

## 退化現象を導入した遺伝的アルゴリズム $GA^d$ による構造最適化

高濱 徹行<sup>†</sup>      阪井 節子<sup>††</sup>      市村 匠<sup>†</sup>      磯道 義典<sup>†</sup>

Structural Optimization by Genetic Algorithm with Degeneration ( $GA^d$ )

Tetsuyuki TAKAHAMA<sup>†</sup>, Setsuko SAKAI<sup>††</sup>, Takumi ICHIMURA<sup>†</sup>,  
and Yoshinori ISOMICHI<sup>†</sup>

あらまし 退化を伴う遺伝的アルゴリズム  $GA^d$  を提案する。退化とは、生物において何らかの器官や組織が機能・構造などを失い、次第に衰退・縮小することをいう。本研究では、損傷遺伝子をもつ個体は、その遺伝子に対応する形質の発現が不完全になったり、発現しなくなるため、退化現象を起こすと仮定する。損傷遺伝子とは、環境の影響により突然変異を起こし、正常な遺伝子が不可逆的に変化した遺伝子である。 $GA^d$  では、遺伝子型から表現型への変換関数を損傷遺伝子に依存する関数として定義することにより、退化現象を表現する。また、生存に有効ではない遺伝子は、損傷しても個体の環境への適合度が低下しないと考えられる。これを利用し、完全に損傷した遺伝子は削除されたとみなすことにより、有効性の低い遺伝子を削減する機能を実現する。この機能を用いていくつかのモデルにおけるパラメータ構造を最適化することにより、 $GA^d$  がパラメータ数を削減できる汎用的で効率の良い方法であることを示す。

キーワード 退化, 損傷遺伝子, 遺伝的アルゴリズム, 構造最適化, AIC

### 1. ま え が き

モデル推定とは、観測データなどの各種データに含まれる変数間の関係を同定するモデルを決定することである。モデルは、モデルを記述するパラメータの数とパラメータの意味付けとを規定するパラメータ構造及びパラメータ値により特徴づけられる。例えば、階層型ニューラルネットワークでは、階層数、各階層のニューロン数がパラメータ構造に対応し、結合荷重やしきい値がパラメータ値に対応する。

モデルを決定する際、パラメータ数を大きくすれば、観測データに対する近似誤差を小さくすることが可能であるが、そのモデルは必ずしも未知のデータに対して良い近似モデルとはならない。したがって、モデル推定では、パラメータ値の最適化だけではなく、パラメータ構造の最適化を行うことが非常に重要である。

構造最適化の研究は、このパラメータ構造の最適化

に焦点を当てた研究である。構造最適化の研究には以下のようなものがある。

(1) いくつかのパラメータ構造において近似誤差が最小となるようにパラメータ値を最適化し、何らかの評価基準に基づいてその中から最適なものを選択する方法 [1]。この方法では、近似誤差にパラメータ構造の複雑さを加味した評価尺度である AIC (Akaike Information Criterion) [2], [3], MDL (Minimum Description Length) Principle [4], [5], GPE (Generalized Prediction Error) [6] などが利用される。

(2) パラメータ構造を変化させる方法を組み込み、パラメータ値を最適化しながらパラメータ構造を最適化する方法。パラメータ数を増加させる生成法、パラメータ数を減少させる削除法、両者を組み合わせた方法がある。モデル記述系としては、ニューラルネットワーク [7] ~ [13], RBF (Radial Basis Function) ネットワーク [14], ファジー推論ルール [15] ~ [17] が対象とされている。

(3) パラメータ値を最適化する過程でパラメータを縮退させ、パラメータ数を削減する方法。石川 [18], [19] は、ニューラルネットワークにおいて近似誤差に結合荷重の絶対値を小さくする項を加え、結合荷重を縮退させる方法を提案している。

<sup>†</sup> 広島市立大学情報科学部知能情報システム工学科, 広島市  
Faculty of Information Sciences, Hiroshima City University,  
3-4-1 Ozukahigashi, Asaminami-ku, Hiroshima-shi, 731-  
3194 Japan

<sup>††</sup> 広島修道大学商学部経営学科, 広島市  
Faculty of Commercial Sciences, Hiroshima Shudo Uni-  
versity, 1-1-1 Ozukahigashi, Asaminami-ku, Hiroshima-shi,  
731-3195 Japan

(1) は様々の記述系に適用可能であるが、多数の構造においてパラメータを最適化する必要があり、計算量が多くなるという問題点が指摘されている(2)、(3)の研究は記述系ごとの特徴を利用したものが多く、様々な記述系への適用可能性(汎用性)については十分でない面がある。

これに対して、我々は広い範囲の記述系に対応できる、汎用性の高い構造最適化法 MGGA (Genetic Algorithm with Mutant Genes)[20] を提案した。MGGA は通常の遺伝子に加えて無効な変異遺伝子 (mutant genes) を導入し、変異遺伝子に対応する形質の発現を抑制することによりパラメータを縮退させる方法である。しかし、MGGA には以下の問題がある。

(1) 安定性: MGGA では正常遺伝子から変異遺伝子への一方向の変異を仮定しているため、長い世代にわたって最適化を行うと、パラメータを削減しすぎて誤差を十分に減少できない場合がある。

(2) 解の表現範囲と精度: MGGA は遺伝子の値と変異遺伝子かどうかを示す情報とを対にして 0/1 のビットストリングで表現された個体を用いており、モデル記述系で通常利用されている実数値を表現するために固定小数点表現を用いている。しかし、固定小数点表現では、パラメータの表現可能範囲を指定しておく必要があるため、事前に適切なパラメータの表現範囲を決めるのが困難な場合には対応しにくい。また、一般にパラメータの表現範囲は広い方が望ましいが、固定小数点表現においてビット数を増やさずに表現範囲を大きくとるためには精度を落とさねばならないという問題がある。

(3) 実行効率: MGGA では個体の適合度を評価するたびごとにビットストリングを実数に変換して最適化を行うため実行時間が掛かる。

これらの問題を解決するため、本研究では新しい構造最適化手法  $GA^d$  (Genetic Algorithm with Degeneration)[21] を提案する。 $GA^d$  では、MGGA を以下のように拡張することにより、問題の解決を図る。

(1) 損傷度の導入: MGGA では一つの遺伝子は有効な正常遺伝子が無効な変異遺伝子かの二つの状態のみをとることができた。しかし、遺伝子はその中間的な状態もとり得ると考えられるため、 $GA^d$  では有効な遺伝子と無効な遺伝子の中間的な性質を有する遺伝子を仮定する。このため、遺伝子がどの程度損傷し、どの程度正常な状態から変異したかを示す尺度である損傷度を導入する。MGGA においては遺伝子が変異

