

連鎖不平衡係数を用いた交叉法の選択

橋口 悟[†] 川野泰雄^{††} 古谷博史^{††}

遺伝アルゴリズムにおいて最も重要な操作の1つである交叉の役割を連鎖不平衡係数を用いて研究した。線形の適応度関数をもつ OneMax 問題を研究対象として取り上げた。連鎖不平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かどうかを意味する。また、連鎖不平衡は適応度の分散、さらに平均適応度の増加、すなわち進化速度に影響を及ぼす。そこで、連鎖不平衡係数を用いて交叉法の違いによる進化速度の変化を計算した。一点交叉と一様交叉の2つの交叉法においては、一様交叉の方が良いという結果を得た。

Choice of Crossover Method by Use of Linkage Disequilibrium Coefficients

SATORU HASHIGUCHI[†], YASUO KAWANO^{††} and HIROSHI FURUTANI^{††}

Crossover is one of the most important operators in genetic algorithms. We studied the role of crossover by the use of linkage disequilibrium coefficients. We apply our method to OneMax problem, which has a linear fitness function. Linkage disequilibrium is an important notion in genetics and shows the correlation between different loci. It affects the variance of the fitness and also the increase of the average fitness, the speed of evolution. We used linkage disequilibrium coefficients to obtain the changes in evolution speed for different crossover operators. We compared the one-point and uniform crossover operators, and found that uniform crossover gives the better performance than one-point crossover.

1. はじめに

交叉は突然変異、選択とともに遺伝アルゴリズム (Genetic Algorithms, 以下 GA と略す) における重要な操作の1つであるが、その役割の解明が最も遅れている操作でもある。交叉には2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。本論文では、交叉の問題を連鎖不平衡¹⁾ という概念を手がかりに考えてみた²⁾。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 問題を例にとり³⁾、GA における交叉と連鎖不平衡の関係を交叉方法の違いにより検討する。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」⁴⁾をもとに進化速度と分散の関係、そして分散と連鎖不平衡との関係に注目し、交叉の方法によってそれらの値がどのように変化を調べ、各交叉法の性質を数値実験により明らかにする。

連鎖不平衡係数は、スキーマ頻度を用いて表される。

そのため、GA における連鎖の変化を調べるためには、スキーマの進化を理論的に予測することが重要である。最近、理論的研究の進展によりスキーマの進化を定量的に表すことが可能になった。この研究では、そうした研究の成果を利用しているが、その詳細については論文5)~7)を参照されたい。

2. 数学モデル

本論文では、GA における交叉の役割について考察する。交叉は一点交叉と一様交叉の2種類について調べることにし、交叉を行う親を選択する方法として適応度比例選択を用いる。

ここでは、これから議論する GA の枠組みを紹介し、使用する記号の定義を与える。集団の構成要素は世代ごとに入れ替わるものとし、進化の過程は差分方程式により記述する。集団の個体数は時間的に一定とする。個体は ℓ ビット固定長の2進ビット列で表し、個体の遺伝子型と呼ぶ。すなわち、 i 番目の遺伝子型 B_i に対する2進ビット列は

$$i = (i_1, i_2, \dots, i_{\ell-1}, i_{\ell})$$

と表される。ここで

$$i_k = 0 \text{ or } 1 \quad (k = 1, \dots, \ell)$$

[†] 京都教育大学大学院
Graduate School, Kyoto University of Education

^{††} 京都教育大学教育学部
Faculty of Education, Kyoto University of Education

は、 i の第 k ビット目における値を表す。また、この i_k の値を第 k 遺伝子座における対立遺伝子の種類と見なすこともある。遺伝子座の種類の総数は $n = 2^\ell$ である。2 進ビット列 i に含まれるビット 1 の数を

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

と表す。

集団の分布は相対頻度

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

により表す。ここで N_i は第 i 遺伝子型をもつ個体数を表す。

3. 連鎖不平衡係数と交叉

3.1 連鎖不平衡

交叉の役割を考察する上で非常に重要な概念として連鎖不平衡 (linkage disequilibrium) がある⁴⁾。連鎖不平衡とは、2 個所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ない場合は連鎖平衡という。一般に、交叉は遺伝子座間の連鎖を弱める働きをし、交叉を繰り返して適用すると最終的に連鎖平衡の状態に収束する。これは遺伝子の頻度分布に注目して定義される。

いま $\ell \geq 2$ の個体からなる集団を考え、その中の 2 つの遺伝子座 m, m' に注目する。集団が、遺伝子座 m においてビット値 1 をとる相対頻度を $p_m^{(1)}$ と表す。また、遺伝子座 m, m' においてビット値 1 をとる相対頻度を $p_{m,m'}^{(11)}$ とする。連鎖不平衡を定量的に表すため、連鎖不平衡係数を定義する。連鎖不平衡係数としてよく使われているのは D 係数で

