遺伝的アルゴリズムによるニューラルネットワークの構造決定

正員 村田 純 一 (九州大学)
 非会員 中村 紀 幸 (九州大学)
 正員 平澤 宏太郎 (九州大学)
 非会員 田 中 慶 (九州大学)
 非会員 佐々木 美 幸 (九州大学)

Neural Network Structure Design Using Genetic Algorithms

Junichi Murata, Member, Noriyuki Nakamura, Non-member, Kotaro Hirasawa, Member, Kei Tanaka, Non-member, Miyuki Sasaki, Non-member (Kyushu University)

For the purpose of automatic and effective neural network structure design, a designing method is proposed based on Genetic Algorithms (GA). Neural Network structures are encoded on chromosomes in such a way that a number of different but related network structures result by changing a parameter in the decoding process. The GA finds an optimal chromosome that provides good network structures for a family of data sets. When a change takes place in the environment around the neural network, by changing the parameter, we can obtain a new neural network suitable for the new environment without re-running the GA again.

The proposed method is described by using, as an example, a problem of finding good neural network structures for data sets with different noise magnitudes. To obtain an accurate but non-overfitted neural network for the noisy data set, we introduce a relevant fitness function, a method for noise magnitude estimation, and a systematic way to determine the control parameter value of the decoding process. By incorporating these techniques in the network structure designing method mentioned above, we obtain a neural network which has a good generalization ability for each of data sets with different noise magnitudes.

キーワード：ニューラルネットワーク構造，ニューラルネットワークのクラス，遺伝的アルゴリズム，コーディング

1. はじめに

ニューラルネットワーク (NN) は脳を構成する神経回路網を抽象化したシステムであり，非線形近似能力，学習能力を有し，さまざまな分野への応用が広く研究されている。その際，NN の内部構造は NN の性能を左右するにも関わらず，その決定は組合せ最適化問題となるため，現実には試行錯誤と経験的知識によるところが大きい。一方，遺伝的アルゴリズム (GA) は，生物の進化過程を抽象化したアルゴリズムであり，問題を人工的な遺伝子列 (染色体) に表現 (コーディング) し，交叉，突然変異，選択という生存競争に基づく探索原理により，組合せ最適化問題の解を効率的に探索することが可能である⁽¹⁾。そこで NN の内部構造決定に GA を用いる。

NN の内部構造を決定する方法の多くは，与えられた問題や対象に対しては，ある一つの優れた NN 構造を見つけることは可能だが，対象の変化や環境の変化が起きた場合

には，NN の構造自体が変化することを許していないので，良好な結果が得られない⁽²⁾。対象や環境が変化した場合に，もう一度 GA を適用するのは非効率的である。ここでは，階層型ニューラルネットワークを対象として，遺伝子型 (染色体) と表現型 (NN 構造) の間の対応をパラメータによって制御できる遺伝的アルゴリズム (GA) による NN の構造決定法を提案する。この GA は単一の染色体を決定するが，この染色体は単一の NN 構造を表現するのではなく，NN のあるクラスを記述する。したがって，類似の問題群に共通した骨格部分のみからなる NN 構造や，個々の問題に応じて肉付けされた NN 構造など，異なる構造を与えることができる。例題として，それぞれ異なる大きさのノイズを含む複数のデータセットを，過適合による汎化能力の低下を招くことなく近似する問題を用いて，この NN のクラスの構造を決定する機能が有効に働くことを示す。

図1 ネットワーク構造と染色体
Fig. 1. Network structure and chromosome

2. コーディングと接続形成

(2・1) コーディング GA を用いて NN 構造を決定する方法において最も重要なことは、どのように NN の構造を染色体上に表現するかということ、つまりコーディングである。本論文では、問題群の共通部分に相当する NN の骨格の周囲に、個別の問題に対応する肉付け部分を必要に応じて配置する。したがって、骨格部分と対応する肉付け部分の NN 内での距離が近いことが必要である。そこで、本来 2 次元的な構造を持つ NN 構造を、その距離関係を維持したまま 2 次元の染色体に表現することとする⁽¹⁾。これは、GA で一般的な 1 次元染色体では、NN 内で近い要素が染色体上では遠くに配置されてしまい、その相互関係が交叉によって破壊される可能性が非常に高いからである。

