べ、次に非管理学習について説明する。教師の有無によって、取り扱いがかなり異なってくることが理解されるであろう。なお、以下で若干の数式や統計学的概念が現れるが、それ抜きでは実のない哲学論議になると思ったからで、ご容赦願いたい。

2. 人工ニューラルネットのモデル

学習について述べる前に、神経回路の人工モデルについて説明しておこう。実際のニューロンの振舞いは、細かく見れば複雑であるが、取り扱いを簡単にするために、極めて簡単な演算素子としてモデル化される。工学者という人種は目的志向が強く、本質的でないと思われるものは強引に切り捨てる傾向があるが、この場合もそうである。

図1にニューロン・モデルが示されている。図中

図3 階層形ネットワーク(a)と全結合形ネットワーク(b)

の○が言わば細胞体で、そこに N 個の入力 $S_1, \dots, S_N$ が入っている。それらに荷重 $w_1, \dots, w_N$ が乗算され、合計される。

$$x = w_1 S_1 + \dots + w_N S_N \quad (1)$$

x は膜電位である。 $w_1, \dots, w_N$ はいわゆるシナプス効率で、値が正の場合は興奮性シナプスの、負の場合には抑制性シナプスの荷重を表す(以後、シナプス効率を結合荷重と呼ぶ)。ここで、実際のニューロンに近いモデルを考えようとするならば、膜電位 x がある閾値を超えたときインパルスが1つ出力されたいが、特別な場合を除いて、そのような面倒なモデルは考えない。 x に非線形関数 g を施したものが出力されるとする。すなわち、ニューロンの出力は

$$y = g(x) \quad (2)$$

で与えられる。強いというなら、 y は軸索に出力されるパルスの‘頻度’である。関数 $g(x)$ としては、図2(a)のようなシグモイド形の関数が多く用いられる。

学習のプロセスを数学的に解析するような場合には、さらに単純化して $g(x) = x$ 、すなわちニューロンを線形要素としてモデル化することもある(図2(b))。後の非管理学習の説明においては線形モデル

を仮定している。線形ニューロンは現実のニューロンとはかなり異なったものであり、また応用上もニューラルネットが持ちうる機能を著しく制限することになるが、非線形関数の導入は数学的取り扱いを困難にするので、仕方なくそうするのである。

このような人工ニューロンを組み合わせ、ネットワークを構成するわけであるが、ここでも実際の生物の神経回路などにはあまり頓着しない。たとえば図3に示すような(生理学者からみれば極端に一樣な)ネットワーク構造がしばしば採用される。(a)は階層形のネットワークで、信号は左から右に1方向に流れる。帰還ループを含まないので、数学的には取り扱いやすい。(b)の場合にはすべてのニューロンの間に相互に結合がある。なお、この場合、一部の結合の荷重を0とすることによって、任意のネットワークが表現できることを注意しておこう。

3. 管理学習(教師付き学習)

管理学習のモデルとして最もよく知られているものは、誤差逆伝搬学習である。対象とするネットワークの構造は、何でもかまわないが、通常、図3(a)の階層形ネットワークが採用される。図4の四角で囲まれた部分がニューラルネットである(図では、ニューロン j からニューロン i への結合のみが書かれている)。 s は入力(一般に入力は1つではないのでベクトル)で、 y は出力(ベクトル)である。

図4 管理学習

例えば、手書き文字を認識をするニューラルネットの場合には、TV カメラで得られた画像を適当に数値化してベクトルを作り、それをニューラルネットに入力する。もし、どのように書かれた文字‘あ’に対しても出力 y が(1, 0, 0, 0,...), どのような‘い’に対しても出力 y が(0, 1, 0, 0,...)等となるなら、このネットワークは手書き文字を認識をする装置として使えよう。

