

A-024

推定誤差を有する適応度関数の実数値遺伝的アルゴリズムによる最適化 Real-Coded Genetic Algorithm to Optimize Fitness Function with Sampling Error

手塚 大†
Masaru Tezuka

棟朝 雅晴‡
Masaharu Munetomo

赤間 清‡
Kiyoshi Akama

1. はじめに

実世界の最適化問題では、評価にモンテカルロ・シミュレーションを用いるなど、複数個のサンプルから適応度の推定を行うことがある。最適化に遺伝的アルゴリズムを用いる場合、推定に用いるサンプルが少ないと推定誤差のために個体の選択に誤りが発生し、最適解を得るまでの世代交代数が増加する。サンプル数を増やすと推定誤差が小さくなるが、サンプリングに要する時間が増える。実問題では計算に使える時間が限られている。この限られた時間を、サンプリング回数と世代交代数にどのように配分すると最も効率的に最適化が行えるかが課題となる。

本報告ではトーナメント選択を用いる実数値遺伝的アルゴリズムで、推定誤差による選択誤り確率と、最良個体が集団全てを占めるのに要する時間である take-over 時間の関係から、最も選択効率が良いサンプリング回数を示す。

2. Take-over 時間と選択効率

1 個体あたりのサンプリング回数を s 、世代交代数を t 、集団の個体数を n とする。また、遺伝的アルゴリズムで 1 個体を生成するのに要する計算時間を α 、サンプリング 1 回に要する時間を β とすれば、総計算時間 $T_e = nt(\alpha + s\beta)$ となる。

遺伝的アルゴリズムで選択のみを適用し、集団中の $n-1$ 個体が最良個体となるまでの時間を take-over 時間という。ランダムに選ばれた 2 個体のうち適応度の高い個体が確率 p ($0.5 < p \leq 1.0$) で選択される確率的トーナメント選択での take-over 時間 t^* は次式で表される [1]。

$$t^* = \frac{2}{2p-1} \log_2(n-1) \quad (1)$$

take-over 時間を最適化に要する時間の指標とし、take-over が多く発生するほど選択が進むと考える。

また、確率的トーナメント選択の選択誤りが、適応度の推定誤りにより生じるものとする。確率 p は s に依存し、サンプリング回数が多いほど $p(s)$ が 1.0 に近づく。

つまり、選択効率は次式で表される。

$$\eta = \frac{t}{t^*} = \frac{(2p(s)-1)T_e}{2(\alpha+s\beta)n \log_2(n-1)} \quad (2)$$

ここで、 $\alpha \ll \beta$ で $\alpha = 0$ とみなせる場合を考える。総計算時間 T_e と集団サイズ n が決められているとすると

$$\eta = \frac{2p(s)-1}{s} \quad (3)$$

が最大のときに選択効率 η も最大となる。

†(株)日立東日本ソリューションズ
‡北海道大学

3. 期待値の最大化・最小化

個体 i から得られるサンプルが正規分布 $N(\mu_i, \sigma_i^2)$ に従うとし、期待値を最大化あるいは最小化する場合を考える。期待値の s 個のサンプルからの不偏推定は、個体 i の k 番目のサンプルを h_{ik} とすると

$$\hat{\mu}_i = \frac{1}{s} \sum_{k=1}^s h_{ik} \quad (4)$$

である。この推定値は正規分布にしたがう。したがって、個体 i の真の期待値が個体 j よりも大きい場合 ($\mu_i > \mu_j$) に、不偏推定も正しく個体 i が大きくなる ($\hat{\mu}_i > \hat{\mu}_j$) 確率 $p(s, i, j)$ は、

$$p(s, i, j) = \Phi\left(\frac{\mu_i - \mu_j}{\sqrt{\frac{\sigma_i^2 + \sigma_j^2}{s}}}\right) = \Phi(\lambda_{ij}\sqrt{s}) \quad (5)$$