図1左の破線で囲んだ階層型 NN の中間層を図右のようにエンコードする。図右の A 内の各マスは各ノードに対応しており、その中 (図中の B) にノードの種類と、そのノードが他のノードと接続を持つ傾向の強さが記される。例えば、図1ではノードの種類 (シグモイド関数をもつノードまたは線形関数をもつノード) を S, L で表し、このノードが後段の S あるいは L のノードと接続を持つ傾向の強さ (以下では接続強度と呼ぶ) を O_S, O_L 、前段のノードとの接続強度を I_S, I_L で表している。接続強度の取りうる値は 0 から m までの整数値とする。

線形ノードは本来 NN 内での信号処理に寄与しないが、図1のように予め 3 層の中間層を用意していても、GA によってある層のノードがすべて線形ノードと決定された場合、中間層は 2 層で良かったことが分かる。このように層数の自動決定を行う場合は線形ノードを考慮することが有用である。

(2・2) 接続形成 図1左のように NN のノードは格子状に配置されているものとし、上下または左右に隣接するノード間の距離を 1 と定義する。任意のノード間の距離は、一方のノードから出発して垂直または水平方向への移

動のみによって他方のノードに至る最短経路の長さで定義する。簡単のためこの節ではノードの種類別 (S または L) を省略して記述する。ある層内のノード i と染色体上に記されたその接続強度 O_i 、このノードより後段のノード j とその接続強度 I_j 、ノード間の距離 r_{ij} 、およびノード番号 i, j に依存しないしきい値 b が与えられた時、

$$\frac{O_i I_j}{r_{ij}} \geq b : \text{接続する}$$

$$\frac{O_i I_j}{r_{ij}} < b : \text{接続しない}$$

に従ってノード i, j 間の接続の有無を決定する。これにより、しきい値 b を変えることで同じ染色体、つまり同じ I_j, O_i の値から、骨格のみからなる NN や複雑に接続された NN といった異なる構造の NN を作成することができる。すなわち、染色体と NN 構造の対応が 1 対多になっており、この対応関係をしきい値 b が制御していることになる。

ここで、ノード間の距離が最小 ($r = 1$) でも接続しないケース、最大 ($r = R$) でも接続するケースがそれぞれありうるという条件を考慮して

$$0 \leq \frac{O_i I_j}{r_{ij}} \leq \frac{m^2}{R} \dots\dots\dots (1)$$

が成り立つ必要がある。さらに l 種類の異なる NN 構造を構成する必要がある場合は、 $O_i \cdot I_j / r_{ij}$ は、少なくとも l 個の異なる値をとる必要があるので $l \leq m^2 / R$ が要請され、これより

$$m \geq \sqrt{lR} \dots\dots\dots (2)$$

とならなければならない。

(2・3) 遺伝的操作 各個体は、図1右の A のようにマスが縦横に並んだ形の染色体で表現され、各個体の染色体のサイズは共通である。交叉は、1 点交叉の単純な拡張であり、図の A に描かれている縦線あるいは横線 1 本を選び、親となる 2 個体を、選んだ線で切断して接合し直すことによって行う。したがって、交叉はノードに対応するマスの内部には影響を与えない。これに対して突然変異は、全ノードのノード種別と接続強度を区別せずにこの中から確率的に一つを選び、その値を変更することによって行う。

3. 構造決定

(3・1) 例題 構造決定の手順は対象とする問題に依存する部分が多いので、例題を用いて説明する。取扱う例題は、異なる大きさのノイズを含む複数種類のデータセットに対し、それぞれを十分な汎化能力をもって表現する NN を構成することである。まずいくつかのデータセットを用いて、それぞれのデータセットにとって適切な NN を与える共通・単一の染色体を、GA によって求める。上記とは別の新しいデータセットに対しては、GA を適用するのではなく、先に得られた染色体と適切なしきい値 b とか