さて、入力 s が与えられたとき、このニューラルネットに出力してほしい目標出力 y^d がわかっているとす。例えば、入力が‘あ’なら、 y^d は(1, 0, 0, 0,...)である(教師はそれを知っている)。初めは、結合荷重 w_{ij} には勝手な値があてがわれているので、実際の出力 y は y^d にはならない。誤差逆伝搬学習においては、ネットワークの出力が目標出力に近づくように、ニューロン間の結合荷重を少しずつ修正する。

そのために、まず、ネットワークの実際の出力 y と目標出力 y^d の差を表す誤差を定義する。

$$E = |y - y^d|^2 \quad (3)$$

誤差は、結合荷重 w_{ij} 等の関数である。誤差関数 E が小さく(願わくば0に)なれば、出力は目的の値に近いわけであるから、その最小化を考えればよい。それには最急降下法と呼ばれる数学手法が利用できる。すなわち、誤差関数を結合荷重で微分して、それに比例した量だけ結合荷重を減少(負の方向に増大)させればよい。

$$\Delta w_{ij} \propto -\frac{\partial E}{\partial w_{ij}} \quad (4)$$

結合荷重の修正は、入力を変更することに行う。(4)の微分計算は、まともに行うと面倒であるが、少し工夫すると極めて簡単になる。そこでは、すべてのニューロンに対して、逆伝搬誤差という量が導入さ

れる。この量は、出力誤差が小さくなるためには各ニューロンの膜電位 x がどれだけ変化すればよいかを示すもので、ネットワークを逆向きにたどって計算される。ニューロン j からニューロン i への結合荷重 w_{ij} に対する修正量は、ニューロン i に対する逆伝搬誤差とニューロン j の出力の積として与えられる。

この学習モデルは1985年頃に提案されたものである。今になってみると馬鹿馬鹿しいほど単純な発想に基づくものであるが、応用面では最も多用される学習モデルである。その理由は、以下の3点にある。

- (i)与えられた入出力関係を持つニューラルネットワークを作るという形で定式化される工学的問題は非常に多い。
- (ii)アルゴリズムが簡単である。
- (iii)このモデルを実際の問題に適用してみると結構期待した結果が得られる。

しかし生理学的に見るなら、このモデルが自然でないのは、1つの結合荷重 w_{ij} をどう変えればよいかは、それに直接関与している2つのニューロンの状態値(y_i と y_j)のみからは決定できず、逆伝搬誤差も必要になることである。例えば、図3(a)の階層形ネットワークにおいて誤差逆伝搬学習を実現しようとする、図の神経経路以外に、逆伝搬誤差を伝える逆向きの情報伝達経路が必要になり、また、その情報を利用して結合荷重の修正量を計算するメカニズムがなければならない。そのようなものが生体中存在する可能性を否定することはできないが、少なくとも現在まで見つかっていない。

4. 非管理学習(教師なし学習)

非管理学習においては、目標出力 y^d は与えられない。ニューロン間の結合の荷重を更新する局所的規則のみが与えられる。その学習によってネットワークがどのような機能を持つようになるかは、ネット

ワークに入力されるパターンの統計的性質に依存する。

4-1 ヘブ学習

非管理学習のモデルとして、古くからよく知られているものは、ヘブ(Hebb, D. O.)によって提唱されたヘブ則である。いろいろなバリエーションが提案されてきたが、その典型的なものは、1つのシナプスの荷重 w はそれを通して入ってくる入力 s とニューロンの出力 y の積に比例して増加するというものである(図5)。式で表現するならば

$$\frac{dw}{dt} = ays \quad (a \text{ は正の定数}) \quad (5)$$

入力 s は変化するので ys も時々刻々変化するが、(係数 a が小さい場合には)ほぼ ys の平均値、すなわち y と s の相関値に比例して w が増加することになる。従って、この学習を相関学習と呼ぶこともある。

局所的な情報のみを利用して結合荷重が更新されるという意味で、この学習則はシナプスにおける学習のモデルとしては自然なものと考えられ、実際に神経生理学の分野でも(細かい点とはともかく広い意味での)ヘブ則の存在は認められているようである。なお、式(5)のままでは、 w は発散する(シナプス効率が無限に増加する)ことになるので、 w が大きくなると成長速度が小さくなるよう変形して、ある値に止まるようにするのが普通である。

