

# 線形回帰分析による多層ニューラルネットの 1 L-7 中間層ニューロン数決定方法

鹿山 昌宏      ○阿部 重夫      武長 寛      諸岡 泰男  
(株)日立製作所      日立研究所

## 1. はじめに

近年ニューラルネットの研究が盛んに行なわれており、実用上の観点からは学習能力に優れたラメルハート型のモデルに期待が大きい。しかし適用上の課題も多く、とりわけネットワークにおける中間層ニューロン数に関しては、試行錯誤的なシミュレーションで決定する場合が大半である。本報では線形重回帰分析を用いて中間層ニューロンの線形関係を抽出することにより、必要十分なニューロン数を与えるアルゴリズムを検討した結果を報告する。

## 2. 実験システム

立案した手法の評価は、ニューラルネットを数字認識に適用したシステム(以下実験システム)を例に行なった。図1に用いたネットワークの構成を示す。画像から抽出した計12個の特徴量を取り込むための入力ニューロンと、0~9の数字を識別するための出力ニューロンを持つ3層構造のネットワークである。200個の代表的なサンプルを教師データとして学習を行ない、収束条件を教師信号に対して±1%の範囲とした。ネットワークの性能を評価するため、1430個の実機データにより認識率を測定した。

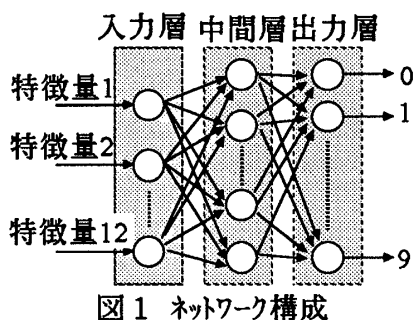


図1 ネットワーク構成

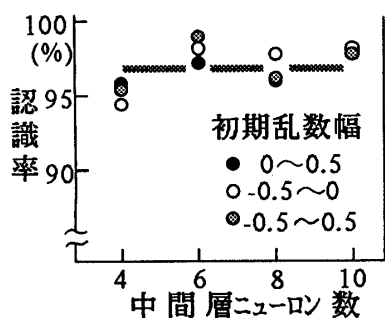


図2 中間層ニューロン数と認識率の関係

## 3. 中間層ユニット数の決定方法

### 3.1 中間層のニューロン数と汎化能力

3層構造で立案した手法の4層構造以上への適用は比較的容易と考えられるため、本研究では3層構造に限定して以下の検討を進める。

中間層のニューロン数は、少な過ぎるとマッピング能力の不足から学習が収束しない。また反対に多過ぎると中間層の描く超空間の曲面が不必要に複雑となり、汎化能力の低下や誤出力を招く恐れがあると考えられる。中間層のニューロン数と汎化能力の関係についてはこれまでも種々報告されているが、これらを総合すると中間層ニューロン数の増大はニューラルネット系の性能を直接向上させるものではないと推察される。図2に実験システムにおいて中間層ニューロン数を変化させたときの認識率を示す。図では3通りの初期乱数幅から学習した結果をそれぞれ示している。本結果でも、ややデータにバラツキはあるものの、認識率のニューロン数依存性は観察されない。一方応用面からは、リアルタイム性との兼ね合いよりネットワーク規模は最小であることが望ましい。中間層ニューロンの必要最小数は、入出力データの個数や非線形性等に依存して決定される値と考えられるが、理論的に値を特定するのは容易ではない。本研究では適当数に値を設定して学習を終了したニューラルネットの中間層ニューロンのふるまいに着目して必要最小数を求めることを考えた。

### 3.2 ニューロン縮減の考え方

一般に過剰な中間層ニューロンを設定すると、それらのいくつかに線形関係が生じ、類似の動作を始めることが報告されている<sup>2)</sup>。これはニューロン数が増えると結果的に一つのニューロンが入出力マッピングのために果たす役割が減少することを意味している。これらより各ニューロンにおける入出力マッピングの負担仕事量を、図3(a)の各部の面積で模式的に表現できる。すなわちマッピングに本来必要な仕事量に対して各ニューロンは、図のような線形関係の程度に応じた重複部分を有しながらその一部を分担して負担していると考えられる。したがって本来必要な仕事量を示す太線内の面

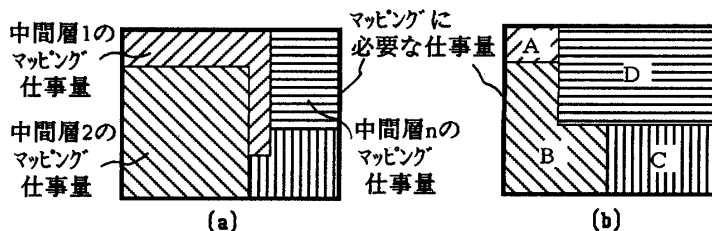


図3 中間層のマッピング仕事量模式図

積を評価できれば、ニューロン数をこれに対応した個数とすることができる。

本研究では線形重回帰分析<sup>3)</sup>を用いた統計的手法によりニューロン間の線形性を評価し、これを基に中間層数を決定することを検討した。

### 3.3 ニューロンの縮減方法

中間層ニューロン数をN個とし、各ニューロンの出力値を $Y_1 \sim Y_n$ とする。学習に用いた200個のサンプルを用い、まずこれらの一つである $Y_i$ を、他の $N-1$ 個の中間層の出力の線形和として、

$$\hat{Y}_i = a_0 + a_1 Y_1 + a_2 Y_2 + a_3 Y_3 + \dots + a_n Y_n$$

(ただし  $a_i Y_i$  を除く)

..... (1)