ら NN 構造を決定する。GA 適用時には、適応度の最大化によって適切な NN 構造が得られ、その後はしきい値の設定によって NN 構造の適切さが確保される。ただし、この二つの適切な NN 構造を得る手段の間には一貫性がなければならない。以下に構造決定に必要な要素と手順を述べる。

なお、提案手法の他に想定される適用例としては、例えば、電力やガスの需要予測モデルなどが考えられる。電力やガスの需要は冷暖房需要の多い夏・冬には気温に対して高い依存性を示すが、春や秋には依存性は低くなる。依存性の低い要素まで考慮すると予測精度が低下するので、季節に応じて予測モデルの構造（気温の情報をどの程度まで利用するか）を変更する必要がある。このことは、本手法を用いれば、季節あるいは気温によってしきい値 b を変更することにより可能となる。

(3.2) NN 構造の適切さ—適応度 適切な NN は、汎化能力の高い NN、すなわち、学習用のデータを精度よく近似するだけでなく、未学習のデータあるいは、学習用/未学習を問わず全データを生み出す元となっている源をよく表すものである必要がある。このためには、学習用のデータにだけ適合することを避けるために、NN は未定係数の少ない構造であることが望ましい。そこで、NN 構造の良さの評価として、AIC(Akaike's Information Criterion)⁽⁴⁾ に類似の、学習用データに対する誤差と NN 構造の複雑さの双方を加味した次式の C を用いる⁽⁵⁾。

$$C = \{2n - (\sum_{i=1}^n \sigma_i^2) / \sum_{i=1}^n \sigma_i^2\} + \alpha T \log\left(\frac{1}{pT} \sum_{j=1}^p \|e_j\|^2\right) \dots\dots\dots (3)$$

ここで、 n は NN 内の全信号（NN 内の全ノードが出力する信号）の個数、 e_j は第 j 出力ノードの誤差ベクトル、 p は出力ノード数、 T は学習用データ数であり、 σ_i は学習用データを入力したときに得られる全信号を並べて構成される $T \times n$ 行列 S の特異値である。また、 α は正の定数である。(3) 式の $(\sum_{i=1}^n \sigma_i^2) / \sum_{i=1}^n \sigma_i^2$ は、NN 内の全信号のうちの線形独立な信号の個数に対応している。したがって $2n - (\sum_{i=1}^n \sigma_i^2) / \sum_{i=1}^n \sigma_i^2$ は、NN 内の全信号数 n に線形従属な信号数を加えたものとなり、冗長な信号である線形従属な信号数分だけのペナルティを考慮した NN の複雑さの指標となっている。

NN 構造の評価にあたっては NN の重みの学習が必要となる。これは Back-propagation 法を用いて行う。

GA は複数のデータセットを一括して取扱う。そこで、各データセットに対する NN 構造の個別評価 C を上記にしたがって計算したのち、その値を合計したものを元にこれらの NN 構造を生み出している染色体の適応度を定義する。

(3.3) しきい値の設定 しきい値も、得られる NN が学習用データに対して過適合とならないという意味で適切なものになるように設定する。ここでは、ノイズを含んだデータを対象としているため、ノイズを含まない真の信号

のみを表現できる NN を構成することとする。そこで、NN の記憶容量、つまり NN の接続数を真の信号成分の大きさに比例させることにする。接続数は可能最大接続数 $\times$ 接続率であるので、接続率 P_c を（真の信号成分の大きさ）/（信号全体の大きさ）に比例するように定める。このためには、しきい値と接続率、接続率と真の信号成分の比率の関係、および真の信号成分の比率の測定法について明らかにする必要がある。なお、NN の接続数と記憶容量が正しく対応するように、以下では本質的な信号処理に寄与しない線形ノードは考えず、シグモイド関数ノードのみを取扱うことにする。

