連載講座

ニューラルネットの基礎数理 (1)

上坂　吉則

1. はじめに

ニューラルネットに関する研究は約半世紀前の McCulloch and Pitts [9] に遡ることができる．この論文は，ニューロンを論理素子と見立てたとき，ニューラルネットはどんな論理関数をも実現できるという，ニューロンの万能性を示したものである．と同時に，数学者と生理学者の共同研究によるもので，今日でいう学際的な仕事のはしりでもある．

その後，60年代にパーセプトロンと呼ばれる世界最初の学習機械が心理学者 Rosenblatt[12] によって提唱されると，ニューラルネットへの関心が世界的な規模で高まることになる．日本でも多くの研究者がこの分野に参入し，パターン認識，学習，認知，記憶，生体情報，バイオニックスといった言葉が通信工学や計算機工学のなかで頻繁に聞かれるようになった．しかし，約10年後にはひと握りの研究者を残して，大部分は当時興隆しつつあった人工知能（A I）やコンピュータサイエンスへと移っていく．

その主たる原因は Minsky and Papart の書物[10]の初版においてパーセプトロン限界説が唱えられたからだとよくいわれているが，これは必ずしも当を得ていない．実際，彼らが論じたのはごく特殊なパーセプトロンの族の能力評価であり，後にふれるように，ニューラルネットは順序機械としても万能性をもっていることが古くから知られている[8]のである．むしろ，パーセプトロンに代表されるこの種の研究が狙った課題が，認識や学習などいわゆるブレイン・サイエンスの中でも難問中の難問に属していたからだといえよう．こうした課題は一朝一夕で片がつくとは考えられない．

この事情は，80年代に入ってからのいくつかの新しいアプローチの提唱に始まるニューロブームにおいても変わるものではない．このフィーバーが多くの人々の関心を呼び，ふたたび多数の研究者を呼び寄せているが，再度氷河期に陥らないためには，扱っている課題が内包している困難さをよく認識しておくことが肝要だと思われる．

それはともかく，今日ニューラルネットの研究は非常に活発化している．その結果，ニューラルネットのタイプも多岐にわたり，したがってその様相を一口で捉えることは難しくなってきている．そこで，この小文では種々のニューラルネットをその原理的側面から眺めることによって分類し，その中から基本的な3つのニューラルネット：学習認識機械，決定論的最小値探索機械，確率論的最小値探索機械を取り上げることにする．そして，本誌の読者は数理的な取扱いにも慣れ親しんでおられると思われるので，これらのニューラルネットの数理的枠組みを紹介することによってその基本的な仕組みを理解することを手助けできればと考えている．紙数の制約から詳細な証明などは一部省略せざるを得ないが，これらについては拙書[17]を参照していただきたい．

人工のニューラルネットは“ニューロンモデル”と呼ばれる多入力一出力の情報処理素子から成っている．これは生体がもっている本物のニューロンのある種の性質を取り入れ，他の性質を捨象することによって得られるいわば“ニューロンもどき”であり，次のように3つの観点から分類することができる．

1°　扱う情報が2値をとる**デジタルタイプ**と連続値をとる**アナログタイプ**，

2°　情報の伝達に遅れのない**静的タイプ**と内部に記憶をもつ**動的タイプ**，

3°　出力が入力に対して一意に定まる**決定論的タイプ**とランダムに定まる**確率論的タイプ**．

したがって原理的には8種類のニューロンモデルを考えることができる（個々のニューロンモデルの動作の詳細はそれらが使われるニューラルネットのところで説明するのでそちらを参照されたい）．

うえさか　よしのり　東京理科大学　理工学部
〒278 野田市山崎2641

一方，回路のトポロジーあるいはアーキテクチャから見ると，ニューラルネットは情報が入力側から出力側へと一方向に流れる**フィードフォワード型**とループを持つ**フィードバック型**に大きく分けることができる．

フィードバック型のニューラルネットに静的モデルを用いても，情報が回路の中を一瞬の内にかけめぐってしまうので，議論のしようがない．また，フィードフォワード型のニューラルネットに動的モデルを用いてもあまり面白いことは期待できない．というわけで，静的モデルを素子とするフィードフォワード型回路が4種類と動的モデルを素子とするフィードバック型回路が4種類，計8種類のニューラルネットが原理的に考えられることになる．

