特集 ニューラルネット

計測と制御
Vol. 30, No. 4
(1991年4月)

≪解 説≫

階層型ニューラルネットワークの原理的機能

船橋賢一*

1. はじめに

近年，ニューラルネットワークという神経回路網に学ぶ情報処理方式が広く話題になっている．ニューラルネットワークの研究は，1943年の McCulloch-Pitts のニューロンモデルの研究を端緒としており，1957年の Rosenblatt によるパーセプトロンの研究により活発化した．このブームは，1969年の Minsky-Papert によるパーセプトロンの限界に関する研究の後，沈滞したといわれている．今回のブームのきっかけは1986年の Rumelhart らによる階層型ネットワークに対する Back-Propagation アルゴリズムの定式化であると思われ，当初の目的の認知心理学への応用を超えて情報工学への応用が盛んになされるようになった．この背景には，コンピュータの進歩のおかげで大規模なシミュレーションが可能になったことが挙げられる．またそれ以前（1982年）の Hopfield による相互結合ネットワークの最適化問題への応用もこの方面に多くの人の関心を向けた．当初，この方面の研究は，コネクショニズムといわれていたが，最近はニューロコンピューティングといわれることが多い．

Back-Propagation アルゴリズムは，最近，相互結合型（recurrent）ネットワークにまで拡張されたが，応用上では，階層型ネットワークは最もよく用いられている．ニューラルネットワークの理論的研究は Mc Culloch-Pitts 以来行われてきているが，現在種々の応用が試みられる中で，昔からの問題と共に，新たに基本的な問題が浮かび上がってきているように筆者には思われる．本解説では，特に階層型ニューラルネットワーク（multilayer feedforward neural network）に関して，数学的な観点から原理的な能力に関して最近得られた結果で，応用とも密接に関わるものをサーベイする．

* 豊橋技術科学大学情報工学系
豊橋市天伯町字雲雀ケ丘 1-1

キーワード：階層型ニューラルネットワーク（multilayer neural network），ユニット（unit），入出力写像（input-output mapping），近似理論（approximation theory），多変量解析（multivariate analysis）．

2. 階層型ニューラルネットワーク

本解説でいうニューラルネットワークとは，現実の神経回路網ではなく，人工的な神経回路網モデル，すなわち非線形・並列処理的な，神経回路網にヒントを得た工学的なシステムである．

ニューラルネットワークは機構の面から大別して，階層型ネットワーク（フィードバックを含まないもの）と相互結合型ネットワーク（フィードバックを含むもの）の2種に分類される．

また学習方式の面から分類すると，教師なし学習方式と教師付き学習方式の2つに分けられる．Hebb の学習規則といわれるものは教師なし学習方式であり，いわゆる Back-Propagation アルゴリズムは教師付き学習方式である．

階層型ネットワークについて以下詳細を述べよう．階層型ネットワークは多層パーセプトロンとも呼ばれ，入力層，出力層およびいくつかの中間層（hidden 層）からなり，入力から出力の方向に層間の結合がある．**図1**に3層のネットワークを示す．各層は，ユニットとよばれる情報処理素子からなり，ユニットの入出力関係は，$\{x_i\}$ をユニットへの入力，y を出力，

