

遺伝的アルゴリズムの収束時間に関する考察

6Q-5

丹羽 竜哉, 田中 勝

電子技術総合研究所

1. はじめに

遺伝的アルゴリズム(GA)は、生物における遺伝や進化の現象を模してはたらくアルゴリズムである。遺伝に関しては、集団遺伝学の分野で様々な数学的解析がなされており、遺伝的アルゴリズムの解析にも役立つものが多い。

我々は既に遺伝的浮動(Random Genetic Drift: 遺伝子選択の偶然性などの理由で、集団中の遺伝子頻度が揺らぐ現象)について解析を行なっている[4][5]。本稿では、遺伝的浮動と共に観察される収束(吸収)時間の変動について、とくに染色体の長さに対して考察する。また、突然変異率 μ と集団サイズ N の関係式 $N\mu = 1/2$ の妥当性について、実際にGAを動かすことにより確認する。

2. 平均収束時間

ライト-フィッシャー・モデル 集団サイズを世代に関わらず一定 N とし、各個体は長さ l の染色体を持ち、対立遺伝子数が2で各遺伝子の適応度には差がない場合を考える。世代交代は不連続で、 N 個から N 個を重複を許して選ぶという方法で行なわれる場合、集団中に含まれる各遺伝子の数を状態とする状態遷移行列 Q は

$$Q_{ij} = \binom{N}{j} \left(\frac{i}{N}\right)^j \left(1 - \frac{i}{N}\right)^{N-j}$$

となる。このモデルのことを、ライト-フィッシャー・モデル(Wright-Fisher model)と呼ぶ。

平均収束(吸収)時間 遺伝的浮動による揺らぎのため、適応度に差がない場合でも、最終的に集団中から一方の遺伝子が消滅する。ライト-フィッシャー・モデルの場合、集団中から一方が消滅するまでの平均時間(平均吸収時間)は、状態遷移行列 Q から第0行第0列第 N 行第 N 列を取り去った行列を $\tilde{Q}$ とし、 ξ を全ての要素が1のベクトル、 I を単位行列として、 $(I - \tilde{Q})^{-1}\xi$ を計算することにより求められる。また[1]によれば、集団サイズ N について連続近似することによって、

$$\sum_{j=0}^{\infty} \frac{P_{2j}(1-2p) - P_{2j+2}(1-2p)}{(j+1)(2j+1)} N$$

A Study of Convergence Time on Genetic Algorithms
Tatsuya NIWA and Masaru TANAKA
Electrotechnical Laboratory

(ただし、 P_j はルジャンドル多項式、 p は初期状態の成分比を示す)となることがわかる。これを $p = 1/2$ の時について数値に直すとおよそ $1.386N$ となり、数値シミュレーション[2][3]で示された値 $1.4N$ にはほぼ一致する。

3. 長い染色体の場合

一般にGAでは長い染色体を扱うが、これを交叉の影響を含めて上記の様な数学的解析をすることは、状態数が爆発的に増加するため、事実上不可能である。ところが、一様交叉の場合に限り、各々の遺伝子座は独立に遷移すると考えられるので、ライト-フィッシャー・モデルの解析結果より計算が可能である。

集団サイズ $N = 2$ で染色体の長さが l のときを考える。時刻 t までに全部の遺伝子座について吸収(convergence)状態になる確率は $(1 - 1/2^t)^l$ であるから、ちょうど時刻 t に全部の遺伝子座について吸収状態になる確率は

$$\left(1 - \frac{1}{2^t}\right)^l - \left(1 - \frac{1}{2^{t-1}}\right)^l$$

従って、平均吸収(収束)時間 E は

$$E = \sum_{t=1}^{\infty} t \left\{ \left(1 - \frac{1}{2^t}\right)^l - \left(1 - \frac{1}{2^{t-1}}\right)^l \right\}$$

これを整理すると、

$$E = \sum_{t=0}^{\infty} \left\{ 1 - \left(1 - \frac{1}{2^t}\right)^l \right\}$$

を得る。ここで連続近似を行ない、 $\tilde{E}$ とする。

$$\tilde{E} = \int_0^{\infty} \left\{ 1 - \left(1 - \frac{1}{2^t}\right)^l \right\} dt + \frac{1}{2}$$

これを計算すると

$$\tilde{E} = \frac{1}{\log 2} \sum_{i=1}^l \frac{1}{i} + \frac{1}{2}$$

となる。さて、公式によれば、

$$\sum_{i=1}^l \frac{1}{i} = \psi(l) + \frac{1}{l} + \gamma$$

$$\approx \log l + \gamma + \frac{1}{2l} - \sum_{i=1}^{\infty} \frac{B_{2i}}{2i l^{2i}}$$

(ただし, γ はオイラー定数, ψ はディガンマ関数, B_i はベルヌーイ数) であるから,

$$\bar{E} \simeq \frac{1}{\log 2} \left(\log l + \gamma + \frac{1}{2l} - \frac{1}{12l^2} + \frac{1}{120l^4} + \cdots \right) + \frac{1}{2}$$

となり, l が十分大きいとき $E \simeq \bar{E}$ は $\log l$ に比例することがわかる. すなわち, 数値計算による予想 [3] は正しいことが確認できる.

