

A-12 階層形ニューラルネットワークにおける 非線形シナプスの選択学習法

A Selective Learning Algorithm for Nonlinear Synapses in Multilayer Neural Networks

電子・情報システム専攻 房川 実

第 1 章 序論

ニューラルネットワークはさまざまな応用分野に用いられ、優れた能力を示している。そこでネットワークの設計時においては、学習の収束性やネットワーク規模などが重要な問題となってくる。各応用問題においては、それぞれに適した活性化関数が存在し、それらの関数を用いることによって学習の収束性の向上、ネットワーク規模の最小化が可能となる。従来のネットワークでは固定されたシグモイド関数を用い、結合係数の学習で問題を解こうとしていた。しかし、最近では任意の活性化関数を実現できる表現方法に基づいて、問題に最適な活性化関数を学習する方法が提案されている [1]。また、従来のユニットの入力は結合しているユニットからの線形和であったが、これを非線形化することにより伝達特性の自由度を高める方法とその学習方法が提案された [2]。非線形特性を高次多項式で表現する場合、高次元入力に対して項数が非常に多くなり学習の収束性及び回路規模の点で問題があった。

本論文では、遺伝的アルゴリズムに基づき、学習中の内部情報も考慮して高次多項式の全ての項から必要とされる項を選択することにより、学習の収束性の向上、ネットワーク規模の縮小を図る方法を提案する [3],[4]。また、高次多項式の項選択と活性化関数の学習を同時に行なう方法も提案する。関数近似及びパターン分類の問題に対して計算機実験を行ない、提案したネットワーク構成及び学習法の有効性を確かめる。

第 2 章 階層形ニューラルネットワーク

本論文では階層形ニューラルネットワークについて考える。ネットワークは図 1 に示す構造とする。学習には最急降下法に従った方法を用い、出力誤差の 2 乗和によって学習の収束を判断するものとする。

第 3 章 活性化関数

活性化関数とは、ユニットの入力を非線形変換する関数である。従来は固定された 1 個のシグモイド関数を用いていた。しかし、活性化関数を学習可能な複数の基本関数の線形和で表し、各基本関数を学習することによって、最適な活性化関数を実現することが提案された [1]。ガウス関数や三角関数などは複数のシグモイド関数で合成できるため、本研究では基本関数としてシグモイド関数を用いることとする (図 2)。

一般的なシグモイド関数: $f(x)$

$$f(x) = \frac{1}{1 + \exp(-x)}$$

に対して、 a_m, b_m, c_m, d_m を学習可能な係数とすると、提案モデルの活性化関数 $f_p(x)$ は次式で表される。

$$f_p(x) = \sum_m \left\{ \frac{a_m}{1 + \exp(-(b_m x + c_m))} + d_m \right\}$$

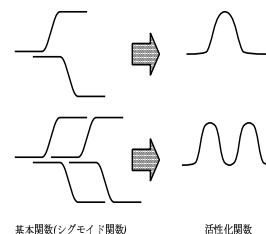
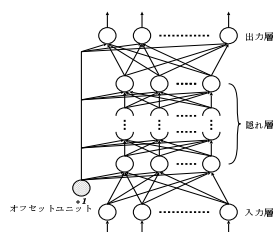


図 1: 階層形ニューラルネットワークの構造 図 2: シグモイド関数による合成

第 4 章 ユニットの非線形特性

従来の階層形ニューラルネットワークでは、ユニットへの入力は下位の層からの重み付き線形和である。このときユニット出力でパターン分類を行なう場合は入力パターンを線形分離することになる。このとき、ユ

ユニットの入出力関係は信号空間の超平面上で $f_p(x)$ を実現している。入力を非線形化することにより、超平面を自由度の高い曲面に拡張することができる。これにより、信号空間における領域形成の自由が高くなり、複雑なパターン分類を少ない素子で実現することが可能となる。非線形の表現方法として、入力の高次多項式を用いる。入力を $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ とすると、ユニットの入出力関係は次式で表される。

$$\begin{aligned} u &= g_p(\mathbf{x}, \mathbf{a}) = \sum_l \sum_i a_{li} x_i^l \\ &= a_0 + a_{11}x_1 + a_{12}x_2 + \dots \\ &\quad + a_{21}x_1^2 + a_{22}x_2^2 + \dots \\ y &= f_p(u) \end{aligned}$$