しているかどうかを示す  $\{0, 1\}$  の情報を用いていたが、損傷度はこれを区間  $[0, 1]$  に拡張したものと考えることができる。

(2) 遺伝子損傷・修復の導入: MGGA では正常な遺伝子が無効になるという一方向の単純な状態変化を仮定した。しかし、遺伝子は完全に無効になるのではなく、ある程度損傷したり逆にある程度修復されることもある。 $GA^d$  ではこのような状況を模倣するために、損傷度の変化を確率的に実現する確率関数を導入する。これにより、MGGA の問題(1)が解決される。

(3) 実数型遺伝子の導入: MGGA では一つの遺伝子が二つの状態のみをとるという制限のため、遺伝子に実数値を直接用いることは現実的でなかった。しかし、 $GA^d$  では損傷度を実数値で表現するため、遺伝子も実数値で表現することは自然であり、効果的であるとえられる。本論文では、完全に有効な状態にあるときに遺伝子が有する遺伝子型である正常型と損傷度の対で一つの遺伝子を構成する。これにより、MGGA の問題(2)、(3)が解決される。

本論文では、 $GA^d$  が MGGA と同じく汎用性の高い構造最適化手法であり、MGGA よりも安定かつ効率よく構造最適化が可能となることを示す。2. で、 $GA^d$  を定義し、3. で単純な最適化問題により  $GA^d$  を調べる。4. で多項式モデル、5. でニューラルネットワークのパラメータ構造の学習に関する数値実験を行い、6. でその結果について検討する。7. はまとめである。

## 2. $GA^d$

自然界においては、進化の過程で不要な器官などを失う退化と呼ばれる現象が知られている。これは、その器官に関与する遺伝子が何らかの原因で損傷し、正常な遺伝子とは異なる損傷した遺伝子(以下損傷遺伝子と呼ぶ)が子孫に伝わったためと考えられる。損傷遺伝子を有する個体では、その遺伝子に対応する形質の発現が不完全になったり、発現しなくなるなどの現象が起きる。

$GA^d$  では、遺伝子型から表現型への変換関数を損傷遺伝子に依存する関数として定義することにより、このような現象を実現する。更に、正常な遺伝子から損傷遺伝子への確率的な変化を導入することにより、生存に寄与しない有効性の低い遺伝子を削減する機能を実現する。この削減機能を利用することにより、複数のパラメータにより記述されるモデルにおいて、不

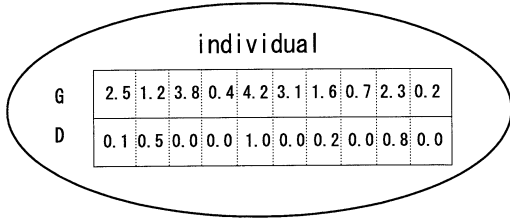


図 1  $GA^d$  における個体の表現  
Fig. 1 Representation of an individual in  $GA^d$ .

要なパラメータを削減し、パラメータ数を最適化する、すなわち構造最適化に应用することができる。

2.1 損傷遺伝子と個体の表現

$GA^d$  において重要な働きをする損傷遺伝子は、突然変異により正常な遺伝子と異なる状態になった遺伝子である。突然変異には置換、挿入、欠損など様々な種類があるため、損傷遺伝子がとり得る状態をあらかじめ想定することは困難である。そこで本研究では、正常な遺伝子が本来有している状態である正常値 (normal value) とその遺伝子がどの程度損傷しているかを示す損傷度 (damaged rate) により、遺伝子を表現する。損傷度は区間  $[0, 1]$  上の値をとり、正常な遺伝子は損傷度 0、形質を全く発現しなくなった遺伝子は損傷度 1 とする。

$GA^d$  では各遺伝子は正常値と損傷度の対で表現されるため、遺伝子型  $G^d$  は、以下のように表現される (図 1 参照)。

$$G^d = (g_1, d_1)(g_2, d_2) \cdots (g_\ell, d_\ell) \quad (1)$$

ただし、 $g_i$  は  $i$  番目の遺伝子の正常値、 $d_i$  は損傷度、 $\ell$  は遺伝子数である。

遺伝的アルゴリズム (GA) における各個体は環境に適応している度合を示す適合度をもつ。遺伝子型から表現型への変換関数を  $h$ 、適合度関数を  $f$  と仮定すれば、個体の適合度は  $f(h(G))$  で与えられる。ただし、 $G$  は正常値の並び  $G = g_1g_2 \cdots g_\ell$  である。これに対して、 $GA^d$  においては、正常値の並びが同じであっても損傷度により表現型が変化することになる。したがって、表現型への変換関数  $h^d$  は正常値の並び  $G$  と損傷度の並び  $D$  に依存し、個体の適合度は  $f(h^d(G, D))$  で与えられることになる。ただし、 $D = d_1d_2 \cdots d_\ell$  である。

2.2 不可逆的突然変異と交叉

損傷度が小さい遺伝子は、生物のもつ遺伝子修復機

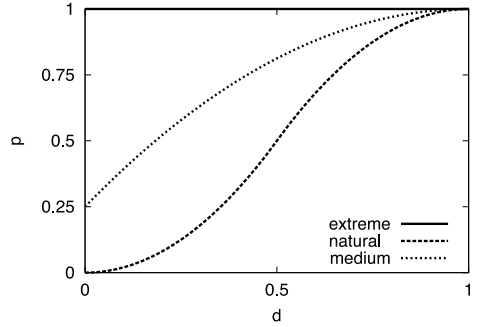


図 2 損傷度増加確率関数  
Fig. 2 Function of probability that a damaged rate increases.

