

ONEMAX 問題における成功確率

杜 一飛^{†1} 古賀 仁信^{†2}
 坂 本 眞 人^{†1} 古 谷 博 史^{†1}

本研究では、遺伝的アルゴリズム (GA) の計算性能を評価する指標として成功確率を取り上げ、その解析を行った。ここで、成功確率を集団中に少なくとも一つの最適解が存在する確率と定義する。そのため、Markov 連鎖を利用して定常分布を求め、実験で得られた 1 次スキーマ分布と比較した。最後に、ONEMAX 問題の 1 次スキーマ分布を利用して最適解の出現分布を求めた。

Success probability in ONEMAX problem

YIFEI DU,^{†1} KIMINOBU KOGA,^{†2} MAKOTO SAKAMOTO^{†1}
 and HIROSHI FURUTANI^{†1}

In this study, we analyzed the success probability as a measure of the performance of genetic algorithm(GA). We define the success probability as that there is at least one optimal solution in the population. To this end, we calculated the stationary distribution by using Markov chain, and compared it with the first-order schema frequency obtained in numerical experiments. Finally, we calculated the distribution of the optimal solution from the first-order schema frequencies in ONEMAX problem.

1. はじめに

遺伝的アルゴリズム (GA) において、計算で得られた集団の中に最適解が存在するか、ということは切実な問題である。本研究では ONEMAX 問題を取り上げ、Markov 連鎖理論と最適解分布の計算法を利用して成功確率を求め、数値実験と比較した。一般に、GA の進化を理論的に解析することは非常に困難である。そのため、この方法では集団が連鎖平衡にあることを仮定し、1 次スキーマの確率分布を求める。その 1 次スキーマ分布から、集団中の最適解の個数分布を計算した。

2. 数学モデル

2.1 数学的表現

進化のモデルとして Goldberg の Simple Genetic Algorithm (SGA) を用い、ONEMAX 問題について解析した。集団の個体数は世代に依らず一定で N とする。個体の遺伝子型を固定長 ℓ の 2 進ビット列で表す。遺伝子型の番号 i ($0 \leq i \leq n - 1$) と 2 進ビット

列を

$$i = \langle i(\ell), i(\ell - 1), \dots, i(2), i(1) \rangle,$$

のように対応させる。 $i(k)$ は第 k ビットの値である。ONEMAX 問題の適応度関数を

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

と定義する。適応度は全てのビットが 1 のとき最大となる。

2.2 連鎖平衡

連鎖は、集団内での異なる遺伝子座間の統計的相関を意味し、遺伝子座間に何らかの相関があるとき、その集団は連鎖不平衡状態という。逆に相関がない状態を連鎖平衡と呼ぶ。連鎖不平衡は、遺伝子座間に強い相関をもたらす適応度関数を用いた場合、および遺伝的揺らぎの効果により引き起こされる。交叉と突然変異は遺伝子座間の連鎖を弱め、集団を連鎖平衡の状態へ導く。

集団が連鎖平衡の状態であれば、個体の分布は 1 次スキーマの頻度によりのみ依存する。1 次スキーマの進化方程式は

$$h_1(t + 1) = ah_1(t) + b, \quad (1)$$

により与えられる²⁾。 $h_1(t)$ はビット 1 をもつ 1 次スキーマの相対頻度を表す。係数 a と b は

$$a = \left(1 - \frac{1}{\ell}\right)(1 - 2p_m), \quad b = \frac{1}{\ell}(1 - 2p_m) + p_m, \quad (2)$$

によって与えられ、 p_m は突然変異率を表す。

^{†1} 宮崎大学工学部

Faculty of Engineering, University of Miyazaki

^{†2} 宮崎大学工学研究科

Graduate School of Engineering, University of Miyazaki

3. Markov モデル

集団遺伝学における Markov モデルの 1 つに、Wright-Fisher モデルがある⁴⁾。ここでは、Wright-Fisher モデルを ONEMAX 問題に適用する。