(3.4) 真の信号成分の大きさの推定 入力データを x_i 、出力データを y_i ($i = 1, \dots, N$) とし

$$y_i = \eta_i + \varepsilon_i \dots\dots\dots (4)$$

y_i : 全信号

η_i : 真の信号成分

ε_i : ノイズ信号成分

と表せるとする。ノイズ ε_i については、平均 0、分散 σ^2 の確率変数とする。信号の大きさを信号の平均値まわりの変動量と定義すると、望ましい接続率は、 $\bar{\eta}$, $\bar{y}$ を η_i , y_i の平均値として、

$$P_c = K \frac{\frac{1}{N} \sum_{i=1}^N (\eta_i - \bar{\eta})^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2} \dots\dots\dots (5)$$

となる。ただし K は比例定数である。

ここで、未知の $(\eta_i - \bar{\eta})^2$ を既知量による推定値で置きかえることを考える。ここでの基本的考え方は、ノイズの平均は 0 であるので

$$(\text{真の信号の平均値}) \simeq (\text{全信号の平均値})$$

であることを用いて、図 2 のようにデータをいくつかの領域に分割し、その分割区間中での各々のデータの平均値を求め、隣り合う区間の平均値を結ぶ線上に真の信号成分が存在するとする近似である。

y_i の変動量の期待値および各領域ごとの変動量の総和の期待値は、 n_k を k ($k = 1, \dots, K$) 番目の領域中のデータの個数、 $\bar{\eta}^k, \bar{y}^k$ を k 番目の領域中の η_i, y_i の平均値とすると

$$E\left[\sum_{i=1}^N (y_i - \bar{y})^2\right] = E\left[\sum_{i=1}^N (\eta_i - \bar{\eta} + \varepsilon_i - \bar{\varepsilon})^2\right] = \sum_{i=1}^N (\eta_i - \bar{\eta})^2 + (N-1)\sigma^2 \dots\dots\dots (6)$$

図2 真の信号成分の推定

Fig. 2. Estimation of true signal components

図3 2入力の場合の真の信号成分の推定

Fig. 3. Estimation of true signal components for two-input cases

$$E\left[\sum_{k=1}^K \sum_{i=1}^{n_k} (y_i - \bar{y}^k)^2\right] = \sum_{k=1}^K \sum_{i=1}^{n_k} (\eta_i - \bar{\eta}^k)^2 + (N-K)\sigma^2 \quad \dots (7)$$

となる。ここで $\Delta\eta^2 = \sum_{k=1}^K \sum_{i=1}^{n_k} (\eta_i - \bar{\eta}^k)^2$ の推定値が得られれば、これから期待値をデータから得られる量に置換え、未知である η, σ^2 に関する項を消去することにより、(5) 式右辺の推定値が

$$\frac{\sum_{i=1}^N (y_i - \bar{y})^2 - \frac{N-1}{N-K} \left\{ \sum_{k=1}^K \sum_{i=1}^{n_k} (y_i - \bar{y}^k)^2 - \Delta\eta^2 \right\}}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (8)$$

と得られる。

2入力の場合、図3のように、 x^1, x^2 軸それぞれを、データに応じた適切な幅 Δ^1, Δ^2 で分割した各分割区間内のデータの平均値 $\bar{y}^{kl}$ を真の信号の平均値 $\bar{\eta}^{kl}$ とみなすと、周囲のデータの平均値の4点 ($\bar{y}^{k+1,l}, \bar{y}^{k-1,l}, \bar{y}^{k,l+1}, \bar{y}^{k,l-1}$) によって四つの三角形を作り、それぞれの三角形が作る平面上に真のデータ成分 η_i が存在すると近似する。ただし、隣り合う領域がない場合は、水平な三角形を考える。以上より

$$\frac{1}{n_k} \sum_{i=1}^{n_k} (\eta_i - \bar{\eta}^{kl})^2 \simeq \frac{1}{\Delta^1 \Delta^2}$$

$$\int_{\bar{x}_1 - \frac{\Delta^1}{2}}^{\bar{x}_1 + \frac{\Delta^1}{2}} \int_{\bar{x}_2 - \frac{\Delta^2}{2}}^{\bar{x}_2 + \frac{\Delta^2}{2}} \{ \hat{\eta}(x^1, x^2) - \bar{\eta}^{kl} \}^2 dx^2 dx^1 \quad \dots (9)$$