能により、正常な遺伝子に修復されることが多い。しかし、損傷度が大きい遺伝子は、正常な遺伝子に戻らず、逆に損傷が進み、その働きを完全に失ってしまうことがある。このように損傷度の変化にはある種の不可逆性が存在する。 $GA^d$  では、損傷度は不可逆的突然変異により変化し、損傷度が大きくなれば修復が非常に困難になるという方向的な変化を起こし、正常値は GA と同様に可逆的突然変異により可逆的に変化すると仮定する。

本研究では、損傷度の変化を確率的にとらえることを提案する。損傷度が増加するとき、損傷度が増加する確率は、現在の損傷度  $d \in [0, 1]$  から確率  $p \in [0, 1]$  への写像、損傷度増加確率関数  $P_{dam}$

$$P_{dam} : d \in [0, 1] \rightarrow p \in [0, 1] \quad (2)$$

によって  $p = P_{dam}(d)$  で与えられる。減少する確率は、 $1 - p$  となる。

損傷度は、小さい場合ほどより修復され減少する確率が高くなり、大きい場合ほどより損傷が進み増加する確率が高くなるという振舞いが自然であると考えられる。しかし、本論文では様々な可能性を探るために、以下の 3 通りの関数について調べる (図 2 参照)。

$$\text{extreme}(d) = 1 \quad (3)$$

$$\text{natural}(d) = \begin{cases} 2d^2 & (d \leq 0.5) \\ 1 - 2(1-d)^2 & (d > 0.5) \end{cases} \quad (4)$$

$$\text{medium}(d) = 1 - 0.75(1-d)^2 \quad (5)$$

式 (3) は、遺伝子は常に損傷を受け、修復はしないことを表す最も極端な関数 (extreme) である。式 (4) は、損傷度が 0.5 以下の場合には損傷度が小さいほど

修復される確率が大きく、0.5 以上の場合には損傷が大きいほど損傷する確率が大きくなる、より自然な関数 (natural) である。式 (5) は、(3) と (4) の中間であり、損傷度が少し大きくなると損傷する確率がかなり大きくなる関数 (medium) である。(4) の場合よりも積極的にパラメータ数を削減する場合に有効であると考えられる。

なお、extreme は MGGA と同様に正常な遺伝子から無効な遺伝子への一方向の変化を実現している。 $GA^d$  はその他の様々な変化を実現可能なため、MGGA の拡張となっていることは明らかである。

### 2.3 $GA^d$ のアルゴリズム

$GA^d$  のアルゴリズムの概要を以下に示す。

(1) 初期集団の生成: 初期個体集団をランダムに生成する。初期個体の損傷度については、通常区間  $[0, 1]$  でランダムに決めるが、 $P_{dam}(0) > 0$  の場合には MGGA と同様にすべて 0 とすることも可能である。

(2) 選択: 次世代の個体の親となる個体を選択する。構造最適化では推定誤差及び構造に対する評価関数を最小化することが多い。本研究では、個体の選択方法として、頑健な方法として知られている線形ランキング選択 [22] を用い、推定誤差を最小化する。個体集団中で関数値の小さい順にランク付けを行い、個体  $i$  のランクが  $r_i$  ( $r_i = 1, 2, \dots, N$ ) であったと仮定する。このとき、個体  $i$  の選択確率  $p_i$  を以下のようにする。

$$p_i = \frac{1}{N} \left( \eta^+ - (\eta^+ - \eta^-) \frac{r_i - 1}{N - 1} \right) \quad (6)$$

ただし、 $N$  は個体集団サイズ、 $\eta^- = 2 - \eta^+$ 、 $\eta^+$  は最大期待値 (maximum expected value) であり、最良の個体が平均的個体より何倍選択されるかを指定する、区間  $[1.0, 2.0]$  上の値である。

(3) 交叉: 交叉確率  $P_c$  で親の交叉を行い、子を生じさせる。MGGA と同様に正常値と損傷度の対が同時に子に継承される。なお、交叉しない場合には親がそのまま子になったとみなす。

(4) 可逆的突然変異: 生成された子の正常値に対して、可逆的突然変異確率  $P_{rm}$  で突然変異を起こす。

(5) 不可逆的突然変異: 生成された子の損傷度に対して、不可逆的突然変異確率  $P_{im}$  で損傷度を変化させる。損傷度が変化する場合、損傷度が増加する確率は  $P_{dam}$ 、減少する確率は  $1 - P_{dam}$  となる。

(6) 世代交代: 現在の集団を子に置き換え、(2) へ戻る。

$GA^d$  のアルゴリズムを以下に C 言語風に記述する。

```
GAd()
{
  t=0;
  初期個体集団 P(t) を生成;
  while(!収束条件を満足) {
    P'=P(t) から個体を (ランキング) 選択;
    for(each 個体対 p, q in P') {
      確率 Pc で p と q を交叉;
      for(all 遺伝子座 i in p, q) {
        確率 Prm で gi を可逆的突然変異;
        確率 Pim · Pdam(di) で di を増加させ、
        確率 Pim · (1 - Pdam(di)) で di を減少させる
      }
    }
    t = t + 1;
    P(t) = P';
  }
}
```

### 2.4 遺伝子型から表現型への変換

遺伝子型から表現型への変換関数  $h^d$  としては様々なものが考えられるが、本論文では損傷度に比例して形質の発現が減少する線形の変換関数を提案する。モデルの実数値パラメータを  $\mathbf{p} = (p_1, p_2, \dots, p_i, \dots, p_n)$  とし、 $p_i$  を表現する遺伝子の正常値を  $g_i$ 、損傷度を  $d_i$  とすると、パラメータ  $p_i$  は以下のように定義できる。

$$p_i = (1 - d_i)g_i \quad (7)$$

例えば、 $g_i$  が固定された区間  $[g^-, g^+]$  上の実数値をとると仮定した場合、この定義によれば、損傷度が 0 のときには  $p_i$  は通常の範囲  $[g^-, g^+]$ 、損傷度が 0.5 のときには  $[0.5g^-, 0.5g^+]$  の値をとり、損傷度が 1 のときには  $p_i = 0$  となる。すなわち、 $p_i$  は、損傷がなければすべての形質を発現でき、損傷が大きくなると発現する形質の範囲が制限される。なお、この変換関数では表現型の値が 0 のときにパラメータを無効とみなす、それ以外の場合には、別の変換関数を定義する必要がある (文献 [20] 3.3 節参照)。

### 2.5 モデル記述系と遺伝子表現

ニューラルネットワークやファジールールなどのモデル記述系を遺伝子で表現するとき、一つの遺伝子を以下のように様々なレベルの要素に対応させることができる。

● 細胞・素子レベル：何らかの機能を実現する細胞や素子を一つの遺伝子で表現する．例えば，ニューラルネットワークでは一つのニューロンが，ファジールールでは一つのルール全体あるいは前件部の一つのファジー集合がこのレベルに相当する．

● パラメータレベル：細胞や素子を記述するパラメータを一つの遺伝子で表現する．例えば，ニューラルネットワークでは一つの結合係数やしきい値が，ファジールールでは三角型ファジー集合の中心位置や広がりなどのパラメータがこのレベルに相当する．