式(5)の見かけは簡単であるが、これによってネットワーク全体がどのように変わっていくかを知ることが、管理学習の場合よりやっかいである。ここでは、非常に簡単な例を用いて、どのようなことが起こるのかを見てみたい。

図5で、入力は s 以外にはないとする。線形ニュー

図5 ヘブ則と反ヘブ則

ロンを仮定すると、出力は

$$y = ws \quad (6)$$

これをヘブ則(5)に代入すると

$$\frac{dw}{dt} = (as^2)w \quad (a > 0) \quad (7)$$

$as^2 > 0$ であるので、 w の初期値が正であるとする、その後 w は次第に大きくなっていく。その結果、 s と y の間の相関値もだんだん大きくなる。ある程度 w が大きくなると成長が抑制されるメカニズムを組み込んでおくと、あるところで w は一定の値になる。

この例では、結合荷重 w が次第に大きくなるという何の変哲もないことになったが、興味深いのは入力が複数の場合である。入力が2個の場合を見てみよう(図6)。この場合も、 w_1 、 w_2 は共に大きくなっていくが、入力 s_1 、 s_2 の分布のかたより具合で、大きくなる速度が異なり、その結果、このニューロンはある‘機能’を持つようになる。

得られる機能を理解するには、統計学でよく知られている主成分分析についての知識が必要になるので、若干の説明をしておこう。ニューロンに逐次加えられる入力ベクトル $s = (s_1, s_2)$ が図7の楕円のように分布しているとする(楕円でなくてもよいのであるが、直感的に分かりやすいのでそうする)。楕円の長軸方向のベクトル $u^{(1)}$ と短軸方向のベクトル $u^{(2)}$ は、この分布を特徴づけるベクトルで、それぞれ第1主成分ベクトル、第2主成分ベクトルと呼んでいる。2つのベクトルは必ず直交している。

さて、ヘブ学習の話に戻ると、 w_1 と w_2 は同じように大きくなるのではなく、荷重ベクトル $w = (w_1, w_2)$ は第1主成分ベクトル $u^{(1)}$ の方向に大きくなっ

図6 2入力の場合(ヘブ学習)

図7 主成分ベクトル

ていくのである。適当に増加が抑制されるメカニズムを組み込んでおくと、あるところで落ち着く。この結果、このニューロンは次のような性質を持つ。式(1)からわかるように出力 y は荷重ベクトルと入力ベクトルの内積になっているので、入力 s が与えられたとき、それが $u^{(1)}$ 方向を向いた(図7に示したような)ベクトルならば、このニューロンは強く反応するが、それとは直交した $u^{(2)}$ 方向を向いている入力ベクトルに対してはまったく反応しない($y=0$)。図7のような入力分布の場合、 $u^{(1)}$ 方向の入力は頻繁に現れるが、 $u^{(2)}$ 方向の入力パターンは滅多に出現しない。すなわち、この学習によって、頻繁に現れる入力パターンに対して特に強く反応する機能が獲得されることとなるのである。

情報論的には次のような説明もできる。もとの入力は2つであるので、入力は2次元の情報をもっているが、出力は1つであるので1次元の情報しかもたない。すなわち、このニューロンによって、2次元情報が1次元情報に圧縮されるわけである。図7のような入力分布の場合、通常 $u^{(1)}$ 方向の変動が意味があるので、この方向の入力変化に対して出力が最も大きく変化するような形で情報圧縮がおこなわれるのが、情報の損失が少なくてよい。ヘブ学習によると、そのような情報圧縮が行われるように、結合荷重が変化するわけである。

4-2 反ヘブ学習

理論上は、ヘブ則の逆すなわち反ヘブ則も重要である(著者は、この学習則が生体に存在するのかわか、寡聞にして知らない)。反ヘブ則とは、基本的には、シナプスを通して入ってくる入力とニューロン出力の積に比例して、結合荷重が減少する(抑