世代 t において遺伝子型 1 の個体数が $N_1 = i$ のとき、次世代 $t+1$ で $N_1 = j$ となる条件付き確率は

$$P_{i,j} = P(j|i) = \binom{N}{j} p_i^j (1-p_i)^{N-j} \quad (3)$$

$$p_i = ay + b = a \left(\frac{i}{N} \right) + b$$

と表される。

世代 t において遺伝子型 1 の個体数が i 個となる確率を $\mu_i(t)$ とする。この $\mu_i(t)$ をベクトル形式で次のように表す。

$$\mu^T = (\mu_0, \mu_1, \dots, \mu_N)$$

また、正規化条件は

$$\sum_{i=0}^N \mu_i(t) = 1 \quad (4)$$

となる。

進化方程式はベクトル形式

$$\mu(t+1)^T = \mu(t)^T P \quad (5)$$

で与えられる。Markov 連鎖は定常分布 π に収束する。

$$\lim_{t \rightarrow \infty} \mu(t) = \pi \quad (6)$$

4. 成功確率

本研究では成功確率 S を集団中に少なくとも 1 つの最適解が存在する確率と定義する。一方、集団中に最適解が存在しない状態の確率 F を失敗確率と呼ぶ。また、集団中の長さ ℓ の部分列のうち j 個が部分最適解である確率を $S_j^{(\ell)}$ と表す。成功確率と失敗確率は

$$S = \sum_{j=1}^N S_j^{(L)}, \quad F = S_0^{(L)} \quad (7)$$

から計算できる。このように、 S と F を求めるには $S_j^{(\ell)}$ が必要である。詳しくは論文 3) を参照されたい。

5. 計算と実験の結果

数値実験では、適応度の初期値を $L/2$ に設定し、1 次スキーマの分布と成功確率を調べた。集団のサイズ $N = 100$ 、ビット長 $L = 20$ 、突然変異 $p_m = 0.01$ とし、交叉は一様交叉を用いた。実験の結果と理論計算の結果を比べた。

図 1 と図 2 に 1 次スキーマの分布と最適解の出現分布を示し、ほぼ同じ振る舞いをすることを確かめた。

6. まとめと今後の課題

今回の研究では、成功確率の理論値を求め、その結果が ONEMAX 問題における数値実験の結果とほぼ一致することを確認した。この計算方法で、ある程度遺伝的アルゴリズム (GA) の計算性能を評価する指

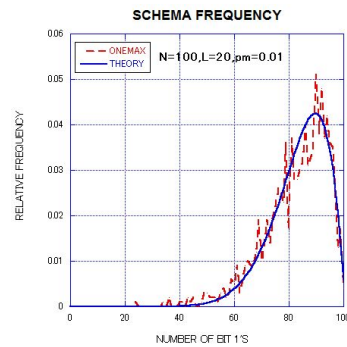


図 1 1 次スキーマの分布, ONEMAX : 実験結果, THEORY : 計算結果

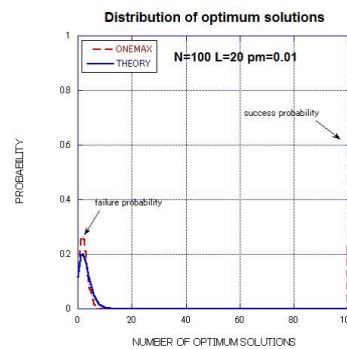


図 2 最適解の出現分布, ONEMAX : 実験結果, THEORY : 計算結果, 成功確率と失敗確率を示した

標としての成功確率を解析できることが分かった。今後の課題として、様々な設定による数値実験を行い、理論計算と比較していく必要がある。また、より複雑な問題に対し、今回の方法を拡張していきたい。

参考文献

- 1) 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「数理モデル化と応用」, **43**, pp.1050-1060, 2002
- 2) 古谷博史: 選択における連鎖不平衡の効果-OneMax 問題のスキーマ解析, 情報処理学会論文誌「数理モデル化と応用」, **45**, pp.12-21, 2004
- 3) 古谷博史, 張玉安, 坂本真人: 遺伝的アルゴリズムにおける最適解出現分布のマルコフ連鎖による研究, 情報処理学会論文誌「数理モデル化と応用」, **2**, pp.54-63, 2009
- 4) Q. Ma, Y. Zhang, K. Koga, K. Yamamori, M. Sakamoto, H. Furutani: Stochastic analysis of OneMax problem by using Markov chain, The Seventeenth International Symposium on Artificial Life and Robotics, pp.19-21, 2012