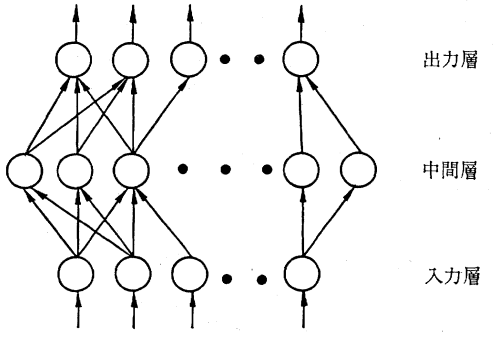

図1 3層ニューラルネットワーク

JL 0004/91/3004-0280 © 1990 SICE

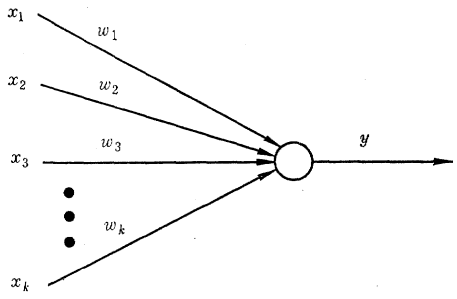

図 2　ユニット

w_i をそのユニットへの結合の重みとすると

$$y=\phi(\Sigma w_i x_i-\theta)$$

と表わされる（**図 2** 参照）．ここで ϕ は出力関数，θ はしきい値とよばれる．出力関数 ϕ は，通常，定数でない単調増加，有界な関数が用いられるが，出力層においては，線形関数 $\phi(x)=x$ を用いる場合もある．ユニットはニューロンの非常に粗いモデル化と考えてよい．結合の重みは正，負どちらの値もとることができ，それぞれ興奮性の結合，抑制性の結合に対応する．以下，出力関数として，連続，定数でない単調増加，有界なものをシグモイド関数とよぶ．シグモイド関数として通常

$$\phi(x)=1/(1+\exp(-x))$$

がよく用いられる．

連続な出力関数をもつユニットからなるアナログ入力の n 入力 m 出力階層型ネットワークは連続写像

$$f: R^n \to R^m$$

を決める．これをネットワークの入出力写像という．

階層型ネットワークに対する教師付き学習アルゴリズムとして Back-Propagation アルゴリズムが知られている．$\{x_i\}$ を学習サンプルにおける入力データ，$\{y_i\}$ を理想出力（教師データ）とする．f をネットワークの入出力写像とするとき，誤差関数

$$E=\Sigma \| f(x_i)-y_i \|^2$$

に対して E を最小化する，結合係数（重み）およびしきい値を最急降下法で求めることを考える．これには，E の重み w_{ij}，しきい値 θ_i による偏微分，$\partial E/\partial w_{ij}$，$\partial E/\partial \theta_i$ を求めなければならない．

これらを求めるアルゴリズムが，Rumelhart ら[1]によって与えられた．これが Back-Propagation アルゴリズムであるが，ここでは詳細を述べない．

Back-Propagation の初期の応用で最も有名なものは，Sejnowski-Rosenberg[2]による NETtalk であり，これは英語のスペリングを入力としてその発音を出力する 3 層ネットワークである．

3. 階層型ニューラルネットワークの原理的機能

ユニットの出力関数が，ヘビサイド関数 $H(x)$ ($H(x)=0, x<0$；$H(x)=1, x\geqq 0$) の階層型ネットワークは Rosenblatt 以来パーセプトロンと呼ばれている．入力が 0 と 1 の多層パーセプトロンは任意の論理関数を実現できることは，すでに McCulloch-Pitts により知られていた．中間層のないパーセプトロンの学習アルゴリズムは Rosenblatt により与えられた．ユニットの出力関数が可微分の場合の階層型ネットワークを，（現代）パーセプトロンと呼ぶ人もいるが，一般に中間層をもつ場合の教師付き学習アルゴリズムとして，Back-Propagation アルゴリズムがある．

パターン認識への階層型ネットワークの応用に関して，中間層なくしては線形分離可能なカテゴリーしか識別できないが，中間層の導入により種々の応用でよい性能が確認されてきた．この場合，中間層 1 層では，凸あるいは凹な識別境界が形成でき，中間層 2 層では，凸でも凹でもない識別境界が形成できることが論理的な解釈でわかることを Lippmann[3]が述べている．

その後 Lippmann らは，任意の識別境界が，中間層 1 層で形成できる可能性をシミュレーションで示した．

つぎの定理[5]は，これらのことを数学的に保証する．

《**定理 1**》 (Funahashi)

$\phi(x)$ をシグモイド関数（単調増加，定数でない有界な連続関数）とする．K を R^n の有界閉集合，$f(x_1, \cdots, x_n)$ を K 上の連続関数とする．そのとき任意の $\varepsilon>0$ に対して，ある自然数 N，定数 c_i, θ_i ($i=1, \cdots, N$)，w_{ij} ($i=1, \cdots, N$；$j=1, \cdots, n$) が存在して，

$$\max_{x\in K} \left| f(x_1, \cdots, x_n) - \sum_{i=1}^{N} c_i \phi\left(\sum_{j=1}^{n} w_{ij} x_j - \theta_i \right) \right| < \varepsilon$$