で近似する。一般に $Y_i$ が他のニューロンと緊密な線形関係にある場合良い精度で近似できるが、そうでない場合近似の精度は悪くなる。定係数  $a_1 \sim a_n$  は良く知られた線形重回帰分析の手法を用いて一意的に決定できる。本研究では $Y_i$ が他のニューロンの出力 $Y_1 \sim Y_n$  ( $Y_i$ を除く)による線形式で表現できる割合の指標として、 $Y_i$ と $\hat{Y}_i$ の重相関係数に着目した。この値を本報では $Y_i * \hat{Y}_i$ で表す。 $Y_i * \hat{Y}_i$ は0~1の値をとり、 $Y_i * \hat{Y}_i$ が1のとき $Y_i$ が他のニューロンの出力で完全に記述できたことを示す。逆に $Y_i * \hat{Y}_i$ が0のとき $Y_i$ と $\hat{Y}_i$ は無相関であることと対応している。

ここで図3(a)の太線で囲まれた部分の面積に相当するニューロン数を、本研究では以下の方法で求めた。簡単のため図3(b)に示す4ニューロンの場合を例に説明する。まず $Y_1$ を $Y_2 \sim Y_4$ の線形和で表した $\hat{Y}_1$ と $Y_1$ の重相関係数を求める。この値を $C_1$ すると $C_1$ は $Y_1$ が $Y_2 \sim Y_4$ で線形に記述できる割合を示している。したがって $1 - C_1$ の値により $Y_1$ の独立分を求めることができ、これは図(b)のA部の面積で模式的に表される。次に $Y_2$ を $Y_3, Y_4$ の線形和で表し、同様の操作を行なう。このとき $1 - C_2$ は図4のB部の面積に相当している。さらに $Y_3$ と $Y_4$ の重相関係数を求め、 $1 - C_3$ としてC部の面積を求める。最後に残った $Y_4$ に関しては $C_4 = 0$

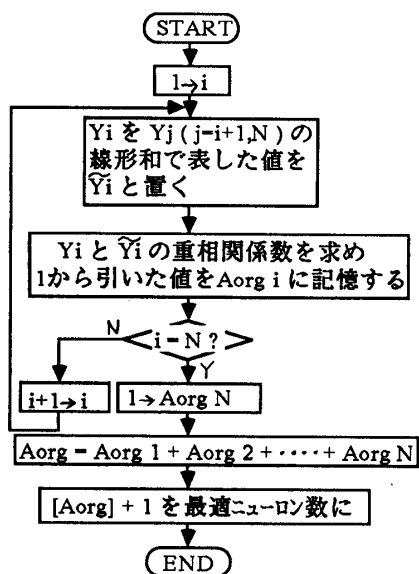


図4 中間層ニューロン数決定アルゴリズム

となり、独立分はD部の面積となる。A~Dの面積を加えると、太線で囲まれた部分の面積となり、これは重複動作分を排除した中間層ニューロン数と対応している。図4に以上を一般化したアルゴリズムを示す。アルゴリズムの最後でニューロン数を整数化している。

### 3.4 実験システムによる評価結果

中間層ニューロン数を4, 6, 8, 10の4通りに設定した実験システムに、開発アルゴリズムを適用して最適ニューロン数を評価した。図5に結果を示す。矢印は縮減の効果を示している。ニューロン数が4のとき評価値は3~3.4であるが、6以上のとき飽和した値となり、設定したニューロン数に依存しなくなる。この値は4前後であり、4~5が中間層ニューロン数の最適値と評価できる。このニューロン数で学習が収れんし、ネットワークの性能(認識率)も低下しないことは、図2より確認できる。実機では中間層を多めに設定したネットワークを用いた1回の学習でニューロン数を評価し、この値を用いた再学習により最適構成のネットワークを得ることができる。

つぎに学習の収束条件を変化させた結果を示す。収束条件を弱くすると超空間の非線形性が薄められるため、必要な中間層ニューロン数が少なくて済むと予想される。図6の評価結果でも、収束条件が緩やかになるに従い、必要なニューロン数が減少している。これより本研究で着目したパラメータは、学習に供した入出力データ間の非線形特性だけでなく、中間層ニューロンが描く超空間の非線形形状に寄与する諸因子の影響をも包含した統計量と考えられる。したがって本アルゴリズムで求まる中間層の値は、いずれの条件で構築されたニューラルネットに対しても統一的に適用可能な評価技法である。

### 4. おわりに

線形重回帰分析を適用した統計的手法により、入出力のマッピングに必要な中間層の個数を評価するアルゴリズムを立案した。

### 5. 参考文献

- 1) D.E.Rumelhart et al, "Parallel Distributed Processing", Vol.1,2 MIT Press
- 2) 関ほか, "ニューラルネットワークによる電力ケーブルの故障診断", 電学論D, Vol.110 No.3
- 3) 芳賀, "回帰分析と主成分分析", 日科技連

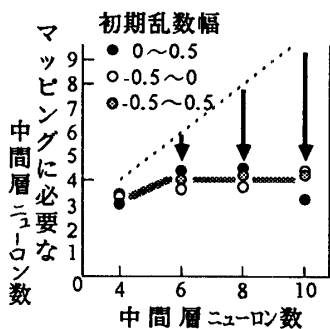


図5 中間層ニューロン数評価結果

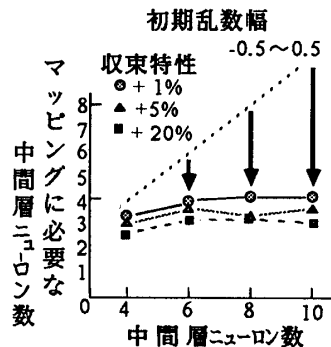


図6 学習収束条件の影響