● ビットレベル：パラメータを複数のビットで表現するときのビットを一つの遺伝子で表現する．パラメータを表現するバイナリストリングの1ビットがこのレベルに相当する．

MGGA ではビットレベルの遺伝子表現に対応したが， $GA^d$  はそれ以外のレベルへの対応も可能である．本論文では，一つの遺伝子を一つの（実数値）パラメータに対応させるというパラメータレベルでの対応を想定する．そして，複数の実数値パラメータで記述されるモデルについて構造最適化を試みる．回帰モデル，ニューラルネットワーク，RBF ネットワーク，ファジールールなどはこの種の記述系とみなすことができる．

### 3. $GA^d$ の性質

本章では，単純な球面関数の最適化を行い， $GA^d$  の性質を調べる．

#### 3.1 テスト問題

テスト問題として，関数値に寄与しない変数をもつ式(8)の球面関数  $F_1$  を取り上げ， $F_1$  の最小化を行う．

$$F_1(x_1, x_2, x_3) = (x_1 - 1)^2 + (x_2 - 4)^2 \quad (8)$$

関数  $F_1$  は， $(x_1, x_2) = (1, 4)$  で最小値をとる． $x_1$ ， $x_2$  は関数値に寄与するため，損傷度の低い遺伝子に対応することが望ましい．これに対して， $x_3$  は関数値に寄与しない変数のため，損傷度の高い遺伝子に対応し削除されることが望ましい．

実験の条件は以下のようにした．

● 遺伝子の表現：各変数  $x_i$  を一つの遺伝子で表現し，正常値と損傷度をそれぞれ一つの実数で表現する．

● 初期集団の構成：初期集団中の個体は，正常値を  $[-5.12, 5.11]$  の乱数，損傷度を  $[0, 1]$  の乱数として生成する．正常値の初期値は，MGGA [20] に合わせ

た．なお， $GA^d$  では  $x_2$  の最適値 4 を含まない  $[-1, 1]$  のような範囲の初期値からでも最適化が可能である．したがって  $GA^d$  は MGGA よりも初期値の設定における自由度が高いという特長もある．

● 正常値・損傷度の変化：正常値は可逆的突然変異により， $[-0.2, 0.2]$  の乱数分だけ変化させる．損傷度は不可逆的突然変異により，最大変化が 0.2 となるように， $[0, 0.2]$  の乱数分だけ増加あるいは減少させる．

● その他：個体数  $N = 50$ ，線形ランキング選択の最大期待値  $\eta^+ = 2.0$ ，1 点交叉，交叉率  $P_c = 0.8$ ，可逆的突然変異率  $P_{rm} = 0.1$ ，不可逆的突然変異率  $P_{im} = 0.1$  とする．

損傷度増加確率関数が最適化に与える影響を調べるために， $P_{dam}$  を式(3)，(4)，(5)と変化させ，関数値及び損傷度の変化の様子を観察する．実験では，300 世代までの試行を 20 回繰り返した．なお，ランク戦略を用いた通常の GA と比較するために， $P_{im} = 0.0$ ， $P_{rm} = 0.2$  に設定したときの結果を GA として示す．ただし， $GA^d$  と同等の初期集団を構成するために，正常値を  $[-5.12, 5.11]$  の乱数と  $[0, 1]$  の乱数の積，損傷度を 0.0 として初期個体を生成した．

#### 3.2 実験結果

実験結果を表 1 に示す．Best  $F_1$  は全試行における最小の関数値，Average  $F_1$  は各試行における最良個体の関数値の平均， $d$  of  $x_3$  は最良個体における  $x_3$  に対応する遺伝子の損傷度の平均，lost は最良個体における  $x_3$  の損傷度が 1 となった試行回数である．最適化能力 (Average  $F_1$ ) については，natural，通常の GA，medium，extreme の順に優れているが，ほぼ同程度の結果となっている．更に，extreme 及び medium では 20 回中 19 回と非常に高い頻度で， $x_3$  が不要な変数であることを検出できている．したがって， $GA^d$  は構造最適化のために非常に有効であると考えられる．

各世代における最良の個体の関数値の変化を図 3 に示す．関数値としては 20 回の試行の平均値を用いた．初期段階における最適化の速度は，medium，GA，natural，extreme の順となっているが，最終的には同

表 1  $F_1$  の実験結果  
Table 1 Results of experiment for  $F_1$ .

$P_{dam}$	Best $F_1$	Average $F_1$	$d$ of $x_3$	lost
extreme	1.067e-08	1.769e-07	0.999	19
natural	8.060e-10	9.396e-08	0.584	11
medium	4.755e-10	1.063e-07	0.954	19
GA	6.499e-09	9.487e-08	0.000	0

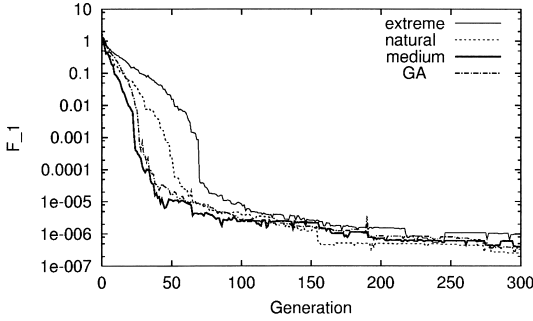


図 3  $P_{dam}$  に対する関数値の変化  
Fig. 3 Graph of function value for several  $P_{dam}$ .

程度の最適化に成功している。

各世代における最良の個体の遺伝子  $x_1, x_2, x_3$  の損傷度の変化を図 4 に示す。損傷度としては 20 回の試行の平均値を用いた。 $x_1$  の損傷度については、natural がやや小さいがいずれも 0.5 程度に収束している。 $x_2$  の損傷度については、natural, medium は 0.1 程度に収束し、extreme は 0.2 程度に収束しており、有効な遺伝子であることが検出できている。 $x_1$  の損傷度と比較して  $x_2$  の損傷度の方がより低い傾向がある。 $x_1$  の最適値 1 は初期集団の正常値の最大値 5.11 と比較して小さいため、損傷度が高い遺伝子でも表現可能であったのに対して、 $x_2$  の最適値 4 は最大値に近いため、損傷度の低い個体でなければ表現が困難であったことが原因であると考えられる。 $x_3$  の損傷度については、extreme はほぼ 1 に、medium も 0.95 以上に収束し、無効な遺伝子であることが検出できた。