が成立する．

この定理は，出力層が線形ユニットからなる 3 層ネットワークによって，任意の連続写像 $f: R^n \to R^m$ が，任意の有界閉集合上で任意の精度で一様に近似的実現されることを示す．同様の結果は，Cybenko[6]，Hornik-Stinchcombe-White[7] によっても得られている．

出力層が，たとえばシグモイド関数 $\phi(x)=(1+\exp(-x))^{-1}$ をもつユニットからなる 3 層ネットワークを用いれば，任意の連続写像 $f: R^n \to (0,1)^m$ が有界閉

集合上で任意の精度で近似的に実現できることは，この定理からわかる．また，より多層のネットワークに対しても同様のことがいえることがこの定理から導かれる．定理1の証明は，Irie-Miyake[4]の Fourier 積分に関する関係式をふまえ軟化作用素および Paley-Wiener 理論[17]を用いてなされる．Cybenko[6]は $\phi(x)$ として単調増加を仮定せず，単に $\lim_{x\to\infty}\phi(x)=1$, $\lim_{x\to-\infty}\phi(x)=0$ を仮定して，関数解析[17]における Hahn-Banach の定理と Riesz の定理を用いて背理法によって同様の結果を証明している．Hornik-Stinchcombe-White[7]では cosine 関数の一部を用いて作ったシグモイド関数 (cosine squasher) と一般的な Stone-Weierstrass の定理[17]を用いて，単調増加ではあるが一般に不連続な $\phi(x)$ に対して同様な結果を証明している．

上記定理1は，つぎの古典的な Weierstrass の近似定理[17]と類似している．

《**定理**》 (Weierstrass)

K を R^n の有界閉集合とする．$f(x_1,\cdots,x_n)$ を K 上の実数値連続関数とする．このとき任意の $\varepsilon>0$ に対して，ある自然数 N と実数 $a_{i1\cdots i_n}$ $(0\leq i_1+\cdots+i_n\leq N)$ が存在して

$$\max_{x\in K}\left|f(x_1,\cdots,x_n)-\sum_{0\leq i_1+\cdots+i_n\leq N}a_{i1\cdots i_n}x_1^{i1}\cdots x_n^{in}\right|<\varepsilon$$

が成立する．

この連続関数の多変数多項式による近似定理は，現在ではより一般的な Stone-Weierstrass の近似定理[17]より導かれる．階層型ネットワークに対する定理1は，Weierstrass の多項式近似定理と同様ではあるが，異なっているのはつぎの点である．Weierstrass の定理では，パラメータ $a_{i1\cdots i_n}$ が線形に入っているが，定理1では，重み w_{ij} がシグモイド関数の中に非線形的に入っている．これにより多項式近似では，次数を制限したとき最良近似が存在するが，定理1のシグモイド関数による近似では，L^2 近似でも最良近似の存在が保証されないこともいえる．

4. 階層型ニューラルネットワークにおける近似理論

3節に述べた階層型ネットワークの原理的機能の研究では，写像を近似するものとして，階層型ネットワークをとらえているが，実際の応用上，ネットワークは，有限サンプルの入出力関係の例から学習によって入出力写像を推定する機能をもっていると考えられる．これを学習による汎化 (generalization) という．汎化の理論的解明は，最近 Vapnik-Chervonenkis の研究をふまえ，Baum-Haussler[8]によって始められた．すなわち，中間層と出力層が，ヘビサイドユニットからなる3層ネットワークに対する汎化の問題を確率論的に定式化し，VC 次元の理論を用いて，有効な汎化が得られるのに十分な学習サンプル数とユニット数および結合の数との関係を与えている．ただし，ヘビサイド出力関数の場合しか扱えないようである．