である。ここで、 $\Phi(z)$ は標準正規分布の ∞ から z までの累積確率であり、 $\lambda_{ij} = (\mu_i - \mu_j) / \sqrt{\sigma_i^2 + \sigma_j^2}$ とおいた。

実際には λ_{ij} は比較を行う個体の組み合わせによって異なるが、仮に全ての個体間で等しいとすると、 $\eta = (2\Phi(\lambda\sqrt{s}) - 1) / s$ が最大するとき、選択効率が最大となる。 s は 1 以上の整数であり、 $s = 1$ のときに最大となる。つまり、期待値の最大化や最小化の問題ではサンプリング回数が $s = 1$ のとき、選択効率が最も高いと考えられる。

4. 分散の最小化

個体 i から得られるサンプルが正規分布 $N(\mu_i, \sigma_i^2)$ に従い、分散を最小化する問題を考える。分散の不偏推定は次式で表される。

$$\hat{\sigma}_i^2 = \frac{1}{s-1} \sum_{k=1}^s (h_{ik} - \hat{\mu}_i)^2 \quad (6)$$

ここで、次に示す確率変数 $F_{ij,s}$ は自由度対 $[s-1, s-1]$ の F 分布に従う。

$$F_{ij,s} = \frac{\hat{\sigma}_i^2 / \sigma_i^2}{\hat{\sigma}_j^2 / \sigma_j^2} \quad (7)$$

したがって、個体 i の真の分散が j よりも小さい ($\sigma_i^2 / \sigma_j^2 < 1.0$) 場合に、分散の不偏推定値も正しく個体 i が小さくなる ($\hat{\sigma}_i^2 / \hat{\sigma}_j^2 < 1.0$) 確率 $p(s, i, j)$ は自由度対 $[s-1, s-1]$ の F 分布の 0 から σ_i^2 / σ_j^2 までの累積確率となる。ここで、サンプリング回数と分散の比ごとの式 (3) の η の値を、表 1 に示す。おおむね真の分散の比が 3 以上ある場合にはサンプリング回数 $s = 2$ で、分散の比が 3 より小さい場合には $s = 3$ で選択効率が最大になると考えられる。

表 1: 分散最小化での $\eta = (2p(s) - 1)/s$

s	σ_j^2/σ_i^2					
	1.001	1.01	1.1	2	3	5
2	1.59	1.58	1.52	1.08	1.67	2.32
3	1.67	1.66	1.59	1.11	1.67	2.22
4	1.59	1.58	1.52	1.04	1.52	1.95
6	1.41	1.41	1.35	0.89	1.24	1.50
12	1.08	1.07	1.02	0.61	0.77	0.82
	$\times 10^{-4}$	$\times 10^{-3}$	$\times 10^{-2}$	$\times 10^{-1}$	$\times 10^{-1}$	$\times 10^{-1}$

5. 数値実験

1 個体あたりのサンプリング回数と選択効率について数値実験を行った。実験では次式で表される Sphere-d 関数 f_s , $d = 10$ を用いた。

$$f_s(x_1, \dots, x_d) = \sum_{i=1}^d x_i^2 \quad (8)$$

実数値 GA を用い、個体 i の染色体を実数値ベクトル $\mathbf{x}_i \in \mathbb{R}^d$ とする。期待値の最小化と、分散の最小化の二つの実験を行った。期待値最小化の実験では式 (9) に、分散最小化の実験では式 (10) に従う乱数を、個体から得られるサンプルとした。