となり、これから $\Delta\eta^2 = \sum_{k=1}^K \sum_{l=1}^L \sum_{i=1}^{n_{kl}} (\eta_i - \bar{\eta}^{kl})^2$ の推定値が得られ、(8),(9) 式から望ましい接続率 P_c を求めることができる。ここで、 $(\bar{x}^1, \bar{x}^2)$ は (k, l) 番目の分割矩形の中心座標、 $\hat{\eta}(x^1, x^2)$ は三角形上にあるとした真の信号の近似値である。

〈3・5〉 NNの接続率としきい値 GAによって染色体を最適化する場合、染色体から具体的なNN構造を導くためにしきい値 b の値を与える必要がある。このしきい値は次の2点を考慮して設定する必要がある。すなわち、(a) 得られるNN構造が真の信号成分の大きさに比例した接続率を持つものとなること、(b) GAによる適応度最大化を著しく制約しないこと、である。(a) は〈3・3〉で述べた内容に相当する。これを実現するには、 b と接続率の間の関係を知る必要がある。しかし、この関係は染色体の中の接続強度 I, O の値によって決定されるため、最適な染色体がGAによって決定される前には、何らかの推定が必要となる。一方、(b)に関連して、例えば大きなしきい値 b を与えると、接続強度 I, O の値が大きくなりすぎない限り、ノード間の接続は行われないことになり、GAが探索するNN構造の範囲を大きく制約することになる。したがって、しきい値 b に対して、可能な接続強度 I, O と距離 r の値の組合せの中で、接続が行われる(行われない)組合せがどの程度存在するかを推測し、これに基づいて、GAによる最適化を大きく制約しないような b を設定する必要がある。この接続が行われる組合せの割合の推定値は、しきい値 b を与えたときの接続率の推定値に他ならないので、(1)の目的のためにも使用できる。

可能な I, O, r の組合せの中で、与えられたしきい値 b に対して接続が行われる組合せの割合 $P_c(b)$ は以下のようにして推定することができる。接続強度のとり値は実際には整数値であるが、ここでは計算の都合上、連続値をとると考える。可能な値をすべて平等に取扱うことは、 I, O を、出現確率が $p(I) = 1/m, p(O) = 1/m$ となる一様分布に従うと仮定することに相当する。

まず、距離 r 、しきい値 b が与えられたときに接続される確率 $P_c(b, r)$ は

$$P_c(b, r) \simeq \begin{cases} \int_{\frac{br}{m}}^m \int_{\frac{br}{m}}^m p(I)p(O)dO dI & (m \geq \frac{br}{m}) \\ 0 & (m < \frac{br}{m}) \end{cases} = \begin{cases} 1 - \frac{br}{m^2} (1 - \log \frac{br}{m^2}) & (m \geq \frac{br}{m}) \\ 0 & (m < \frac{br}{m}) \end{cases} \quad (10)$$

と表せる。また、 n 個のノードが2層に並んでいる場合の可能最大接続数は n^2 であるので、全接続中距離 r の接続の占める割合 $P_r(r)$ は

$$P_r(r) = \begin{cases} \frac{1}{n} & (r = 1) \\ \frac{2n-2r+2}{n^2} & (r \geq 2) \end{cases} \quad \dots (11)$$

図4 $\widetilde{P}_c(b)$ による $P_c(b)$ の近似
Fig. 4. Approximation of $P_c(b)$ by $\widetilde{P}_c(b)$

と表せる。しかし確率密度関数 $P_r(r)$ は $\int_2^n P_r(r)dr + P_r(1) = 1$ を満たさなければならないので

$$P_r(r) = \frac{2n-2r+2}{n} \times \frac{n-1}{n-2} \quad (r \geq 2, n > 2) \quad (12)$$