しかし純粋な関数近似の問題に限っても，階層型ネットワークには数学的な問題が数多く残されている．

たとえば，例外的な関数を除けば，中間層のユニット数の制約下で，少なくとも平均自乗誤差の意味で最良近似が存在するかといった問題は基本的である．また4層のネットワークを用いれば，一般に3層ネットワークを用いるよりも少ない中間ユニットを用いて写像を近似できるのではないかと予想される．著者は，このことを予想しているが[5]，Chester[9]は，その例を示している．3層ネットワークによる任意の連続写像の近似的実現可能性を示す定理1は，3層ネットワークで応用上で十分であるといっているわけではないことに注意されたい．

5. 階層型ニューラルネットワークと多変量解析

階層型ネットワークは，学習によって一種の多変量解析を行う機能をもっていることを示唆する研究が最近行われている．その1つは，入力データの上で恒等写像を近似的に実現する問題で，Cottrell ら[10]による画像のニューラルネットワークによる情報圧縮の研究から始まっている．これは**図3**のような入力ユニット数と出力ユニット数が同じで hidden ユニット数が入力ユニット数より少ない3層ネットワークで，入力パターンと同じ教師パターンを与えて学習させて

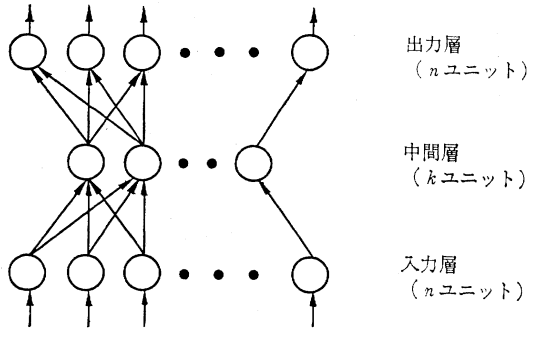

図3　n入力n出力の3層ネットワーク

hidden 層の出力を情報圧縮したパターンとして利用する方法である．もう1つはパターン認識に応用する場合におけるベイズ識別や判別分析との関係である．

前者の理論的解析は，すべて線形ユニットの3層ネットワークの場合に Bourlard-Kamp[12] によってまず行われた (Baldi-Hornik[13]も参照)．教師パターンは，入力パターンと同じであるからこれは，入力データの上で恒等写像を実現する問題として定式化される．この場合，主成分分析 (Principal Component Analysis) と関連がある．

《**定理2**》 (Bourlard-Kamp)

R^n 内のデータ $\{x_i, i=1,\cdots,N\}$ に対して，n 入力 n 出力で中間層が k ($<n$) ユニットをもつ，すべて線形ユニットからなる3層ネットワークの入出力写像を f としたとき，$\|\cdot\|$ をユークリッドノルムとすれば，恒等写像の近似誤差

$$(1/N)\sum_{i=1}^{N}\|x_i-f(x_i)\|^2$$

の最小値は，データ $\{x_i\}$ の $K-L$ (Karhunen-Loève) 変換による k 項近似（(すなわち主成分分析による第 k 主成分までの近似）における平均二乗誤差になる．

この定理では，すべて線形ユニットをもつ3層ネットワークの場合であるから，入出力写像は affine 写像の結合になる．問題とすべきは，中間層のユニットがシグモイド出力関数をもつ場合どうなるかであるが，船橋[20]は，つぎの結果を得た．

《**定理3**》(Funahashi) 中間層のユニットが，一般に非線形の出力関数をもつとき，ネットワークの入出力写像 f に対して恒等写像の近似誤差

$$(1/N)\sum_{i=1}^{N}\|x_i-f(x_i)\|^2$$

は，$K-L$ 変換による k 項近似の誤差以上であり，C^1 級のシグモイド関数のときは，いくらでも，$K-L$ 変換による k 項近似による誤差に近づきうる．

したがって，恒等写像の近似的実現に関しては，3層ネットワークでは中間層の非線形性は，$K-L$ 変換より性能を悪くするといえる．この性質は，用いる学習アルゴリズムが何であっても成立する．また，ネットワークは Back-Propagation アルゴリズムによる学習が理想的に進めばデータの k 次元の主成分を k 個の中間層のユニットで取り出すようになることがわかる．