第 5 章 高次多項式の有効な項の選択法

ユニット入力部の非線形関数は高次多項式で表現すると、例えば、 N 次元入力で 2 次の多項式の項数は $N(N+3)/2$ となる。従って、入力信号が高次元になる場合は項数が莫大となる。さらに、問題に対して全ての項が必要とされる場合は少ない。このため有効な項を選択して、非線形関数を表現する項数を少なくし、学習効率を上げると同時にハードウェアの小型化を考える。

5.1 遺伝的アルゴリズムに基づく選択法

本論文では、高次多項式の有効な項を選択するために、遺伝的アルゴリズム (GA) に基づき、学習中のネットワーク内部情報を考慮し選択する。優性個体の親の選択基準として、出力誤差の 2 乗和を用い、子の形成する遺伝子の選択基準としては、各項の結合荷重を用いる。提案するネットワークモデルを図 3 に示す。高次多項式のブロックで入力の高次項が生成される。項数制限のブロックでは有効な項のみが出力される。隠れユニットの活性化関数は学習可能な関数が用いられる。図 3 のネットワークにおいて、 m 次多項式の出力は一般に次式で表される。

$$x_1^{l_1} x_2^{l_2} \dots x_{N_I}^{l_{N_I}} \quad 0 \leq l_i \leq m \quad l_1 + l_2 + \dots + l_{N_I} = m$$

この項に $A_\alpha = 1$ or 0 をかける。項数制限ブロックの出力として、

$$\hat{x}_i = A_\alpha x_1^{l_1} x_2^{l_2} \dots x_{N_I}^{l_{N_I}} \quad A_\alpha = 1$$

が取り出される。さらに、隠れユニットの出力は次式となる。

$$u_j = \sum_{i=0}^{M-1} w_{ji} \hat{x}_i \quad y_j = f_j(u_j)$$

隠れ層から出力層の関係も同様である。また、 A_α は遺伝子を表し、染色体は次のように表される。

$$C = [A_1, A_2, \dots, A_m]$$

5.2 選択アルゴリズムのフロー

各学習において個体数を K とし、 $A_\alpha = 1$ となる遺伝子の数、すなわち有効な項数を M 個に制限する。従って、染色体における遺伝子の総数を L とすると、このうち M 個を $A_\alpha = 1$, $L - M$ 個を $A_\alpha = 0$ とする。

Step 1

学習開始時においては、初期集団として M 個の $A_\alpha = 1$ となる遺伝子の位置とネットワークの結合荷重をランダムに生成して、 K 個の個体を作る。

Step 2

Step 1 もしくは Step 3 で生成された個体を用い、結合荷重と活性化関数の同時学習を行う。出力誤差が許容範囲内で、選択された項数が十分に低減されていれば学習は終了する。そうでなければ Step 3 に進む。

Step 3

K 個の個体のうち出力誤差の小さい順に $K' (< K)$ 個選択してこれを優性個体とする。 K' 個から2個の個体 G_1, G_2 を組み合わせる。個体の優性度及び遺伝子に対応する結合荷重の重さにより、遺伝子の選択順を決定する。 G_1 が G_2 より優性であるとき、まず、 G_1 の遺伝子の中で $A_\alpha = 1$ かつ結合荷重が最も大きいものが選ばれる。次に、 G_2 から同様な遺伝子が選択される。さらに、 $G_1 \rightarrow G_2 \rightarrow \dots$ とこの操作を M 回行なう。もし、 G_1 と G_2 で選ばれた遺伝子が全て異なるときは M 個の遺伝子が、また、重複がある場合は M 未満の遺伝子が選ばれることになり、この方法は学習中に $A_\alpha = 1$ となる遺伝子の数、すなわち有効な項数が減少する可能性がある。

