

## 遺伝的アルゴリズムにおける交叉の役割の連鎖解析

古谷 博史, 藤林 由紀, 村田 真知子 京都教育大学

遺伝的アルゴリズムにおける交叉の役割を連鎖解析を用いて研究した。適応度関数が線形で解析が容易な OneMax 問題を研究対象としてとりあげた。連鎖不均衡は遺伝学において重要な概念であり、遺伝子座間に相関があることを意味する。交叉の効果を連鎖不均衡を計算することによって予測できることを示した。

### Linkage Analysis of Crossover in Genetic Algorithms

HIROSHI FURUTANI, YUKI FUJIBAYASHI, MACHIKO MURATA  
KYOTO UNIVERSITY OF EDUCATION

The paper describes an investigation of crossover in genetic algorithms (GAs) by means of linkage analysis. We consider the GA applied to One Max problem because its fitness function is linear and easy to analyze. Linkage disequilibrium is an important concept in genetics, which represents the genetic correlation between loci. It is found that the effect of crossover can be estimated by calculating the degree of linkage disequilibrium.

#### 1. はじめに

遺伝的アルゴリズム (以下 GA と略す) はこれまで多くの分野に適用され、その有効性について疑問を差し挟む余地はあまり無くなったように思われる。しかし、どのようなメカニズムで GA が解を探索しているのか、といった基本的な疑問にはまだ答えきれていない。Holland や Goldberg による GA のテキストでも、かなりの部分がそうした疑問に対する解答 (の試み) にさかれている<sup>1),2)</sup>。

GA で用いられている主要な遺伝的操作としては、選択、突然変異、交叉の三つが挙げられる。これらの操作のうち、最もその役割の解明が遅れているのが交叉である。交叉には二つの個体が関与するため、その解析は複雑になり本質を抽出することが困難になる。生物でも交叉は性的問題と密接に関連しているが、その存在意義について十分な説得力をもつ理論はまだないと言ってよい。

本論文では、交叉の問題を連鎖不均衡という概念を手がかりに考えてみた。そのため、GA においてこれまでたびたび研究されてきた OneMax 問題とよばれる問題を例にとり、GA における交叉と連鎖不均衡の関係を検討した。ここではさらに、遺伝学者 Fisher が導いた「自然選択の基本定理」をもとに進化速度と集団の分散の関係に注目する<sup>3)</sup>。進化速度や集団の分

散については後で定義するが、Fisher の定理を大まかに述べると、進化速度と集団の分散は比例する、ということである。したがって、進化を速めるためには分散を大きくする必要があり、そのとき重要な働きをするのが交叉である。

連鎖不均衡は、生物学において性的問題を研究する上で必要となる遺伝学概念である<sup>4),5)</sup>。連鎖不均衡とは 2 個所の遺伝子座の遺伝子が互いに相関をもつことを意味し、遺伝子座間に相関が無い場合を連鎖均衡という。一般に交叉は、遺伝子座間の相互作用を断ち切る働きをし、交叉を繰り返し適用すると最終的に連鎖均衡の状態に収束する。最後に数値実験により、OneMax では、交叉により連鎖均衡状態になり、その結果集団の分散が増加し、進化速度が増大することを明らかにする。

#### 2. 方法

この論文で用いた GA は、Goldberg のテキストに記載されている Simple Genetic Algorithm にほぼ準じたものである<sup>2)</sup>。集団中の個体数  $N$  は非常に大きいものとし、確率的な揺らぎは無視する。数学的な取り扱いの詳細については、論文<sup>6),7)</sup>を参照されたい。

長さ  $\ell$  のビット列  $B_i$  により  $i$  番目の遺伝子型を表現する

$$B_i = \langle i_\ell, i_{\ell-1}, \dots, i_2, i_1 \rangle.$$

ここで  $i_k$  ( $k = 1, \dots, \ell$ ) は第  $k$  番目のビットを表わす。このビット列を整数  $i$  の 2 進数表現とみなして  $B_i$  と  $i$  を同一視する。遺伝子型は全部で  $n = 2^\ell$  種類ある。

$$|i| = \sum_{k=1}^{\ell} i_k$$

により  $B_i$  中のビット 1 の総数を表わすことにする。集団の分布は相対頻度

$$x_i(t) = \frac{N_i}{N} \quad (0 \leq i \leq n-1), \quad (1)$$

により表わす。ここで  $N_i$  は第  $i$  遺伝子型をもつ個体数を表わし、 $N$  は集団の個体数を表わす。頻度  $x_i(t)$  は規格化の条件を満たす

$$\sum_{i=0}^{n-1} x_i(t) = 1. \quad (2)$$

## 2.1 Walsh 変換

ここで Walsh 関数と Walsh 変換について簡単にふれておく<sup>8)</sup>。長さ  $\ell$  のビット列  $i$  に対する第  $j$  番目の Walsh 関数を次式で定義する