これは，Cottrell-Munroのシミュレーション結果[11]とも対応する．

K-L 変換を一般に越える性能を得るには，たとえば中央層を絞った5層ネットワークを用いればよい．このことは，定理1の3層ネットワークによる任意の連続写像の近似的実現性が示唆している．5層ネットワークの使用は，片山ら[18]および入江-川人[19]により最初，提案された．最近この方向での実験的研究が，いろいろためされている（たとえば，Usui-Nakauchi-Nakano[14]参照)．

パターン認識におけるニューラルネットワークの応用においては，カテゴリー数が，m 個ある場合には，m 個の出力ユニットを用い，それぞれの出力ユニットに各カテゴリーを対応させ，入力に応じて対応するユニットに1が出力され他のユニットは0が出力されるように学習が行われる．いま，入力の次元が n で中間層は k ($<n$) ユニットからなる3層ネットワークの場合を考えよう．つぎの結果が Gallinari-Thiria-Soulie[15]によって得られている．

《**定理4**》(Gallinari-Thiria-Soulie) すべて線形ユニットからなる3層ネットワークで，中間層のユニット数が級間分散行列の階数 ($<m$) に等しいとき，出力と理想出力の平均二乗誤差を最小化することは，判別分析 (Discriminant Analysis) と等価である．

この定理では，hidden ユニット数がカテゴリー数より小さいとしており，またすべての線形ユニットであるため，実際的な場合を扱っているといえないが，hidden ユニット数が，カテゴリー数より多いシグモイドユニットからなる3層ネットワークの能力は，判別分析以上であることを示していると解釈してよい．また，3層ネットワークで，シグモイドユニットを用いる場合，中間層のユニット数を十分大にすれば，ベイズ識別の能力にいくらでも近づきうることが Asoh-Otsu[16]に述べられているが，これは，不連続関数でも可積分ならば，平均二乗誤差の意味で3層ネットワークで任意の精度で近似できることが定理1からいえ，このことをベイズ識別関数に適用することでただちに導かれる．Asoh-Otsu は中間層のユニット数がカテゴリー数 m より小さい場合に非線形判別分析との関係の研究を試みている．

今後調べなければならないことは，中間層のユニット数が限られた場合，ただしカテゴリー数 m より小さいと限らない場合に，中間層のシグモイド特性がどうきいてくるか，また中間層の内部表現は何かということを理論的に解明することであろう．これによって，ニューラルネットワークと多変量解析との関係がより明らかにされると考えられる．

ともあれ一種の非線形多変量解析手法としての階層型ネットワークの理解は，今後も応用，理論とも進展していくと思われる．この側面は，ユニットの出力関

数をヘビサイド関数から連続なシグモイド関数に置き換えたために生じたといえる.

6. おわりに

この小文では階層型ニューラルネットワーク理論の数学的側面として，関数近似理論および多変量解析との関係について紹介し，今後の展望について述べた.

筆者に与えられたテーマは，数学的な観点から階層型ネットワークの原理的な機能について解説することであるので，実際のニューラルネットの学習アルゴリズムという側面から離れて，上記の近似理論および多変量解析との関連で，厳密に証明されることに内容を絞って解説した．ニューラルネットワークのコンピュータシミュレーションによる知見のみでは扱うタスクによっても相違があるため，数学的に厳密な証明を伴わなければニューラルコンピューティングの学問体系をうちたてることは不可能であると筆者は考える.

多くの先駆者の研究にもかかわらずニューラルネットワークは理論的な学問体系がまだしっかりしていないのが現状である．これはシグモイド関数の非線形性による困難から生ずると筆者は考える．階層型ネットワークは，フィードバックのないシステムのためその機能は写像の近似的実現でしかないにもかかわらず，本解説で述べたようにその範囲のなかでも興味ある事実がわかってきた．しかしまだ理論的には出発点といえる結果しか得られていない．こうした事柄は，階層型ネットワークのみならず，相互結合型ネットワークにおいても基礎的な重要性をもつと考えられる.

今後は，工学のみならず非線形科学全般の中でニューラルネットワーク理論を位置づけて研究していくことが必要と思われる.

謝辞　日頃この方面の研究に関してご支援賜る本学の臼井支朗教授に感謝します.