この結果から、無効な遺伝子を検出しやすいのは、extreme, medium であり、最適化に優れているのは natural, medium であると推定できる。したがって、 $F_1$  では、最適化を行うと同時に、無効な遺伝子を検出し、退化現象を起こしやすい関数は medium であると考えられる。本研究では、無効な遺伝子を検出し、退化現象を起こすことを目的としているため、この後の実験では、medium を使用することにする。

#### 4. 多項式モデルの構造最適化

目的変数を説明変数の多項式で表現する多項式モデルについて構造最適化を行う。

##### 4.1 学習課題

ここでは坂元ら [3] の例を取り上げる。訓練データは、式 (9) の 2 次の多項式モデルを真の構造とし、 $\varepsilon_i$

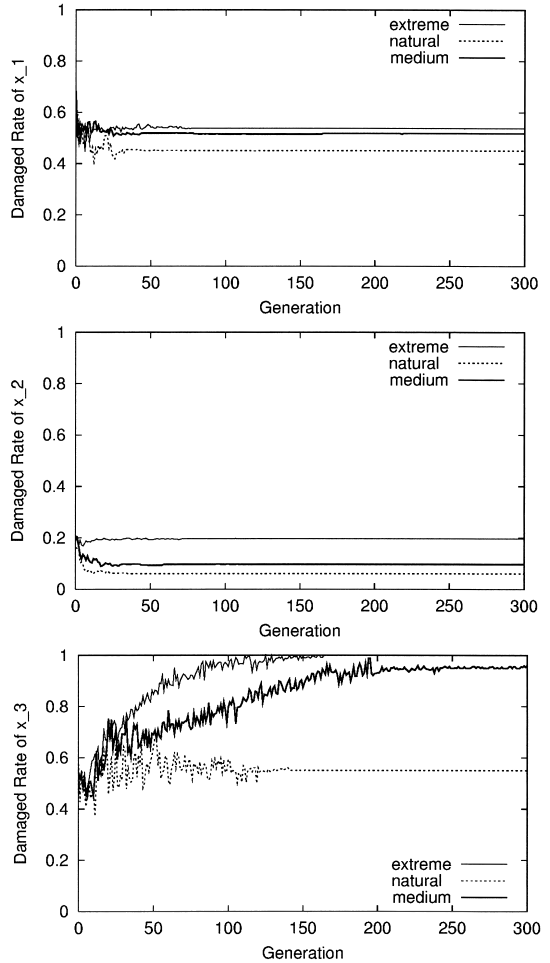


図 4  $P_{dam}$  に対する損傷度の変化  
Fig. 4 Graph of damaged rate of  $x_i$  for several  $P_{dam}$ .

に平均 0、分散 0.01 の正規乱数を与えて得られたものである。

$$y_i = 0.05 - 0.4x_i + 0.8x_i^2 + \varepsilon_i \quad (9)$$

ここでは式 (10) の多項式モデルを導入し、 $a_i$  をパラメータとして遺伝子にコーディングする。

$$y_i = a_0 + a_1x_i + a_2x_i^2 + \dots + a_mx_i^m \quad (10)$$

通常多項式モデルでは、 $m$  を変化させ、 $a_0 \sim a_m$  までの係数をすべて用いる。 $GA^d$  では損傷遺伝子が出現することにより、自然に任意の  $a_i$  を削除することができる。

坂元らによれば、AIC の観点から最良のモデルは、

表 2 GA<sup>d</sup> による多項式モデルの構造最適化の結果  
Table 2 Results of structural optimization of polynomial model by GA<sup>d</sup>.

Trial	Model	$\sigma^2$	AIC
1	$+0.5065x^3$	0.00719	-19.0682
2	$+0.5064x^3$	0.00719	-19.0682
3	$+0.5066x^3$	0.00719	-19.0682
4	$-0.2224x + 0.7008x^2$	0.00652	-18.1389
5	$+0.7317x^4 - 0.1995x^5$	0.00771	-16.2990
6	$+1.0746x^4 - 0.5695x^5$	0.00695	-17.4432
7	$-0.0612x + 0.5870x^3$	0.00696	-17.4248
8	$+1.2905x^4 - 0.8066x^5$	0.00665	-17.9198
9	$-0.6248x^2 + 2.3696x^3$ $-1.2793x^4$	0.00602	-17.0201
10	$-0.3487x + 0.8545x^2$	0.00613	-18.8275
Average	1.8 parameters	0.00685	-18.0278

式 (11) であり,  $\sigma^2 = 0.00592$ ,  $AIC = -17.21$  である.

$$y = 0.03582 - 0.49218x + 0.97237x^2 \quad (11)$$

#### 4.2 実験

GA<sup>d</sup> を用いて多項式モデルのパラメータ構造を学習する.  $g_i$  の初期値,  $d_i$  の初期値をそれぞれを区間  $[-5.12, 5.11]$ , 区間  $[0, 1]$  の乱数とし,  $m = 5$  とした. すなわち, 6 個の実数型遺伝子で 6 個のパラメータを表現することになる. 適合度は教師データとの平均 2 乗誤差  $\sigma^2$  とし, 適合度を最小化した. GA<sup>d</sup> の設定は, 個体数  $N = 50$ ,  $\eta^+ = 2.0$ , 1 点交叉,  $P_c = 0.8$ ,  $P_{rm} = 0.2$ ,  $P_{im} = 0.2$ , medium を用いた.  $g_i$  は可逆的突然変異により区間  $[-0.15, 0.15]$  の乱数分変化する.  $d_i$  は不可逆的突然変異により区間  $[0, 0.15]$  の乱数分増加あるいは減少する. 実験では, 世代数 200 までの試行を 10 回行い, 結果を表 2 に示した.

10 回の試行のうち, 5 パラメータが削除されたのが 3 回, 4 パラメータが削除されたのが 6 回, 3 パラメータが削除されたのが 1 回であり, 平均 4.2 個のパラメータを削除した. GA<sup>d</sup> が十分なパラメータ削減力を有することが示された. GA<sup>d</sup> により求められたモデルは, AIC の最良値が  $-19.0682$ , 平均が  $-18.0278$  という良好な結果が得られており, 8 回の試行において式 (11) よりも AIC が優れたモデルを発見した. 試行 No.10 において,  $-0.3487x + 0.8545x^2$  という真のモデルに近いモデルも得られている. なお, すべての試行において定数項  $a_0$  が削除されているが, これは真のモデルにおいて定数項は小さく, 定数項を削除しても他のパラメータにより補正することができたためと考えられる.