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

と定義される。2 つの遺伝子座が独立、すなわち連鎖平衡であれば $p_{m,m'}^{(11)} = p_m^{(1)} p_{m'}^{(1)}$ となり、 $D[m, m'] = 0$ が成り立つ。

\hat{C} は交叉率 $\chi = 1$ における交叉の効果を表すものとする、連鎖不均衡係数に対する効果は次式で与えられる

- 一様交叉の場合

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

- 一点交叉の場合

$$\hat{C}D[m, m'] = \left(1 - \frac{|m - m'|}{\ell - 1}\right)D[m, m']. \quad (3)$$

いずれも連鎖不平衡係数 D の絶対値を小さくする効果がある。したがって、交叉が十分に働けば、 $D = 0$

と仮定することができる。また、高次の連鎖不平衡係数も全て 0 になれば

$$x_i = \prod_{m=1}^{\ell} p_m^{(i_m)}$$

と表すことができる。これを連鎖平衡状態と呼ぶ。

3.2 Fisher の定理

平均適応度の時間変化から、Fisher の自然選択の定理について述べる。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

$$x_i(t+1) = \frac{f_i}{\bar{f}(t)} x_i(t) \quad (i = 0, \dots, n-1). \quad (4)$$

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

$$\bar{f}(t) = \sum_{i=0}^{n-1} f_i x_i(t). \quad (5)$$

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⁴⁾。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。

$$\Delta \bar{f}(t) = \frac{1}{\bar{f}(t)} \text{VAR}(f), \quad (6)$$

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

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

$$v(t) \equiv \Delta \bar{f}(t) = \bar{f}(t+1) - \bar{f}(t) \quad (7)$$

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。

4. OneMax 問題

OneMAX 問題の適応度は

$$f_i = |i| = \sum_{k=1}^{\ell} i(k), \quad (8)$$

と定義される。この場合、前述の分散は、

$$\text{VAR}(f) = V_A + V_I$$

と表すことができ³⁾、

$$V_A = \sum_{k=0}^{\ell} p_k^{(1)} (1 - p_k^{(1)})$$

$$V_I = 2 \sum_{k < k'} D[k, k']$$

となる。

V_A は $p_k^{(1)}$ にのみ依存し、交叉の影響は受けない。一方、 V_I は 2 つの遺伝子座間の相関に依存し、 $D[k, k']$ の関数である。したがって、交叉によって値が変化する。

スキーマ理論の研究から、1 次のスキーマに対する進化方程式を導くことができる⁸⁾。連鎖平衡状態において、 $p_k^{(1)}$ のビット位置依存性を無視し、

$$P_1(t) = p_k^{(1)}$$

と近似する。OneMax 問題における $P_1(t)$ の進化方程式は

$$P_1(t+1) = P_1(t) + \frac{1}{\ell} \{1 - P_1(t)\}$$

であり、その解は

$$P_1(t) = 1 - (1 - \frac{1}{\ell})^t \{1 - P_1(0)\} \quad (9)$$

となり、 $t \rightarrow \infty$ で $P_1(t) \rightarrow 1$ に収束する。

5. 結果とまとめ

これまで述べてきた理論を数値実験により検証する。特に交叉法によって、進化過程にどのような影響がみられるか、連鎖不平衡係数の変化に注目して調べる。今回はビット長 $\ell = 30$ 、個体数 $N = 200$ とし、乱数を変えながら同じ計算を 100 回繰り返して平均をとった。各ビット列における 1 の出現確率 $Q = 0.25$ とし、交叉率を $\chi = 1$ と $\chi = 0.1$ の場合についての計算を比較し、交叉の効果調べた。なお、今回は突然変異、選択は組み入れず交叉単独で計算を行った。また、親の選択には適応度比例選択を用いた。

図 1 は各世代の平均適応度を交叉法、交叉率を変えて計算したものである。理論値は、連鎖平衡を仮定した近似式 (9) を用いた。交叉法については一様交叉の方が進化速度は速く、また交叉率 $\chi = 1$ すなわち完全交叉の場合が理論値に近い進化速度になっていることが分かる。このことは、交叉により集団が連鎖平衡状態に近づいていくことから理解できる。

図 2 は平均適応度の分散を計算したものである。0 世代目から 70 世代目にかけて一様交叉の方が分散値が大きくなっているのが分かる。Fisher の定理から、一様交叉のほうが平均適応度の進化にとって有利であることを示している。図 1 と併せて比較すると、分散と進化速度は比例関係にあることがいえる。

図 3 は連鎖不平衡係数 D を隣り合ったビット列 [1, 2] について計算したものである。一点交叉については、隣り合ったビット列の場合、交叉点とならない場合があ