4. 突然変異率

突然変異がある場合には一度消滅した遺伝子が復活する場合も起こりうる. しかし, 突然変異率が大きすぎると分布が中央部分に寄り, アルゴリズムとして収束しにくくなる. 前著 [5] において, 突然変異率 μ が集団サイズ N に対し $\mu = 1/2N$ を満たすとき状態 (集団における遺伝子の数) の頻度分布が一様となるため, 遺伝的アルゴリズムの突然変異率の参考値となることを理論的に示した. 今回, これに関して実際に GA を動かすことにより確認を行なった.

GA の各世代毎の状態の統計を十分に長い世代に渡って取り, その頻度分布を見るという方法で行なった. まず, 染色体長 1 ビットで適合値なしの場合, 突然変異率 $\mu = 1/2N$ で定常状態が一様になるかを調べた. 両端 (集団全体が一種類の遺伝子で満たされた状態) において頻度がやや小さくなる傾向があるものの, 全体ではほぼ一様であることが確認できた (結果の図示は省略する).

次に, 適合値を評価する場合について, 上と同様の方法で調べた. 染色体長が 2 ビットの場合について, 図に示す. 図 a の黒丸で示した適合値は (0,0):5, (1,0):3, (0,1):4, (1,1):6 という値で, 一種のだまし関数である. 図 b は, GA を 1000000 世代まで繰り返した時の頻度分布である. ただし, 交叉は一様交叉 (1/2 の確率で組み換えが起きる) で, 集団サイズは 10 である. 適合値の差をやや強調し過ぎる傾向があるものの, 図 a の四つの黒丸を結ぶ曲面に沿うような頻度分布になっていることが観察できる.

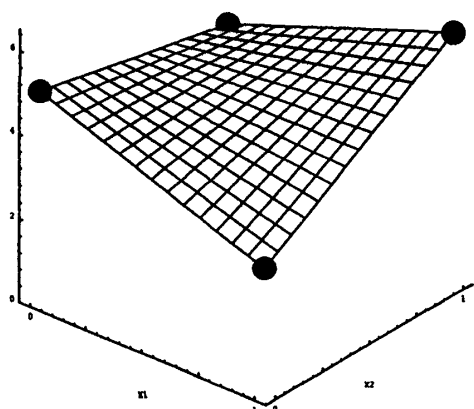


図 a 適合値

5. おわりに

染色体長 l が長い場合の平均収束 (吸収) 時間が $\log l$ に比例することを, 集団サイズ $N = 2$ の場合について解析的に示した. ただし, $N > 2$ の場合については今後の課題である. また, 突然変異率を $\mu = 1/2N$ とすると, 集団の状態の頻度分布が適合度に沿う形となることを確認した. この結果は, GA の染色体の構造や数々の GA パラメータの設定に役立つものと考えられる.

謝辞 本研究に関し様々な議論や資料を下さった電総研の方々, また本研究の機会を与えて下さった太田部長, 諏訪部長, 大津部長他関係者の方々に感謝の意を表す.

参考文献

- [1] M. Kimura, "Diffusion Models in Population Genetics", *J. Appl. Prob.*, 1, 177-232, 1964.
- [2] D. E. Goldberg and P. Segrest, "Finite Markov Chain Analysis of Genetic Algorithms", *Proceedings of the 2nd International Conference on Genetic Algorithms*, 1-8, 1987.
- [3] H. Asoh and H. Mühlenbein, "On the Mean Convergence Time of Evolutionary Algorithms Without Selection", to appear in *Parallel Problem Solving from Nature 3 (proceedings)*, 1994.
- [4] M. Tanaka and T. Niwa, "Markov Chain Analysis on Simple Genetic Algorithm", *Technical Report*, ETL-TR-94-13, 1994.
- [5] 丹羽, 田中, "遺伝的アルゴリズムにおける遺伝的浮動の一考察", 情報処理学会第 49 回全国大会論文集 (2), 243, 1994.

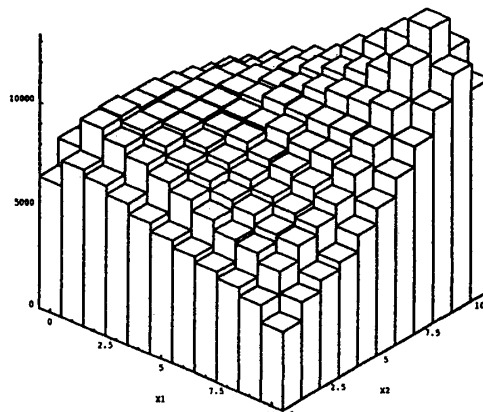


図 b 状態頻度分布