## 5. ニューラルネットワークの構造最適化

階層型ニューラルネットワークを GA<sup>d</sup> で最適化することにより構造を学習する. 結合荷重に着目し, 結合荷重を表現する遺伝子の損傷度が増えると結合荷重が減少し, 損傷度が 1 となった場合にはその結合が削除されたと考える. しきい値についても同様である.

### 5.1 学習課題

式 (12) の関数を学習するために, 2 入力 1 出力 3 層の階層型ニューラルネットワークを用いた.

$$\text{neuro}(x_1, x_2) = x_1 x_2^2 \quad (12)$$

訓練データは,  $\{(x_1, x_2) | x_1, x_2 = 0.0, 0.2, \dots, 1.0\}$  の 36 点において, 関数値に平均 0, 分散 0.01 の正規乱数を加えた値を用いた. 中間層を 6 ニューロン, すなわち, 最大パラメータ数を  $25 (= 3 \times 6 + 7)$  とした.  $g_i, d_i$  の初期値をそれぞれを区間  $[-10, 10]$ , 区間  $[0, 1]$  の乱数とした. 適合度は教師データとの平均 2 乗誤差  $\sigma^2$  とし, 適合度を最小化した. GA<sup>d</sup> の設定は, 個体数  $N = 50$ ,  $\eta^+ = 2$ , 1 点交叉,  $P_c = 0.8$ ,  $P_{rm} = 0.1$ ,  $P_{im} = 0.1$ , medium を用いた.  $g_i$  は可逆的突然変異により区間  $[-0.2, 0.2]$  の乱数分変化する.  $d_i$  は不可逆的突然変異により区間  $[0, 0.1]$  の乱数分増加あるいは減少する. 実験では, 世代数 300 までの試行を 10 回行った.

### 5.2 実験結果

各試行について, 削除されたパラメータ, 有効パラメータ数  $P_{eff}$ , 平均 2 乗誤差  $\sigma^2$  及び AIC を表 3 に示した.  $nw$  は結合荷重,  $n\theta$  はしきい値がそれぞれ  $n$  個削除されたことを示している. 参考のために, 他のニューロンと全く結合していないニューロン数を  $nN$  で示した. なお, 出力が他のニューロンに接続されていない場合には, 実質的にはそのニューロンを削除できる. 括弧内は, この知識を用いた場合の実質的な結果である.

実質的な有効パラメータ数は平均 9.7 であり, 平均 15.3 パラメータが削除された. 実質的な AIC の良好な試行から, 中間層ニューロンは 2 あるいは 3 個が適当であり, 更に結合荷重及びしきい値を数個削減可能であることがわかる. 実質的ニューロン数の最も少ない試行 No.4 の構造を図 5 に示す. ただし, 結合強度の絶対値に比例して結合を太く表している.

表 3  $GA^d$  によるニューラルネットワークの構造最適化の結果 (括弧内は実質的な結果)

Table 3 Results of structural optimization of the neural networks by  $GA^d$ .

Trial	Reduced param.	$P_{eff}$	$\sigma^2$	AIC
1	9w, 1 $\theta$ (12w, 3 $\theta$ / 3N)	15 (10)	2.878e-03	-76.46 (-86.46)
2	9w, 6 $\theta$ / 2N (10w, 6 $\theta$ / 3N)	10 (9)	2.086e-03	-98.05 (-100.05)
3	10w, 1 $\theta$ (10w, 4 $\theta$ / 3N)	14 (11)	2.714e-03	-80.57 (-86.57)
4	10w, 4 $\theta$ / 2N (13w, 5 $\theta$ / 4N)	11 (7)	2.859e-03	-84.69 (-90.69)
5	12w, 4 $\theta$ / 3N	9	2.024e-03	-101.12
6	9w, 4 $\theta$ / 1N (10w, 4 $\theta$ / 2N)	12 (11)	1.809e-03	-99.18 (-101.18)
7	11w, 5 $\theta$ / 3N	9	1.947e-03	-102.54
8	9w, 4 $\theta$ / 1N	12	3.993e-03	-70.67
9	9w, 6 $\theta$ / 2N	10	1.762e-03	-104.11
10	10w, 3 $\theta$ (11w, 5 $\theta$ / 3N)	12 (9)	1.962e-03	-96.25 (-102.25)
Avg.	13.6 parameters (15.3 parameters)	11.4 (9.7)	0.002403	-91.3651 (-94.7651)

表 4  $GA^d$  と MGGA の比較  
Table 4 Comparison between  $GA^d$  and MGGA.

Problem	Algorithm	$P_{eff}$	$\sigma^2$	AIC	CPU(s)
Polynomial	$GA^d$	1.8	0.00685	-18.0278	1.3
	MGGA	1.8	0.00700	-17.9041	2.8
Neural Net	$GA^d$	9.7	0.00240	-94.7651	7.1
	MGGA	15.6	0.00257	-80.4040	20.2

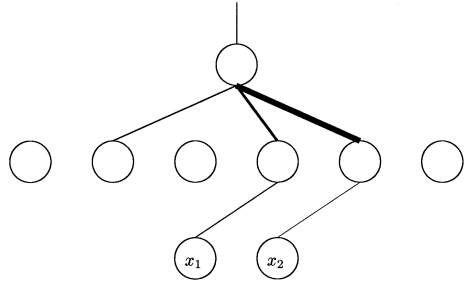


図 6 AIC 最小のネットワーク (10,000 世代)  
Fig. 6 Neural network of minimum AIC.

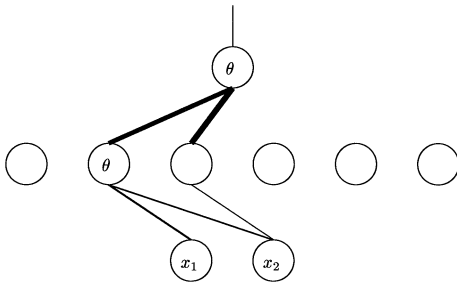


図 5  $GA^d$  によって獲得されたネットワーク (試行 4)  
Fig. 5 Neural network obtained by  $GA^d$  (Trial 4).

## 6. 検 討

### 6.1 MGGA との比較

$GA^d$  と MGGA を安定性, 解の精度, 実行効率で比較する.