この8種類のニューラルネットによってできる典型的なマシン，つまり，ニューロコンピュータとしては次のものを挙げることができる．

(1) フィードフォワード型回路

(a) デジタル型静的決定論的モデルを用いたもの
- 学習認識機械としての古典パーセプトロン (Rosenblatt, 1961)
- 静的連想記憶としてのアソシアトロン (Nakano, 1972)

(b) デジタル型静的確率論的モデルを用いたもの
- (信頼性の低い部品から信頼性の高いシステムを構築できるという理論 (von Neumann, 1956))

(c) アナログ型静的決定論的モデルを用いたもの
- 学習認識機械としての現代パーセプトロン (Rumelhart et al., 1986)
- 静的連想記憶としての一般化されたアソシアトロン (Uesaka & Ozeki, 1972)

(d) アナログ型静的確率論的モデルを用いたもの
- (特になし)

(2) フィードバック型回路

(a) デジタル型動的決定論的モデルを用いたもの
- 動的連想記憶としてのアトラクタマシン(Hopfield, 1982)

(b) デジタル型動的確率論的モデルを用いたもの
- 最小値探索機械としてのボルツマンマシン (Ackley et al., 1985)

(c) アナログ型動的決定論的モデルを用いたもの
- 最小値探索機械としてのホップフィールドマシン (Hopfield & Tank, 1985)

(d) アナログ型動的確率論的モデルを用いたもの
- 最小値探索機械としてのガウシァンマシン (Akiyama et al., 1988)

これらのニューラルマシンを数理的なシステムとして考えたとき，その原理的側面が浮かび上がってくる．デジタル型静的決定論的モデルで構成される学習機械は多変数の論理関数であり，任意の論理関数がこのシステムによって実現できることがわかっている[9]．デジタル型動的決定論的モデルによる連想記憶マシンは本質的には有限オートマトン（あるいは離散力学系）であり，任意のオートマトンがこのシステムによって実現できることがずいぶん前から知られている[8]．デジタル型動的確率論的モデルを用いた最小値探索機械は数学的には有限マルコフ連鎖であり [18]，"焼きなまし"と呼ばれる物理現象に由来するシミュレーティッドアニーリングの手法の範疇に入る[1]．アナログ型静的決定論的モデルから成る学習認識機械は多変数の実数値関数であり，任意の多変数関数が任意の精度でこのシステムによって近似できることがわかっている[5]．アナログ型動的決定論的モデルによる最小値探索機械は勾配系と呼ばれる古くから知られた力学系（あるいは微分方程式系）によく似たシステムである[15]．そしてこれにガウシァンノイズを加えて改良した，アナログ型動的確率論的モデルによる最小値探索機械は確率微分方程式で記述されるタイプのシステムである[3]．この他に，主として甘利によって精力的に展開されてきた統計力学的アプローチがあるが，これについては文献[4]に優れた解説を見ることができる．

このように，ニューラルネットはさまざまな数理的側面を持っているが，ここでは関数近似としての現代パーセプトロン，力学系としてのホップフィールドマシンおよび有限マルコフ連鎖としてのボルツマンマシン（とアニーリング）の3つについて，その基本的な仕組みを紹介し，今後解かれるべき課題について考えてみたいと思う．

2. 学習認識機械

開区間 (0,1) の n 個の直積集合を X で表わす．つまり，X の元 x は値が (0, 1) に属する n 個の実数の組 $(x_1, \cdots, x_n)$ として書くことができる．このとき，次のように定められる X から (0,1) への関数：

(2.1) $f(x) = \sigma(w_0 + w_1 x_1 + \cdots + w_n x_n)$

を考える．ここに，σ は実数全体$\mathbf{R}$から(0,1)への関数：

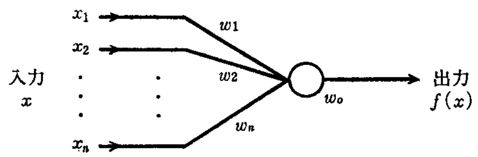

図 2.1 アナログ型静的決定論的ニューロンモデル

$$\sigma(u)=\frac{1}{1+\exp(-u)} \tag{2.2}$$

であり，$w_0, w_1, \cdots, w_n$ は実数を値に持つ定数である．

この関数fを**アナログ型静的決定論的ニューロンモデル**という．情報処理素子としては入力$x=(x_1, \cdots, x_n)$を出力$f(x)$に変換するプロセッサと見ることができる(**図 2.1**)．