(1990年12月11日受付)

参考文献

1) D. E. Rumelhart, G. E. Hinton and R. J. Williams: Learning Representations by Error Propagation, In D. E. Rumelhart, J. L. McClelland and the PDP Research Group (eds.), Parallel Distributed Processing, Cambridge, MA: MIT Press, Vol. 1, 318/362 (1986)
2) T. J. Sejnowski and C. R. Rosenberg: Parallel Networks that Learn to Pronounce English Text, Complex Systems, 1, 145/168 (1987)
3) R. P. Lippmann: An Introduction to Computing with Neural Nets, IEEE ASSP Magazine, 4, 4/22 (1987)
4) B. Irie and S. Miyake: Capabilities of Three-Layered Perceptrons, Proc. of ICNN, Vol. 1, 641/648 (1988)
5) K. Funahashi: On the Approximate Realization of Continuous Mappings by Neural Networks, Neural Networks, 2, 183/192 (1989)
6) G. Cybenko: Approximation by Superposition of a Sigmoidal Function, Math. Control Signal Systems, 2, 303/314 (1989)
7) K. Hornik, M. Stinchcombe and H. White: Multilayer Feedfoward Networks are Universal Approximators, Neural Networks, 2, 359/366 (1989)
8) E. B. Baum and D. Haussler: What Size Net Gives Valid Generalization?, Neural Computation, 1, 151/160 (1989)
9) D. L. Chester: Why Two Hidden Layers are Better than One, Proc. of IJCNN, Washington, D. C., January 15-19, Vol. 1, 265/268 (1990)
10) G. W. Cottrell, P. Munro and D. Zipser: Image Compression by Back-Propagation: An Example of Extensional Programming, N. E. Sharkey (ed.), Advances in Cogntive Science (Vol. 3), Norwood NJ, Ablex (in Press)
11) G. W. Cottrelll and P. Munro: Principal Component Analysis of Image Via Back Propagation, SPIE 1001 Visual Communications and Image Processing '88, 1070/1076 (1988)
12) H. Bourlard and Y. Kamp: Auto-Association by Multilayer Perceptrons and Singular Value Decomposition, Biological Cybernetics, 59, 291/294 (1988)
13) P. Baldi and K. Hornik: Neural Networks and Principal Component Analysis: Learning from Examples Without Local Minima, Neural Networks, 2, 53/58 (1989)
14) S. Usui, S. Nakauchi and M. Nakano: Reconstruction of Munsell Space by a Five-Layered Neural Network, Proc. of IJCNN, San Diego, California, June 17-21, Vol. 2, 515/520 (1990)
15) P. Gallinari, S. Thiria and F. F. Soulie: Multilayer Perceptron and Data Analysis, Proc. of IJCNN, San Diego, California, July 24-27, Vol. 1, 391/399 (1988)
16) H. Asoh and N. Otsu: Nonlinear Data Analysis and Multilayer Perceptrons, Proc. of IJCNN, Washington, D. D., June 18-22, Vol. 2, 411/415 (1989)
17) K. Yosida: Functional Analysis, New York: Springer-Verlag (1968)
18) 片山，大山：自己組織逆伝播ニューラルネットの諸特性，信学会春季全国大会予稿集，SD-1-14, 309/310 (1989)
19) 入江，川人：多層パーセプトロンによる内部表現の獲得 —テーブルルックアップ法と違うのか？—，信学技報 89-157, NC 89-15, 33/40 (1989)
20) 船橋：3層ニューラルネットワークによる恒等写像の近似的実現についての理論的考察，電子情報通信学会論文誌A，J73-A-1, 139/145 (1990)

[著者紹介]

船橋賢一（ふなはし けんいち）君

昭和27年2月25日生．昭和55年名古屋大学理学部数学科卒業．60年同大学大学院博士課程修了．同年シャープ(株)に入社，音声情報処理の研究開発に従事．62年ATR視聴覚機構研究所に出向，ニューラルネットワークの研究に従事．平成元年10月シャープ(株)退社後，現在，豊橋技術科学大学情報工学系講師．理学博士．日本音響学会，電子情報通信学会，日本数学会などの会員.