$$W_{ij} \equiv \prod_{k=1}^{\ell} (-1)^{i_k \cdot j_k}. \quad (3)$$

頻度  $x_i(t)$  の Walsh 変換  $\tilde{x}_i(t)$  を

$$\tilde{x}_i(t) \equiv \sum_{j=0}^{n-1} W_{ij} x_j(t) \quad (4)$$

とし、その逆変換を

$$x_i(t) = \frac{1}{n} \sum_{j=0}^{n-1} W_{ij} \tilde{x}_j(t), \quad (5)$$

と定義する。以下では Walsh 変換  $\tilde{x}_i$  を Walsh 係数とよぶことにする。

すべての  $j$  について  $W_{0j} = 1$  が成り立つことと、規格化条件 (2) から次式を得る。

$$\tilde{x}_0(t) = \sum_{i=0}^{n-1} x_i(t) = 1. \quad (6)$$

$k = |i|$  となる  $\tilde{x}_i$  を  $k$  次の Walsh 係数とよぶ。  $\tilde{x}_i$  の次数を表現したいときは次の表式を用いることにする

$$\tilde{x}_i \equiv \tilde{x}^{(k)}[b_1, b_2, \dots, b_k].$$

ここで  $k = |i|$  であり、  $b_1 < b_2 < \dots < b_k$  は  $i$  の中のビット 1 の位置を表わす。

集団において位置  $m$  のビットの値が 1 である確率  $p_m^{(1)}$  を Walsh 係数を用いて表わす。そのためまず次

の関係に注意する

$$\frac{1}{2}(1 - \tilde{x}^{(1)}[m]) = \frac{1}{2} \sum_j (1 - (-1)^{jm}) x_j.$$

このことを用いて、位置  $m$  における値が 1 であるようなビット列  $j$  を抜き出す演算子が定義できる

$$\delta(1 - j_m) = \frac{1 - (-1)^{jm}}{2}.$$

同様にして、ビット 0 を抜き出す演算子は

$$\delta(j_m) = \frac{1 + (-1)^{jm}}{2},$$

で与えられる。したがって

$$\begin{aligned} p_m^{(0)} &= (1 + \tilde{x}^{(1)}[m])/2 \\ p_m^{(1)} &= (1 - \tilde{x}^{(1)}[m])/2, \end{aligned} \quad (7)$$

となる。

同様にして、位置  $m$  と  $m'$  のビット値がともに 1 である確率  $p_{m,m'}^{(11)}$  も Walsh 係数を用いて表わすことができる

$$p_{m,m'}^{(11)} = \frac{1}{4} \{1 - \tilde{x}^{(1)}[m] - \tilde{x}^{(1)}[m'] + \tilde{x}^{(2)}[m, m']\}. \quad (8)$$

## 2.2 平均適応度の時間変化

遺伝的操作が選択のみの場合の平均適応度  $\bar{f}(t)$  の時間変化を求めてみる。世代  $t+1$  における頻度は  $x_i(t)$  から求まる

$$x_i(t+1) = \frac{f_i}{\bar{f}(t)} x_i(t) \quad (i = 0, \dots, n-1). \quad (9)$$

ここで  $f_i$  は遺伝子型  $B_i$  に対する適応度であり、  $\bar{f}(t)$  は世代  $t$  における集団の平均適応度である

$$\bar{f}(t) = \sum_{i=0}^{n-1} f_i x_i(t). \quad (10)$$

次に、この進化方程式から Fisher が「自然選択の基本定理」と名づけた重要な定理を導く<sup>3)</sup>。進化方程式の両辺に  $f_i$  をかけ、すべての  $i$  についての和をとる

$$\bar{f}(t+1) = \frac{1}{\bar{f}(t)} \sum_i f_i^2 x_i(t).$$

平均適応度の世代あたりの変化率

$$\Delta \bar{f}(t) = \bar{f}(t+1) - \bar{f}(t),$$

は次式で与えられる。

$$\Delta \bar{f}(t) = \frac{1}{\bar{f}(t)} \text{VAR}(f), \quad (11)$$

$$\text{VAR}(f) = \sum_i f_i^2 x_i(t) - \bar{f}(t)^2.$$

平均適応度  $\bar{f}(t)$  の変化率を進化速度の一つの指標とみなし

$$v(t) \equiv \Delta \bar{f}(t) = \bar{f}(t+1) - \bar{f}(t) \quad (12)$$

とする。進化速度  $v(t)$  は適応度の分散  $\text{VAR}(f)$  に比例する。したがって、進化を促進するためには適応度の分散を大きくすればよいことが分かる。