このような関数をどうしてニューロンのモデルと考えるかはニューロンの生理学的性質による．つまり，生体のニューロンは多入力一出力の情報処理素子であり，情報はニューロンの活動を表わすインパルスの頻度で表現されていると考えられているからである．加えて，入力はニューロン固有の定数 $w_0, w_1, \cdots, w_n$ によって重み付きで加算され，それがニューロンの内部電位を定め，内部電位は関数σによって非線形に変換されて出力されることが知られている．つまり，$f(x)$ は入力 x によって引き起こされた出力としてのパルス頻度である．この意味で $w_1, \cdots, w_n$ をニューロンの**重み係数**，w_0(の符号を変えたもの)を**しきい値**と呼んでいるが，ここでは簡単に両者を**重み**と呼ぶことにする．

このようなニューロンの出力を他のニューロンの入力に接続することによってニューロンを素子とする回路，すなわち，**ニューラルネット**ができる．

ここでは簡単のため，n個の入力と1個の出力を持つ回路を考える(**図 2.2**)．すなわち，m個のニューロン$j=1, \cdots, m$に対して：

$$y_j=f_j(x)=\sigma(w_{0j}+w_{1j}x_1+\cdots+w_{nj}x_n) \tag{2.3}$$

が層状に並んでおり，n個の入力が各ニューロンに入り，それらの出力 $y_1, \cdots, y_m$ がもう1つのニューロン：

$$z=g(y)=\sigma(v_0+v_1y_1+\cdots+v_my_m) \tag{2.4}$$

の入力となって，そこから出力zが出ると考える．

これは典型的な2段のフィードフォワード型ニューラルネットであり，その入出力関係は式(2.3)と(2.4)から明らかなように関数$f_1, \cdots, f_m$とgの合成として

$$h(x)=g(f_1(x), \cdots, f_m(x)) \tag{2.5}$$

と与えられ，Xから(0,1)への関数となっている．このhは**パーセプトロン**とも呼ばれる．

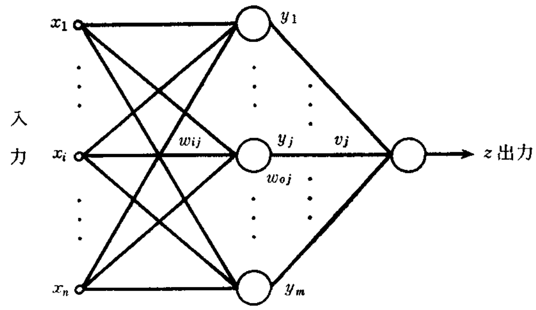

図 2.2 n入力1出力を持つ2段のフィードフォワード型ニューラルネット

パーセプトロンとは“perception”，すなわち，“知覚”という心理学用語に由来した名称であって，図形などのパターンを認識する機械のことである．たとえば，関数hの値域(0,1)を(0,0.5)と[0.5,1)の2つに分割し，そのhによる逆像：

$$\begin{cases} X_1=h^{-1}((0,0.5))=\{x|h(x)\in(0,0.5)\}, \\ X_2=h^{-1}([0.5,1))=\{x|h(x)\in[0.5,1)\} \end{cases} \tag{2.6}$$

をとると，これらはXを2つの部分集合X_1とX_2に分割する．したがって，Xをパターンの集合と考えると，$h(x)$ の値を見ることによって，パターンxが2つのカテゴリ X_1, X_2 のいずれに属しているかがわかる．このことはhによって**パターン認識**ができることを意味しており，hはしばしば**認識関数**と呼ばれる(**図 2.3**)．たとえば，X_1が文字パターン“A”の集合であり，X_2が“A”以外のパターンの集合であれば，hは文字“A”を認識できることになる．

逆に，文字“A”を認識できる関数hを作ることを考えてみよう．この場合，実際的場面ではカテゴリ X_1 が陽に与えられることはなく，いくつかの文字パターンがX_1やX_2の元であるという事例が得られるに過ぎない．いいかえれば，パターン集合Xの“有限”部分集合SとSに属するパターンのカテゴリ名が与えられ，それをも