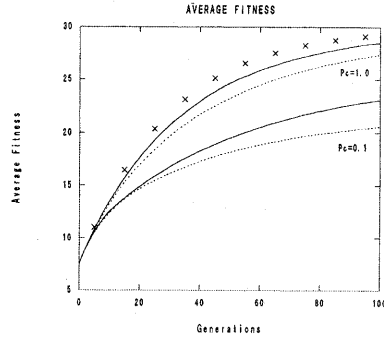


図 1 平均適応度。 $\ell = 30$, $N = 200$, $Q = 0.25$ とし、 $\chi = 0.1$, 1.0 の 2 通りについて計算。100 回の計算を平均値を一様交叉の場合は実線で、一点交叉の場合は点線で、理論値は \times で示す。

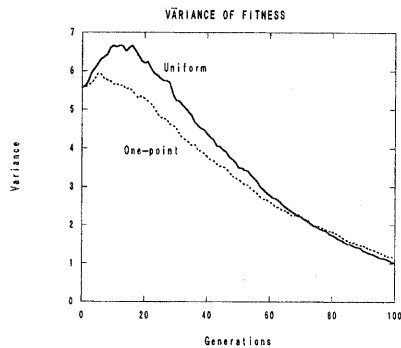


図 2 適応度の分散。 $\ell = 30$, $N = 200$, $Q = 0.25$ とし、 $\chi = 1.0$ の場合について計算。

るため、交叉が有効に働かず、 D の値が負に大きくなる。また一様交叉においては、交叉マスクを作って交叉点を決めるために、交叉点となる確率が $\frac{1}{2}$ と一点交叉よりも大きくなるので、 D は 0 に近づいていく。

図 4 は連鎖不平衡係数 D を一番離れたビット列 [1, 30] について計算したものである。図 3 と比較すると図 4 の方が D の絶対値が小さくなっている。これは、一点交叉が大きな効果をもたらしていることが関係している。一方、一様交叉は、絶対値は一点交叉に比べて大きくなっている。以上より、連鎖不平衡係数で見ると一点交叉に比べて一様交叉の方が効果があることがいえる。

今回は連鎖不平衡係数によって一点交叉と一様交叉の比較を行った。しかし、今回の実験では交叉単独による操作しか扱わなかったため、突然変異や選択と組み合わせた場合に連鎖不平衡係数がどうなるかは不明である。今後は交叉と他の操作と組み合わせ、連鎖

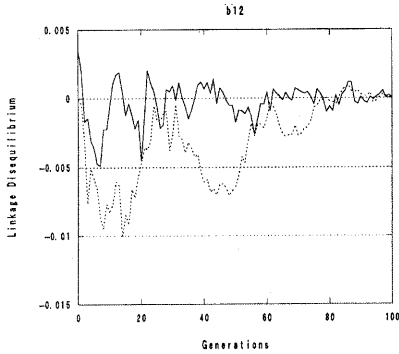


図3 b₁₂における連鎖不平衡係数. パラメータは図2と同じ.

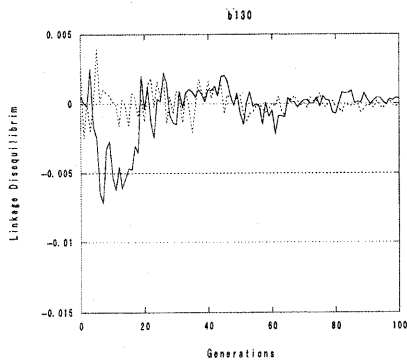


図4 b₁₃₀における連鎖不平衡係数. パラメータは図2と同じ.

不平衡係数や進化過程がどのように変化するかを調べてみたい.

参考文献

- 1) Maynard Smith, J.: Evolutionary Genetics, Oxford University Press, Oxford (1998).
- 2) 古谷博史, 藤林由紀, 村田真知子: “遺伝的アルゴリズムにおける交叉の役割の連鎖解析,” 情報処理学会研究会報告, **2001-MPS-33** (2001) 53-56.
- 3) Furutani, H: “Study of Crossover in One Max Problem by Linkage Analysis,” L. Spector et al. (Eds.) Proceedings of the Genetic and Evolutionary Computation Conference, GECCO-2001, Morgan Kaufmann, (2001) 320-327.
- 4) R.A. Fisher: The Genetical Theory of Natural selection, 2nd edition, Dover, New York (1958).
- 5) Vose, M.D.: The Simple Genetic Algorithms, MIT Press, Massachusetts (1999).
- 6) 古谷博史: “Walsh 変換による突然変異と交叉に対するスキーマ定理の導出,” 情報処理学会論文誌, **43(4)** (2002) 1050-1060.
- 7) 古谷博史: “遺伝的アルゴリズムにおける交叉の Walsh 解析,” 情報処理学会論文誌, **42** (2001)

2270-2283.

- 8) 古谷博史: “遺伝的アルゴリズムのスキーマ定理による解析: 突然変異と交叉の役割,” 情報処理学会論文誌: 数理モデル化と応用, **43** (2002) 35-45.