### 3. 分散と連鎖不均衡

#### 3.1 連鎖不均衡

連鎖不均衡係数を次式で定義する<sup>4)</sup>

$$D[m, m'] = p_{m, m'}^{(11)} - p_m^{(1)} p_{m'}^{(1)}. \quad (13)$$

式 (7), (8) をこの定義に代入し

$$D[m, m'] = \frac{1}{4} \{ \hat{x}^{(2)}[m, m'] - \hat{x}^{(1)}[m] \hat{x}^{(1)}[m'] \} \quad (14)$$

をえる。Walsh 表現を用いた交叉の研究は Vose のテキストなどに記述されている<sup>9)</sup>。それらの結果から連鎖不均衡係数に対する交叉の効果を求めることができる。

1 次の Walsh 係数は交叉に対して不変である<sup>9)</sup>

$$\hat{C}\hat{x}^{(1)}[m] = \hat{x}^{(1)}[m],$$

ここで  $\hat{C}$  は交叉率  $\chi = 1$  における交叉の効果を表わすものとする。

2 次の Walsh 係数に対する交叉の効果も同様に計算することができ、一様交叉の場合<sup>10)</sup>

$$\hat{C}\hat{x}^{(2)}[m, m'] = \frac{1}{2} \{ \hat{x}^{(2)}[m, m'] + \hat{x}^{(1)}[m] \hat{x}^{(1)}[m'] \}.$$

となる。したがって連鎖不均衡係数に対する効果は次式で与えられる

$$\hat{C}D[m, m'] = \frac{1}{2} D[m, m'].$$

#### 3.2 適応度の平均と分散

OneMax 問題の適応度関数を次式で定義する

$$f_i = \sum_{k=1}^l i_k = |i|.$$

この関数から二つの統計量を計算する。一つは平均適応度  $\bar{f}$  であり、他の一つは適応度の分散  $\text{VAR}(f)$  である。世代  $t$  における平均適応度は

$$\bar{f}(t) = E\{|i|\} = \sum_{i=0}^{n-1} |i| x_i(t)$$

で与えられ、分散は

$$\text{VAR}(f) = V\{|i|\} = \sum_{i=0}^{n-1} |i|^2 x_i(t) - E\{|i|\}^2,$$

となる。これらの量は次の母関数  $G(s, t)$  を用いて用意に計算できる<sup>10)</sup>

$$G(s, t) = \sum_{i=0}^{n-1} \exp\{|i|s\} x_i(t).$$

母関数  $G(s, t)$  中の  $x_i(t)$  を Walsh 係数  $\bar{x}_j(t)$  で展開すると以下の計算が容易になる

$$G(s, t) = \frac{1}{n} \sum_{i,j} \exp\{|i|s\} W_{ij} \bar{x}_j(t).$$

母関数  $G$  の具体的な形は次式で与えられる。

$$G = \sum_{j=0}^{n-1} \left( \frac{1+e^s}{2} \right)^{\ell-|j|} \left( \frac{1-e^s}{2} \right)^{|j|} \bar{x}_j(t). \quad (15)$$

平均適応度  $E\{|i|\}$  は母関数から

$$E\{|i|\} = \left. \frac{\partial}{\partial s} G(s, t) \right|_{s=0},$$

として求めることができ、次式をえる

$$E\{|i|\} = \sum_{m=0}^{\ell} p_m^{(1)} = \sum_{m=0}^{\ell} \frac{1 - \hat{x}^{(1)}[m](t)}{2}. \quad (16)$$

同様にして分散は、

$$V\{|i|\} = V_A + V_I, \quad (17)$$

で与えられ、ここで

$$V_A = \sum_{m=0}^{\ell} p_m^{(1)} (1 - p_m^{(1)}), \quad (18)$$

$$V_I = 2 \sum_{m < m'} D[m, m']. \quad (19)$$

となる。 $V_A$  は  $p_m^{(1)}$  にも依存し、交叉の影響は受けない。一方、 $V_I$  は二つの遺伝子座間の相関に依存し、 $D[m, m']$  の関数である。したがって、交叉によって値は変化する。

### 4. 数値実験と考察

OneMax 問題を例にとり上げ、数値的に交叉の効果を検討する。ビット長  $\ell = 8$ 、集団の個体数  $N = 200$ 、突然変異率  $p = 0.01$  として GA を実行した。乱数を変えながら同じ計算を 100 回繰り返し、得られた結果の平均を求めた。初期状態は、 $p_m^{(1)} = 1/\ell$  として集団の個体をランダムに生成した。二通りの交叉係数 ( $\chi = 1, \chi = 0$ ) について計算結果を比較し、交叉の効果を検討した。

図 1 に平均適応度とその分散が、世代とともにどう変化していくかを示した。実線が交叉あり ( $\chi = 1$ )、点線が交叉なし ( $\chi = 0$ ) の結果である。明らかに交叉のある方が速く定常状態に収束していくことが分かる。もう一つ気が付くことは、適応度の分散は交叉がある方が大きく、その差は主に  $t \leq 20$  の領域で見ら