制性シナプスの荷重が増加する)というものである。すなわち、結合荷重を $-v$ とおくと(図5)、 v は次の式に従って変化する。

$$\frac{dv}{dt} = -avs \quad (a > 0) \quad (8)$$

再び、ニューロンはただ1つの入力を受ける場合を考えてみよう。この場合は、ニューロンの出力は

$$y = -vs \quad (9)$$

これを(8)に代入すると、

$$\frac{dv}{dt} = -(as^2)v \quad (10)$$

$-as^2 < 0$ であるので、 v は次第に小さくなり、最後は0になる。その結果、入力 s には無関係に出力 y は常に0、従って y と s の相関値は0になる。

相関が0となるという結果は、1入力の場合にはつまらない話であるが、入力が2つ以上の場合に意味がある。図8のネットワークで見てみよう。ニューロン2は、ニューロン1の出力 y_1 と外部入力 s を受けている。

$$y_2 = s - vy_1 \quad (11)$$

反ヘブ則はニューロン1からニューロン2への結合の荷重に対してのみ適用されるとする。この場合も、簡単な計算で、学習の結果 y_1 と y_2 の間の相関は0になることがわかる。ただし、 v は0にはならない。

このとき、次のような状態になっている。信号 s は、一般に、

図8 2入力の場合(反ヘブ学習)

- ・ 信号 y_i と完全な相関関係にある成分(信号 y_i と大きさは違うが、まったく同じ変動をする成分で、それを y_i 成分と呼ぼう)と
- ・ 信号 y_i とまったく無相関な成分(信号 y_i とまったく無関係な変動をする成分)

の和に分解することができる。学習が進むと、 v の大きさは、 v_{y_i} が s の中に含まれる y_i 成分と等しくなるように変化していく。(11)からわかるように、その結果、ニューロン 2 はニューロン 1 の出力とは無相関な信号のみを出力する。

情報論の観点からは、反ヘブ則の意義は次のように説明される。今ニューロン 1 とニューロン 2 の相関が大きいとすると、それは、どちらのニューロンも同じような振舞いをすることを意味し、言い換えれば、どちらもほとんど同じ機能を持っていることになる。そのような状況は、2つのニューロンが効率的に利用されている状態であるとは言えないであろう。ニューロン間の結合において反ヘブ学習が行われると、2つのニューロンは異なった機能を持つことになるのである。

4-3 脳におけるヘブ学習・反ヘブ学習の役割

以上、ヘブ学習、反ヘブ学習の意味を非常に簡単な例を用いて説明したが、脳におけるそれらの役割は(大胆に言うなら)次のようなことになるだろうか。生物が環境から受け取る入力とは極めて高い次元のものである。たとえば視覚系においては、視細胞の数に等しい入力が入ってくる。そのような高次元の情報をそのままの形で処理することは、ニューロンの数が多いといえども脳にとってかなりの負担であり、何らかの情報圧縮が必要になろう。どのような情報圧縮がよいかは入力パターンの統計分布に依存し、ヘブ学習は、まさに、入力環境に応じた最適な情報圧縮をもたらすのである。また、脳の中の多数のニューロンが外界からの刺激に対して同じ演算をするのでは、無駄が多い。ニューロン間の結合において反ヘブ学習が行われると、各ニューロンの機能が分化する。このように、ヘブ学習も反ヘブ学習も、限られたニューロンの資源を用いて、効率よく情報を処理する機能の形成に役立つのである。

4-4 より複雑なネットワークにおけるヘブ学習・反ヘブ学習

4. 1, 4. 2 節で述べたニューラルネットは極めて簡単なものであったが、ニューロンの個数が増

図9 主成分分析ネットワーク

えてくると、学習は複雑な様相を呈してくるであろう。一般のニューラルネット構造に対して解析を行うことは望むべくもないが、図9のネットワークについては、かなり詳しい数学的解析がなされている³⁾。N個の入力はM個のニューロンすべてに加えられている。ニューロンどうしは互いに(抑制性)結合で結ばれている。 w_{ij} は入力 j からニューロン i への結合の荷重で、ヘブ則に従って変化する。 $-v_{ij}$ はニューロン j からニューロン i への結合の荷重で、反ヘブ則に従う。