という補正を行う。そこで、 $P_c(b, r)$ を r について平均化したものを $P_c(b)$ とすると

$$P_c(b) \simeq \sum_{r=1}^n P_c(b, r) P_r(r) \quad (13)$$

と表せる。

この $P_c(b)$ に基づいて、上記の (a), (b) を満足するようにしきい値 b を設定する。実際には、接続率が希望通りになるようにしきい値を設定するので、(13) 式の逆関数を用いる方が便利である。この逆関数は存在するが解析的に求めることができないので、(13) 式を次のように近似する。

$$\widetilde{P}_c(b) = Ae^{-kb} \quad (14)$$

ただし

$$k = \frac{\log P_c(\frac{m^2}{4}) - \log P_c(1)}{1 - \frac{m^2}{4}}, \quad A = e^k P_c(1)$$

である。 $m=6, n=10$ の場合について $P_c(b)$ と $\widetilde{P}_c(b)$ を図4に示す。図よりこの近似は妥当であることが分かる。これより接続率 P_c が与えられたとき b は

$$b \simeq -\frac{1}{k} \log \frac{P_c}{A} \quad (15)$$

より求めることができる。

(3.6) NN 構造決定の手順 手順は大きく二つに分かれる。手順Iでは、複数のNNの適切な構造を規定することのできる単一の染色体をGAによって求め、手順IIでは、得られた染色体を用いて新しいデータセットに対するNN構造を決定する。

手順Iで求める染色体は、しきい値 b から適切なNN構造への写像関係を規定するものである。したがって、適切な

染色体は、複数の (しきい値, NN 構造) の組に基づき、これらを補間するものとして求める必要がある。そこで、この手順では、ノイズの大きさの異なる二つのデータセット (データセット1, データセット2) を用いる。

手順Iの具体的な内容は以下の通りである。

I-1: 二つのデータセットに対して、(8)式に基づいて、真の信号成分の比率を推定する。設定する接続率 P_c は、比例係数を K として、この真の信号成分の比率に比例することになる。

I-2: (8)式中の K を変えることにより、GAによる最適化を制約しないように望ましい接続率を設定する。すなわち、データセット1, 2に対する望ましい接続率を P_{c1}, P_{c2} ($P_{c1} \leq P_{c2}$ と仮定) としたときに、 $P_{c1}, P_{c2} - P_{c1}, 1 - P_{c2}$ がそれぞれ過度に小さくならないように設定する。これは、両データセットのいずれに対しても接続されるノード対、データセット2に対してのみ接続されるノード対、両データセットともに接続されないノード対の割合に相当しており、これらがいずれも適当に高い確率で生起すれば、GAによる探索範囲を著しく制約していないことになる。

I-3: (15)式を用いて、各データセットに対して望ましい接続率を与えるしきい値 b_1, b_2 を求める。

I-4: GAを用いて適応度が最大となる単一の染色体を決定する。適応度は、染色体に各しきい値 b_1, b_2 を与えて得られる二つの構造のNNの学習を、対応するデータセットを用いて行い、その結果に基づいて、各データセットに対して(3)式を計算することによって得る。

上記の二つとは異なる新しいデータセットに適したNN構造を求める手順IIは、以下のステップから構成される。

II-1: 新しいデータセットの真の信号成分の比率を推定し、手順Iで設定した比例係数 K を用いて望ましい接続率を求め、(15)式からしきい値を決定する。

II-2: 決定されたしきい値を染色体に適用して、NN構造を得、学習を行う。

4. シミュレーション

データの発生源を

$$y = \sin(x_1) + \log(x_2) \quad (16)$$

$$(0 \leq x_1 \leq 2\pi, 1 \leq x_2 \leq 6)$$

とし、これに大きさの異なる以下のノイズが重畳された3種類のデータセットを用いる。

ノイズ小: $-0.01 \sim 0.01$ の一様乱数

ノイズ中: $-0.20 \sim 0.20$ の一様乱数

ノイズ大: $-0.35 \sim 0.35$ の一様乱数

NNの構造は2入力1出力、中間層は2層で1層あたり最大 $n=10$ 個のノードを持つものとする。手順Iではノイズ小のデータセットとノイズ大のデータセットを用い、得