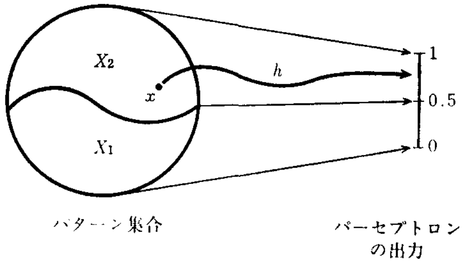

図 2.3 認識関数hによってパターン集合は2つのカテゴリ X_1, X_2 に分割される

とにして認識関数を作り上げるということを考えなければならない．

このことを数学的に捉えれば，Xから$(0,1)$への未知の関数dがあり，有限集合$S \subseteq X$の各元xに対するdの値$d(x)$だけが与えられたとき，X全体にわたってdをできるだけよく近似する関数 $h: X \to (0,1)$ を求めよという，関数補間の問題に帰着される．dを**目標関数**，Sを**学習パターンの集合**という．

関数hの目標関数dに対する近似の良さとしては，通常，学習パターンにおける両者の値の平均2乗誤差：

$$E = \frac{1}{|S|} \sum_{x \in S} [h(x) - d(x)]^2 \tag{2.7}$$

が用いられる．ここに$|S|$はSの元の個数である．学習パターンから認識関数hを設計するには，この誤差Eが最小となるように，重みw_{ij}とv_jを決めればよい．しかし，Eの形から明らかなように，望ましい重みは解析的には求めようがない．そこで逐次的に近似解を求めることになるが，その1つの数値計算法を与えるのが**誤差逆伝搬法**[13]であり，解を求めるプロセスを**学習**と呼んでいる．この種のニューラルネットが**学習機械**と呼ばれるのはこのことに由来している．

具体的には，非線形な関数の最小解を求める手法として古くから知られている**勾配法**ないしは**最急降下法**を用いればよい．すなわち，重みw_{ij}, v_jを時間tの関数と考え，誤差Eをポテンシャルとする力学系：

$$\frac{dw_{ij}}{dt} = -\frac{\partial E}{\partial w_{ij}}, \quad \frac{dv_j}{dt} = -\frac{\partial E}{\partial v_j} \tag{2.8}$$

を用意する．このとき，Eはw_{ij}, v_jを介して時間の関数となるが，w_{ij}, v_jが上の微分方程式系の解であれば，

$$\frac{dE(w_{ij}(t), v_j(t))}{dt} \leq 0 \tag{2.9}$$

が成り立ち，等号が成り立つ条件は

$$\frac{\partial E}{\partial w_{ij}} = 0, \quad \frac{\partial E}{\partial v_j} = 0 \tag{2.10}$$

で与えられることが容易に示される．

したがって，Δを十分小さい正の定数にとり，適当な初期値$w_{ij}(0)$，$v_j(0)$から出発して漸化式：

$$w_{ij}(t+\Delta) = w_{ij}(t) - \Delta \frac{\partial E}{\partial w_{ij}}, \tag{2.11}$$

$$v_j(t+\Delta) = v_j(t) - \Delta \frac{\partial E}{\partial v_j} \tag{2.12}$$

を用いて点列$w_{ij}(k\Delta)$，$v_j(k\Delta)$ $(k=1,2,3,\cdots)$を生成すると，これは誤差Eの最小点（一般には極小点）に収束することになる（**図 2.4**）．

いま，偏微分$\partial E/\partial w_{ij}$と$\partial E/\partial v_j$を実行してみると，

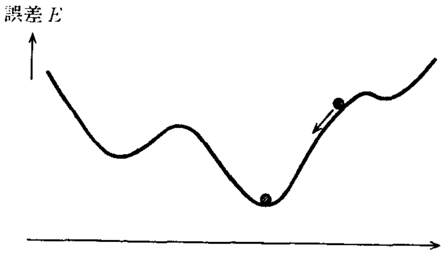

図 2.4 誤差Eの曲面上を移動して力学系の状態はEの極小点に到達する

この漸化式はより具体的に，$i=0,\cdots,n$；$j=1,\cdots,m$に対して

$$w_{ij}^{new} = w_{ij}^{old} - \frac{\Delta}{|S|} \sum_{x \in S} \delta_{1j}(x) x_i, \tag{2.13}$$

$j=0,\cdots,m$に対して

$$v_j^{new} = v_j^{old} - \frac{\Delta}{|S|} \sum_{x \in S} \delta_2(x) f_j(x) \tag{2.14}$$