$$h_{ik} \sim N(f_s(\mathbf{x}_i) + 100.0, 10.0^2) \quad (9)$$

$$h_{ik} \sim N(100.0, f_s(\mathbf{x}_i) + 10.0^2) \quad (10)$$

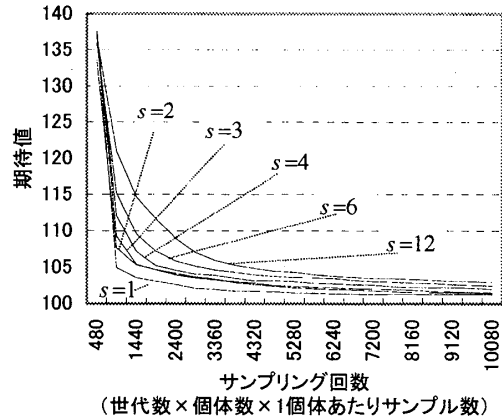
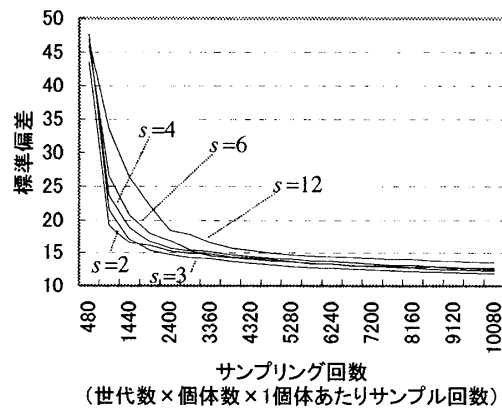
このサンプルから式 (4), (6) により期待値、分散を推定し適応度とする。

交叉には BLX- α [2] を用いた。BLX のパラメータは理論上の最適値とされる $\alpha = 0.366$ [3] とした。正規分布 $N(0, \sigma^2)$ に従う突然変異を個体の各要素ごとに発生させた。交叉率を 0.6, 突然変異率を 0.1, 突然変異の分散は $\sigma^2 = 1.0 \times 10^{-4}$, 集団の個体数を 40 とした。

期待値最小化の実験結果を図 5., 分散最小化の結果を図 5. に示す。それぞれ、各 s について 50 試行を行い、各世代の真の期待値, 真の分散の最良個体について, 50 試行の平均値を求め, 縦軸に示した。なお, 図 5. のグラフは標準偏差を示している。グラフの横軸はサンプリング回数 $n \times t \times s$ で揃えてある。実験結果から, 最も選択効率が高いサンプリング回数は期待値の最小化では $s = 1$ であることが確認された。また分散の最小化では, 最適化の序盤では $s = 2$ が早く最適化が進むがサンプリング回数が 1400 回を超えたあたりから $s = 3$ のほうが良い値となる。これは, 最適化の初めの段階では個体がランダムに生成されるため個体間の分散の比が大きいが, 最適化が進むにつれて個体間の差が小さくなり, 分散の比が小さくなるためと考えられる。

6. おわりに

本報告では, サンプルからの推定値を適応度とする遺伝的アルゴリズムで, 最も選択効率が高くなる 1 個体あたりのサンプリング回数を, take-over 時間をもとに求め

図 1: 1 個体あたりサンプリング回数 s と期待値の最小化図 2: 1 個体あたりサンプリング回数 s と分散の最小化

た。この結果, 期待値の最大化および最小化ではサンプリング回数 1 が最も選択効率が高いことが示された。また, 分散の最小化では, 個体間の分散の比が小さい場合は 3 回, 大きい場合は 2 回のサンプリングが効率が良いことが示された。数値実験により, これらのサンプリング回数の際に選択効率が最大となることを確認した。

参考文献

- [1] David E. Goldberg and Kalyanmoy Deb. A comparative analysis of selection schemes used in genetic algorithms. In Gregory J. E. Rawlins, editor, *Foundations of Genetic Algorithms*. Morgan Kaufman, 1991.
- [2] Larry J. Eshelman and J. David Schaffer. Real-coded genetic algorithms and interval-schemata. In L. Darrell Whitley, editor, *Foundations of Genetic Algorithms 2*. Morgan Kaufman, 1993.
- [3] 樋口隆英, 筒井茂義, 山村雅幸. 実数値 GA におけるシンプレックス交叉の提案. 人工知能学会論文誌, Vol. 16, No. 1, pp. 147-155, 2001.