多項式モデルとニューラルネットワークの構造最適化において  $GA^d$  と MGGA を解の精度と実行効率に関して比較した結果を表 4 に示す. ただし, CPU(s) は, AlphaServer ES40(Alpha 21264A 667 MHz) を使用したときの秒単位の実行時間である. いずれの場合も  $GA^d$  は MGGA より精度の高い解を得ている. 特にニューラルネットワークでは高い精度とより簡潔な構造を両立する解を得ている. また, 多項式モデルでは約 2.1 倍, ニューラルネットワークでは約 2.8 倍  $GA^d$  の方が実行速度が速い. この傾向は, パラメー

タ数が増加するほど, あるいは遺伝子数が増加するほど顕著になると考えられる. したがって, 解の精度及び実行効率は  $GA^d$  の方が優れている.

次に, 安定性を比較するために, ニューラルネットワークの構造最適化において各試行を 2,000 世代まで行った. MGGA では 678 世代ですべての遺伝子が変異遺伝子となってしまった. しかし,  $GA^d$  では損傷率は増加してはいるが非常にゆっくりとした増加であり, 10,000 世代まで実験を行っても平均損傷率は 0.8615 (有効パラメータ数 3.5 程度) までの増加に止まった. MGGA も  $GA^d$  も 300 世代程度で十分に構造最適化が達成できるように実験パラメータを設定したが, より長い世代にわたって最適化を行うと  $GA^d$  の方が明らかに安定した最適化に成功している. したがって,  $GA^d$  は MGGA よりも安定した構造最適化法である.

参考のために 10,000 世代実行したときに得られた AIC 最小のニューラルネットワークを図 6 に示す. これは, 有効パラメータ数 5, 平均 2 乗誤差 0.001567, AIC -118.35 のネットワークであり, 非常に単純な構造が得られている. このように,  $GA^d$  は訓練データの構造を明確化させるという働きをもつことを強調しておきたい.

6.2 ニューラルネットワークにおける構造最適化  
はじめに述べたように,  $GA^d$  と同様の縮退に基づく方法として, 石川はニューラルネットワークのための構造最適化法を提案している [18], [19]. ここでは,



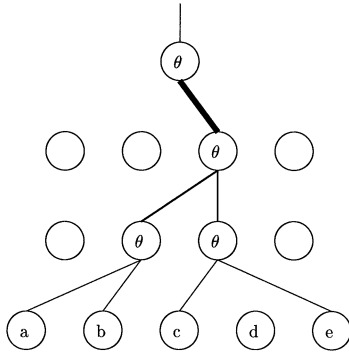


図 7 発見された論理構造  
Fig. 7 Discovered logical structure.

石川の学習課題である式 (13) の論理関数  $L$  を学習することにより,  $GA^d$  が同程度の能力を有することを示すとともに,  $GA^d$  の特長について説明する.

$$L(a, b, c, d, e) = (a \cup b) \cap (c \cup e) \quad (13)$$

なお,  $L$  は関数値と独立な変数  $d$  を含んでいる.

ニューラルネットワークは, 5 入力 1 出力 4 層の階層型ニューラルネットワークとし, 第 2 層, 第 3 層のニューロン数をいずれも 4 とする. したがって, 最大パラメータ数は  $49 (= 6 \times 4 + 5 \times 4 + 5)$  である. 訓練データは, 各変数に 0, 1 を代入した  $2^5 = 32$  個のうちランダムに選択した 26 個のデータを用いる.

$GA^d$  では, 各遺伝子を実数値で表現し,  $g_i, d_i$  の初期値をそれぞれ区間  $[-5, 5]$ , 区間  $[0, 1]$  の乱数とした. 適合度は教師データとの平均 2 乗誤差  $\sigma^2$  とし, 適合度を最小化した.  $GA^d$  の設定は, 個体数  $N = 50$ ,  $\eta^+ = 2.0$ , 1 点交叉,  $P_c = 0.8$ ,  $P_{rm} = 0.06$ ,  $P_{im} = 0.06$ , medium を用いた.  $g_i$  は区間  $[-0.2, 0.2]$  の乱数分変化し,  $d_i$  は区間  $[0, 0.2]$  の乱数分増加あるいは減少する. 実験では, 世代数 2,500 までの試行を 10 回行った.

得られた AIC 最良の個体を図 7 に示す. このニューラルネットワークは明らかに論理関数  $L$  の構造を直接的に表現した構造を有しており, 論理構造の発見に成功している. 6 個の未学習データに対する平均 2 乗誤差は  $8.44e-27$  であり, 未学習データに対しても誤差の少ない精度の高いモデルが得られた.

このように,  $GA^d$  によっても石川の研究と同様の結果を得ることができる. 石川の方法と比較した  $GA^d$  の特長としては, ニューラルネットワークだけでなく様々な記述系に適用できる汎用的な方法であること,

評価関数値のみを利用して最適化を行うため評価関数に微分可能性を仮定しなくてもよいこと, したがって例えば AIC などの情報量基準を直接最適化したり強化学習などにおける構造最適化に適用できる可能性があることが挙げられる.

## 7. む す び

本研究では, 損傷遺伝子を導入した  $GA^d$  を提案した. 損傷遺伝子における確率的变化を提案し, 損傷度に対して線形の表現型への変換関数を仮定することで,  $GA^d$  がパラメータ値を最適化しながらパラメータ構造も最適化できる構造最適化アルゴリズムであることを示した. また, 多項式モデル及びニューラルネットワークの構造最適化を行い, MGGA と比較することにより,  $GA^d$  が汎用的でより効率的かつ精度の高い安定した手法であることも示した.

本論文では, 損傷度増加確率関数について 3 種類の関数を検討した. しかし, ほかにも様々な可能性が考えられるため, 退化現象について考察しながら, どのような関数が適切かを検討していく必要がある. 特に,  $P_{dam}(d) = \text{定数}$  という関数は簡潔でかつ損傷度の増加割合を比較的容易に調整可能であるため, 有望であり, 調査していきたい. また, 本論文では,  $P_{rm}, P_{im}$  は  $1/\ell$  を中心に  $[1/(3\ell), 3/\ell]$  程度の範囲 ( $\ell$  は遺伝子数),  $g_i, d_i$  の変化量は  $[0.1, 0.2]$  程度の範囲に設定して実験を行ったが, 問題によって適切な値は異なるため, これらの設定についてより詳しく調べる必要がある.

今後は,  $GA^d$  をニューラルネットワークやファジー推論ルールの構造最適化に適用し, より実用的な問題への応用を試みる予定である.

謝辞 本研究の一部は, 日本学術振興会科学研究費補助金基盤研究 (C) 2002 課題番号 14580498) による補助のもとで行われた.