と計算される．ここに

$$\delta_{1j}(x) = \delta_2(x) v_j f_j(x)(1 - f_j(x)), \tag{2.15}$$

$$\delta_2(x) = [h(x) - d(x)] h(x)(1 - h(x)) \tag{2.16}$$

であり，$x_0 \equiv f_0(x) \equiv 0$と既約してある．この偏微分の計算の過程で式(2.2)の導関数が

$$\frac{d\sigma}{du} = \sigma(u)(1 - \sigma(u)) \tag{2.17}$$

で与えられることを積極的に利用していることを注意しておく．

ここでδ_2をみると，これはパーセプトロンの出力$h(x)$の目標関数$d(x)$に対する誤差にほぼ相当しており，またδ_{1j}を求めるのにδ_2を用いていることがわかる．つまり，入力側の重みw_{ij}の修正に出力側の誤差を利用している．このように重みの修正のさいに誤差が，入力情報の流れとは逆に，出力側から入力側に伝搬していくことが誤差逆伝搬法の命名の由来になっているわけである．

ここでは典型的な2段のパーセプトロンを取り上げたが，こうした考えは段の数がもっと多くても通用する．また回路が層状になっていなくても，回路にループが存在しない任意のフィードフォワード型のニューラルネットに適用することができる．この意味では誤差逆伝搬法は万能の学習法だということができる．そしてこのことが80年代にふたたび爆発的な関心がニューラルネットに寄せられたことの大きな要因の1つになっているといってよいであろう．

参 考 文 献

[1] Aarts, E. and Korst, J.: *Simulated annealing and Boltzmann machines: A stochastic approach to combinatorial optimization and neural computing*, Wiley, 1988.

[2] Ackley, D. H., Hinton, G. E. and Sejnowski, T. J.: A learning algorithm for Boltzmann machines, *Cognitive Science*, 9 (1985), 147–169.

[3] 秋山，山下，梶浦，安西，相磯：ガウシアンマシンによる組合せ最適化，電子情報通信学会技術研究報告，**MBE 88-183** (1988).

[4] 甘利：神経回路網の数理―脳の情報処理様式―，産業図書，1978.

[5] Funahashi, K.: On the approximate realization of continuous mappings by neural networks, *Neural Networks*, **2**, **3** (1989).

[6] Hopfield, J. J.: Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA*, **79** (1982), 2554–2558.

[7] Hopfield, J. J. and Tank, D. W.: "Neural" computation of decisions in optimization problems, *Biol. Cybern.*, **52** (1985), 141–152.

[8] Kleene, S. C.: Representation of events in nerve nets and finite automata, in "Shannon, C. E. and McCarthy, J. (eds.): *Automata Studies*, Annals of Mathematical Studies No. 34, Princeton Univ." (1956), 3–41.

[9] McCulloch, W. S. and Pitts, W. H.: A logical calculus of the ideas immanent in neural nets, *Bull. Math. Biophys.*, **5** (1943), 115–133.

[10] Minsky, M. L. and Papert, S. A.: *Perceptrons, An introduction to computational geometry, expanded edition*, The MIT Press, 1987.

[11] Nakano, K.: Associatron—a model of associative memory, *IEEE Trans. on SMC*, **SMC-2** (1972), 380–388.

[12] Rosenblatt, F.: *Principles of neurodynamics—Perceptrons and the theory of Brain Mechanisms*, Spartan, 1961.

[13] Rumelhart, D. E., Hinton, G. E. and Williams, R. J.: Learning representations by back-propagating errors, *Nature*, **323**, 9 (1986), 533–536.

[14] 上坂，尾関：連想型記憶の二，三の性質，電子通信学会論文誌，**55-D** (1972), 223–230.

[15] 上坂：2値変数の実数値関数から導かれるエネルギーを持つニューロン回路網の安定性について，電子通信学会技術研究報告，**PRU 88-6** (1988).

[16] 上坂，塚田：シミュレーティッドアニーリングのための受理関数の族について，電子情報通信学会技術報告，**NC 89-66** (1990), 31–36.

[17] 上坂，尾関：パターン認識と学習のアルゴリズム，文一総合出版，1990.

[18] 上坂：ニューロダイナミックスの数理，電子情報通信学会誌，**73**, 2 (1990), 131–136.

× × × × ×