もしニューロン間に相互結合がないとするなら、図1のニューロンが複数個あるのと何らかわりがないので、ヘブ学習によって、すべてのニューロンはN次元入力パターンに含まれる第1主成分を検出する機能を持つようになる。すなわち、すべてのニューロンは同じ出力を出すことになる。ここで、相互結合の荷重 $-v_{ij}$ が反ヘブ則に従って変化すると、出力 $y_1, \dots, y_M$ は互いに無相関になる。このことが、 w_{ij} の変化にも影響を及ぼす。詳しい解析によると、つぎのような状態に落ち着くことが分かっている。各ニューロンはそれぞれ第1主成分、第2主成分、第3主成分等を検出する機能を持つようになる。すべての主成分を抽出するネットワークが、自己組織的に形成されるわけである。

図10のネットワークについても解析がなされている。このニューラルネットワークにおいては、N個の入力はそれぞれ対応するニューロンにのみ加えられている。学習は、相互結合に対してのみ行われ、反ヘブ則に従う。

このネットワークがどのような機能を持つようになるかについて、例を挙げて説明しよう。入力として

図10 新奇性フィルタ

は、人の顔画像(を数値ベクトル化したもの)が次々に与えられるとする。学習時点では、ほとんどの顔は眼鏡をかけていないとする。すると学習後には、眼鏡をかけた顔パターンを提示すると、眼鏡の部分だけが出力される。眼鏡をかけていない顔に対しては何の出力も現れない。すなわち、入力パターン s が珍しい成分を含むとき、それだけが出力されるのである。このような機能を持つネットワークを新奇性フィルタ⁵⁾と呼んでいる。この機能の形成メカニズムは、反ヘブ則の基本的振舞いから説明できるのであるが³⁾、ここではふれない。

また、ヘブ則、反ヘブ則にはいろいろなバリエーションが考えられ、面白いことに、同じ図9、図10のネットワーク構造に対して、学習則を少し変えるだけで、まったく異なった機能を獲得させることができる。興味のある読者は、文献2)を参照されたい。また、これまで各ニューロンはすべて線形ニューロンであるとしたが、非線形ニューロンを要素とするネットワークに対してヘブ則・反ヘブ則を適用すると、さらに多彩な機能形成が見られるであろう。

5. あとがき

本解説では、ニューラルネットの分野でよく知られている管理学習と非管理学習の考え方について述べた。前者については誤差逆伝搬学習を、後者についてはヘブ学習・反ヘブ学習を取りあげた。しかし、もちろん、ニューラルネットの分野で扱われる学習モデルは、ここで述べたものに尽きるわけではない。たとえば、全結合形のネットワーク(図3(b))については、ボルツマン学習と呼ばれる有名なモデルがある¹⁾。

また、非管理学習の説明において、ネットワークの空間的な構造はまったく考慮しなかった。例えば、図9のネットワークで、入力はすべてのニューロンに加えられており、いわゆる受容野というものとはまったく考慮されていない。空間的な構造を持つネットワークについては、特徴マップと呼ばれるモデルがあり^{1, 5)}、大脳皮質のコラム構造の形成に関して示唆を与えるもので、生理学者にとっても面白いモデルかもしれない。

文 献

- 1) 松岡清利(編著): ニューロコンピューティング, 朝倉書店(1992)
- 2) 松岡清利: ニューラルネットの学習と進化, 計測と制御, 843-847 (1993)
- 3) Matsuoka, K., & Kawamoto, M.: Neural Networks, 7, 753-765 (1994)
- 4) Matsuoka, K., & Kawamoto, M.: Neural Networks, 8, 411-419 (1995)
- 5) Kohonen, T.: Self-Organization and Associative Memory, Springer-Verlag (1984)