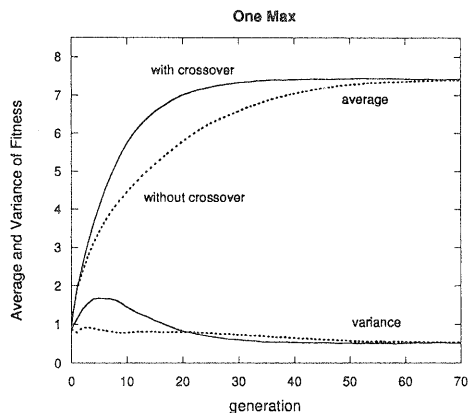


図 1 Average and variance of fitness in One Max problem.  $\ell = 8$ ,  $p = 0.01$ , and  $N = 200$ . Solid lines show the results with crossover, and dotted lines without crossover.

れるということである。

この結果をより詳しく検討するため図 2 に分散の内訳  $V_A$  と  $V_I$  を世代の関数として示した。実線が交叉あり、点線が交叉なしの結果である。分散  $V$  は交叉ありの場合は黒丸で、交叉なしの場合は白丸で表わした。二つの計算の最も顕著な差は  $V_I$  に表れている。交叉ありの計算では、初期 ( $t \leq 10$ ) を除いてほとんど 0 になるが、交叉がないと負の値をとりその絶対値は  $V_A$  の半分程度にもなる。このため結果として分散  $V$  は交叉がないと小さくなってしまふ。

Fisher の自然選択の基本定理によれば、適応度の分散が大きくなるとそれに比例して進化速度 (この場合は平均適応度の変化率) も大きくなる。そして今回の計算では、交叉により適応度の分散が大きくなることが分かった。一般に突然変異があると、Fisher の定理は厳密には成り立たないが、 $p$  が小さい場合は近似的な意味で成り立つ。

本論文では、GA において交叉が有効な働きをするか否かということを知るためには  $V_I$  の値が重要である、ということを示した。すなわち  $V_I$  が負ならば交叉によりその値が小さくなり、結果として分散が大きくなる。今後、他の問題などに今回の解析を適用し、より詳しく検討してみたい。

#### 参 考 文 献

- 1) Holland, J.H.: *Adaptation in Natural and Artificial Systems*, MIT Press, Massachusetts (1992).
- 2) Goldberg, D.E.: *Genetic Algorithms in Search,*

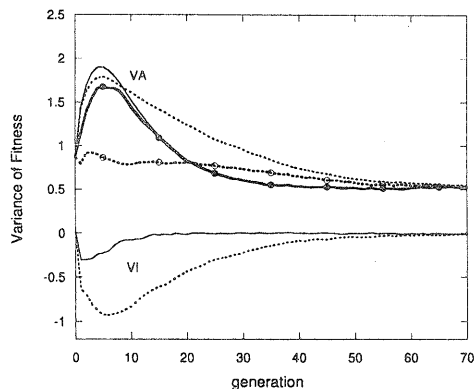


図 2 Variance of fitness  $V$  and its components  $V_A$  and  $V_I$ . Solid lines show the results with crossover and dotted lines those without crossover. Thick solid line with closed circles for the total variance  $V$  with crossover and thick dotted line with open circles for  $V$  without crossover.

*Optimization, and Machine Learning*, Addison Wesley, Massachusetts (1989).

- 3) Fisher, R.A.: *The Genetical Theory of Natural selection*, 2nd edition, Dover, New York (1958).
- 4) Maynard Smith, J.: *Evolutionary Genetics*, Oxford University Press, Oxford (1998).
- 5) Weir, B.S.: *Genetic Data Analysis II*, Sinauer Associates, Massachusetts (1996).
- 6) Furutani, H.: *Application of Eigen's Evolution Model to Infinite Population Genetic Algorithms with Selection and Mutation*, *Complex Systems*, Vol.10, pp.345-366 (1996).
- 7) Furutani, H.: *Analytical Solutions for Infinite Population Genetic Algorithms on Multiplicative Landscape*, *Proceedings of the Genetic and Evolutionary Computation Conference, GECCO-99 Morgan Kaufmann*, Vol.1, pp.204-211 (1999).
- 8) Goldberg, D.E.: *Genetic Algorithms and Walsh Functions: Part I, a Gentle Introduction*, *Complex Systems*, Vol.3, pp.129-152 (1989).
- 9) Vose, M.D.: *The Simple Genetic Algorithms*, MIT Press, Massachusetts (1999).
- 10) Furutani, H.: *Study of Evolution in Genetic Algorithms by Eigen's Theory Including Crossover*, *Proceedings of 2000 IEEE International Conference on Industrial Electronics and Instrumentation*, Nagoya, pp.2696-2703 (2000).