この Step で、 K 個の個体を生成し、 Step 2 に戻る。

(例題) $L = 6, M = 4$ の例を図4に示す。 G_1, G_2 の順に染色体の優性度が高いとする。次世代で $A_\alpha = 1$ となる遺伝子は $G_1(2番) \rightarrow G_2(1番) \rightarrow G_1(6番) \rightarrow G_2(2番)$ の順に選ばれる。2番目の遺伝子が2回選ばれるために有効な項数は $M = 4$ から3に低減している。このように、より重要な遺伝子が残り、最終的には有効項数の数も最適な数にまで減少すると考えられる。

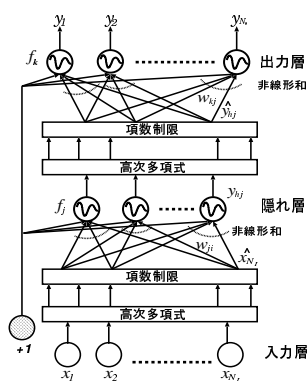


図 3: 提案法のネットワークモデル

$$\begin{array}{l} 1 \quad 2 \\ G_1 = \{0, 1, 1, 1, 0, 1\} \\ 1 \quad 2 \\ G_2 = \{1, 1, 0, 1, 1, 0\} \end{array} \rightarrow G_{1,2} = \{1, 1, 0, 0, 0, 1\}$$

図 4: 項の選択

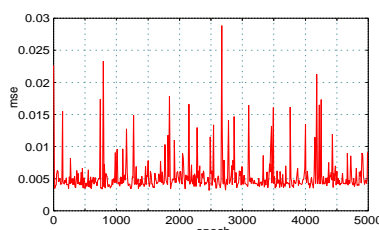


図 6: 関数近似の学習曲線

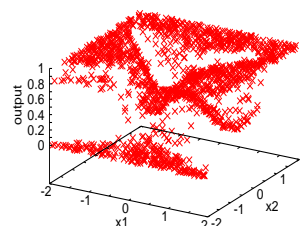


図 5: 目標関数

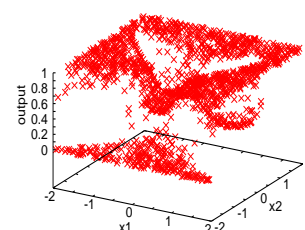


図 7: 学習後の関数

第 6 章 計算機実験 1 -活性化関数を固定-

6.1 関数近似

目標とする関数を図3のネットワークを用いて生成した。隠れ層ユニット数が2個、出力ユニット数が1個である。高次多項式は x^2y, x^4, y^5 を用い、結合係数は乱数とした。目標関数を図5に示す。これを学習するために、5次の高次多項式を用い有効項数 $A_\alpha = 1$ を $M = 6$ とした。各世代の個体数を $K = 10$ とした。学習は第3世代で収束し、選択された項は目標関数と同じであり、最小項数が得られている。学習曲線を図6に、学習後の関数を図7に示す。

6.2 2次元パターン分類

2次元パターン分類の学習データを図8に示す。ここでは、隠れユニット数2個の3層のネットワーク、個体数 $K = 10$ 、有効項数 $M = 8$ とした。図10の学習曲線や図11の学習後の入出力関係の形状から有効に学習されていることが分かる。有効項数は $M = 8$ 個で行なったが、結果として $x^2, y^2, x^4, x^2y^2, y^4$ の5個で収束した。

6.3 収束率の評価

6.2の2次元パターン分類を用いて学習の収束率をGAと比較して評価する。初期個体集団の遺伝子や結合係数を乱数で生成するが、これを10通りについて学習を行ない、その収束状況を調べる。

提案法は10回の試行のうち6回が収束し、選択された項は、いずれも $x^2, y^2, x^4, x^2y^2, y^5$ であった。GAは3回が収束し、選択された項は、 $x^2, y^2, x^4, x^2y^2, y^4, x^4y - x^2, y^2, x^3, xy^2, x^4, x^2y^2, y^4, x^3y^2, x^2y^3, xy^4 -$

$y, x^2, y^2, x^3, x^2y, xy^2, x^4, x^3y, x^2y^2, y^4, x^5, x^3y^2, xy^4$ となり、項の数も6個、10個、13個とバラバラであった。提案法で選択された項は全ての収束解に含まれており、提案法が高い確率で安定に収束することが確認された。