表 1 接続率・しきい値・接続数

Table 1. Connectivity rates, thresholds and numbers of connections

対象データセット	設定接続率 P_c	しきい値 θ	接続数
ノイズ小	0.97	0.26	43
ノイズ大	0.72	0.98	33
ノイズ中	0.82	0.68	35

られた染色体から手順 II に従ってノイズ中のデータセットに適した NN 構造を得るシミュレーションを行った。手順 I における GA の設定は複数回の試行から選んだ適切なものを用いた。その内容は以下の通りである。

- 初期集団の個体数 10
- 交叉によって生み出される個体数 20
- 交叉までに生存し得る個体数 10
- 突然変異率 0.05
- ネットワークの学習回数 300 回
- 終了条件の世代数 200
- 適応度中の係数 $\alpha = 0.3$

また、層間のノードの最大距離 R は 10 で、同一の染色体で少なくとも 3 種類の異なる NN 構造を表現したい ($l = 3$) ので、(2) を満たすように接続強度の最大値 m を 6 と設定した。

手順 I で得られたノイズ小とノイズ大のデータセットに対する NN 構造と、その出力を図 5～8 に示す。ここでは真の信号成分の比率と接続率の比例係数を $K = 1$ と設定した。設定した接続率としきい値、および得られた NN 中の接続数の値を表 1 に示す。図 9, 10 には手順 II を適用して得られたノイズ中のデータセットに対する NN 構造と NN 出力を示す。ここでの目的は、データの発生源の性質を正しく表現できる汎化能力の高い NN を構成することである。したがって、データ中のノイズに惑わされずに真の信号成分に近い NN 出力が得られることが望ましい。図 8, 10 から、ノイズが大きくなっても、得られた NN はノイズに大きく惑わされることなく、滑らかな出力を与えていることが分かる。これは、図 7, 9 に示されているようにノイズが大きくなるにつれて、図 5 に示されるような複雑な NN 構造から、枝葉が取り除かれたノード間接続の少ない簡潔な NN 構造となっていることが寄与している。なお、手順 II-1 において、(15) 式ではなく、GA から得られた染色体を用いて、希望の接続率を与えるしきい値を求めて使用した場合も、ノイズ中のデータセットに対して接続数が 37 と、表 1 に示した結果とほぼ同様の NN が得られた。これから、染色体中の I, O の値を一様分布と仮定して接続率としきい値の関係を求めることが、大きな誤差を生じさせるものではないことが分かる。

次に、汎化能力を定量的に評価するために、真の信号に対する各 NN 出力の誤差を計算した。比較対象として以下のような NN に対しても同様の誤差を計算し、学習用データに対する誤差と合わせて、表 2 にまとめて示した。(A) 各

図 5 ノイズ小のときの NN 構造
Fig. 5. NN structure for data with small noise

図 6 ノイズ小のときの NN の出力
Fig. 6. Output of NN for data with small noise

図 7 ノイズ大のときの NN 構造
Fig. 7. NN structure for data with large noise

図8 ノイズ大のときのNNの出力
Fig. 8. Output of NN for data with large noise

図9 ノイズ中のときのNN構造
Fig. 9. NN structure for data with medium noise

図10 ノイズ中のときのNNの出力
Fig. 10. Output of NN for data with medium noise

表2 NN出力の誤差 (上段:学習誤差, 下段: 真の信号に対する誤差)

Table 2. NN output errors (upper:training errors, lower :errors on true signals)

対象データセット	提案手法によって変化した構造のNN	最適構造のNN (A)	固定構造のNN (B)
ノイズ小	7.52×10^{-5} 6.42×10^{-5}	6.41×10^{-5} 4.79×10^{-5}	6.41×10^{-5} 4.79×10^{-5}
ノイズ大	1.77×10^{-2} 4.42×10^{-3}	1.87×10^{-2} 6.50×10^{-3}	1.71×10^{-2} 6.94×10^{-3}
ノイズ中	7.75×10^{-3} 1.54×10^{-3}	7.09×10^{-3} 1.44×10^{-3}	7.28×10^{-3} 1.60×10^{-3}