## 文 献

- [1] 栗田多喜夫, “情報量基準による 3 層ニューラルネットの隠れ層のユニット数の決定法,” 信学論 (D-II), vol. J73-D-II, no.11, pp.1872–1878, Nov. 1990.
- [2] H. Akaike, “A new look at the statistical model identification,” IEEE Trans. Autom. Control, vol. AC-19, no.6, pp.716–723, Dec. 1974.
- [3] 坂本慶行, 石黒真木夫, 北川源四郎, 情報量統計学, 共立出版, 1983.
- [4] J. Rissanen, “A universal prior for integers and estimation by minimum description length,” The Annals of Statistics, vol.11, no.2, pp.416–431, 1983.

- [5] J. Rissanen, "Stochastic complexity and modeling," The Annals of Statistics, vol.14, no.3, pp.1080-1100, Sept. 1986.
- [6] J.E. Moody, "The effective number of parameters: An analysis of generalization and regularization in nonlinear learning systems," in Advances in Neural Information Processing Systems 4, ed. J.E. Moody, S.J. Hanson and R.P. Lippmann, pp.847-854, Morgan Kaufmann, San Mateo, 1992.
- [7] H. Kitano, "Designing neural networks using genetic algorithms with graph generation system," Complex System, vol.4, no.4, pp.461-476, Aug. 1990.
- [8] 長尾智晴, 安居院猛, 長橋 宏, "遺伝的手法による神経回路網の構造進化," 信学論 (D-II), vol.J75-D-II, no.9, pp.1634-1637, Sept. 1992.
- [9] 長尾智晴, 安居院猛, 長橋 宏, "GA による仮想生物の運動制御用神経回路網の自動生成," 信学論 (D-II), vol.J78-D-II, no.7, pp.1150-1152, July 1995.
- [10] 高橋裕樹, 中嶋正之, "遺伝的アルゴリズムを用いた最適な階層型神経回路網の構造決定に関する一考察," 信学論 (D-II), vol.J79-D-II, no.11, pp.1920-1928, Nov. 1996.
- [11] H. Takahashi and M. Nakajima, "Evolutional design and training algorithm for feedforward neural networks," IEICE Trans. Inf. & Syst., vol.E82-D, no.10, pp.1384-1392, Oct. 1999.
- [12] 押野隆弘, 尾島 潤, 山本真司, "誤差逆伝搬学習における中間層ユニットの逐次削減法," 信学論 (D-II), vol.J76-D-II, no.7, pp.1414-1424, July 1993.
- [13] 松永 豊, 村瀬一之, 山川 修, 谷藤 学, "競合作用により冗長中間層素子を自律淘汰する誤差逆伝搬学習アルゴリズム," 信学論 (D-II), vol.J79-D-II, no.3, pp.403-412, March 1996.
- [14] A. Leonardis and H. Bischof, "An efficient MDL-based construction of RBF networks," Neural Networks, vol.11, no.5, pp.963-973, July 1998.
- [15] 荒木昭一, 野村博義, 林 勲, 若見 昇, "ルールを逐次的に生成するファジィモデリングの一提案," 日本ファジィ学会誌, vol.4, no.4, pp.722-732, Aug. 1992.
- [16] 岸田一也, 宮島廣美, 福元伸也, 村島定行, "学習アルゴリズムによるファジィモデルの構成," 日本ファジィ学会誌, vol.7, no.3, pp.585-593, June 1995.
- [17] J. Yen and L. Wang, "Application of statistical information criteria for optimal fuzzy model construction," IEEE Trans. Fuzzy Syst., vol.6, no.3, pp.362-372, Aug. 1998.
- [18] 石川真澄, "忘却を用いたコネクショニストモデルの構造学習アルゴリズム," 人工知能誌, vol.5, no.5, pp.595-603, Sept. 1990.
- [19] M. Ishikawa, "Structural learning with forgetting," Neural Networks, vol.9, no.3, pp.509-521, April 1996.
- [20] 高濱徹行, 阪井節子, 磯道義典, "変異遺伝子を導入した遺伝的アルゴリズム (MGGA) の提案," 信学論 (D-I), vol.J84-D-I, no.9, pp.1297-1306, Sept. 2001.
- [21] 高濱徹行, 阪井節子, "退化の発現による最適構造の発見,"

人工知能学会研究会資料, SIG-FAI-A002-2, pp.7-12, Sept. 2000.

- [22] J.E. Baker, "Adaptive selection methods for genetic algorithms," Proc. First International Conference on Genetic Algorithms and Their Applications, pp.101-111, Pittsburgh, USA, July 1985.

(平成 14 年 5 月 30 日受付, 10 月 5 日再受付)



高濱 徹行 (正員)

1982 京大・工・電気第二卒。1987 同大学院博士課程研究認定退学。同年福井大学工学部助手。1994 同大講師。1998 より広島市立大学情報科学部知能情報システム工学科助教授。ファジー知識表現と学習アルゴリズム, 機械学習, 推論, 非線形最適化, CAI, 自然言語処理などに関する研究に従事。情報処理学会, 人工知能学会, 教育システム情報学会, 言語処理学会, ISGEC, IEEE 各会員。工博。



阪井 節子

1979 福井大・教育卒。1984 阪大大学院基礎工学研究科数理系後期課程単位取得退学。1986 甲子園大学経営情報学部講師。1990 福井大学教育学部助教授。1998 より広島修道大学商学部経営学科教授。ゲーム理論, 意思決定, ファジー数値計画, GA によるファジー制御, CAI などに関する研究に従事。日本 OR 学会, 日本ファジィ学会, 日本生産管理学会各会員。工博。



市村 匠 (正員)

1992 桐蔭横浜大・工・制御システム卒。1997 同大学院博士課程了。同年広島市立大学情報科学部知能情報システム工学科助手。ニューラルネットワーク, 遺伝的アルゴリズムによる学習システム, 意思決定支援システムの研究に従事。人工知能学会, 日本ファジィ学会, 日本医療情報学会, 人間工学会, INNS, IEEE 各会員。工博。



磯道 義典 (正員)

1962 東大・工・電気卒。電子技術総合研究所パターン情報部, 広島大学総合科学部を経て現在広島市立大学情報科学部知能情報システム工学科教授。信号理論, パターン認識, 学習システムに関する研究を経て現在人工知能, ゲーム論の研究に従事。著書「情報理論」, 「情報学概論」, 「情報化革命」など。人工知能学会, 情報処理学会, 日本神経回路学会, 社会・経済システム学会各会員。工博。