6.4 3次元パターン分類

図11に示すドーナツの内部と外部を分類する問題を取り上げる。学習曲線を図12に、学習後のパターン分類の領域を図13に示す。提案法で項数制限は $M = 12$ 個で行なったが最終的に8個で収束し、 $z, x^2, y^2, z^2, zx, x^3, xy^2, y^2z$ が選択された。この結果から、より高次元のパターン分類へも適用可能であると考えられる。

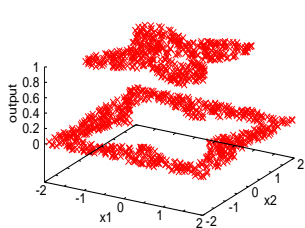


図 8: 2次元パターン分類

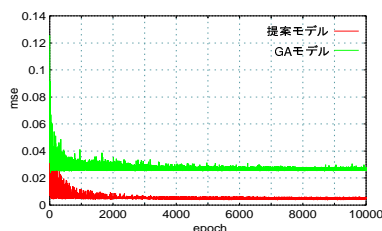


図 9: 提案モデルとGAの学習曲線

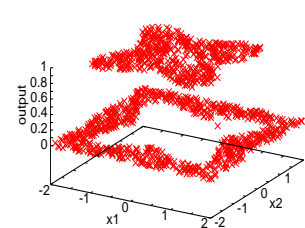


図 10: 学習後の入出力関係

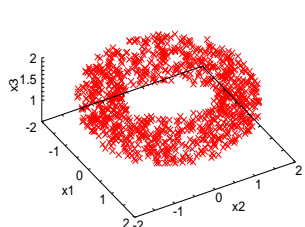


図 11: 3次元パターン分類

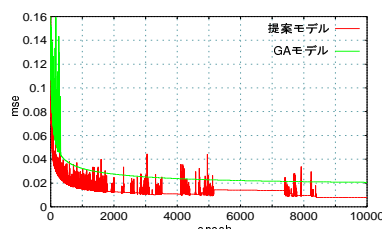


図 12: 提案モデルとGAの学習曲線

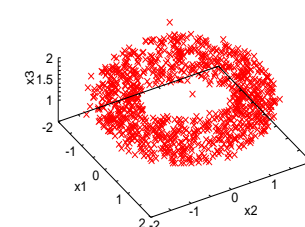


図 13: 学習後のクラス領域

第7章 計算機実験2 -活性化関数を学習-

第6章で取り上げた2次元パターン分類と3次元パターン分類において、活性化関数の学習[1]と有効な項の選択を同時に行なった[2]。2次元パターン分類において、隠れ層ユニット数が2個から1個に減少した。また3次元パターン分類においては、隠れ層ユニット数が5個から3個に減少した。さらに、学習に要する世代数は4世代から2世代に減少した。これらの結果から、有効な項の選択と活性化関数の学習を同時に行なうことにより学習の高速化とネットワーク規模が縮小が可能であることが確認された。

第8章 まとめ

ユニット入力部の非線形化により生成される項の中から、必要とされる項のみを選別する方法を提案した。多くの例題による計算機実験により提案法の有効性が確認でき、活性化関数と同時学習することにより学習の収束性の向上、ネットワークの小型化などが可能となった。

参考文献

- (1) K.Nakayama and M.Ohsugi, "A simultaneous learning method for both activation functions and connection weights of multilayer neural networks", Proc. of IJCNN'98, pp.2253-2257, May 1998.
- (2) K.Nakayama, A.Hirano and .Ido, "A multilayer neural network with nonlinear inputs and trainable activation functions: Structure and simultaneous learning algorithm", Proc. IJCNN'99, pp.1657-1661, July 1999.

口頭発表

- (3) 房川 実, 中山謙二, 平野晃宏: "非線形入力を有するニューラルネットワークの遺伝的アルゴリズムによる学習", 電気関係学会北陸支部連合大会, pp.432, 2000年9月, 金沢
- (4) 房川 実, 中山謙二, 平野晃宏: "階層形ニューラルネットワークにおける非線形シナプスの選択学習法", 信学会, ニューロコンピューティング研究会, 2001年3月, 東京 発表予定