データセットごとにGAを適用して得た、(3)式のCが最小となるという意味で最適な構造を持ったNN。(B)データセットの状況(ノイズの大きさ)が変化してもNN構造を変えないケースとして、上記(A)で得たノイズ小のデータセット用の構造のNNをノイズ中、大のデータセットに対しても使用し、結合重みの値のみを各データセットを用いて学習して得たNN。

個別に最適化されたNNと比べると、真の信号に対する誤差の悪化は30%程度であり、提案手法がまずまずの結果を与えていることが分かる。特に、GAによる最適化を経ずに、しきい値bの設定だけで構造を決定したノイズ中のデータセットについても良好な結果が得られており、問題群に対応した複数のNN構造を産み出す単一の染色体を求める、という本手法の機能が適切に働いていることが分かる。また、(B)の固定した構造のNNと比較すると、(B)において最適構造が用いられたノイズ小のデータセットの場合を除いて、提案手法によって得られたNNは、真の信号に対する誤差が小さく、汎化能力が向上していることが分かる。

5. おわりに

階層型ニューラルネットワークを対象として、遺伝子型(染色体)と表現型(NN構造)の間の対応をパラメータによって制御できる遺伝的アルゴリズムを用い、NNの適用対象や使用環境の変化に応じた適切なNN構造を決定する方法を提案した。遺伝的アルゴリズムによって最適化される染色体は、単一のNN構造を表現するのではなく、NNのあるクラスを記述する。したがって、類似の問題群に共通した骨格部分のみからなるNN構造や、個々の問題に応じて骨格に肉付けされたNN構造などの、異なる構造を与えることができるものである。例題として、それぞれ異なる大きさのノイズを含む複数のデータを、過適合による汎化能力の低下を招くことなく近似する問題を用いて、提案手法が有効に機能することを示した。

(平成9年8月11日受付, 同10年3月10日再受付)

文 献

- (1) 北野宏明 : 遺伝的アルゴリズム, 産業図書 (1993)
- (2) H.Kitano : Designing Neural Networks Using Genetic Algorithms with Graph Generation System, *Complex Systems*, 4, 4 (1990)
- (3) J.Murata, K.Tanaka and K.Hirasawa : Determination and Adaptive Alteration of Artificial Neural Network Structures by a Genetic Algorithm with a Controlled Genotype-Phenotype Mapping, *Proc. of 1996 IEEE Int. Conf. on Systems, Man, and Cybernetics*, 1690-1695 (1996)
- (4) 久米 均, 飯塚悦功 : 回帰分析, 岩波書店 (1987)

村 田 純 一 (正員) 1986年3月九州大学大学院工学研究科博士後期課程電気工学専攻修了。1986年4月九州大学工学部助手。1988年4月同大学工学部助教授。1996年4月同大学大学院システム情報科学研究科助教授。現在に至る。電気学会, 計測自動制御学会, システム制御情報学の各会員。主として学習, 自己組織化システムに関する研究に従事。工学博士。

中 村 紀 幸 (学生員) 1996年3月九州大学工学部電子工学科卒業。1998年3月九州大学大学院システム情報科学研究科電気電子システム専攻修士課程修了。1998年4月三菱電機(株)勤務。

平 澤 宏太郎 (正員) 1964年九州大学工学部電気工学科卒業。1966年3月九州大学大学院工学研究科電気工学専攻修了。1966年4月(株)日立製作所入社日立研究所勤務。1991年8月大みか工場勤務。1992年12月九州大学工学部教授。1996年4月同大学大学院システム情報科学研究科教授。現在に至る。電気学会, 計測自動制御学会, 情報処理学会, 米国IEEEの各会員。工学博士。

田 中 慶 (非会員) 1994年3月九州大学工学部電気工学科卒業。1996年3月九州大学大学院工学研究科電気工学専攻修士課程修了。1996年4月九州電力(株)勤務。

佐々木 美 幸 (非会員) 1997年3月九州大学工学部電気工学科卒業。1997年4月松下電器産業(株)勤務。