

1M-8

GAにおける集団の時間発展に対する 初期集団と集団サイズの影響について

吉澤大樹

橋本周司

早稲田大学理工学部

1. はじめに

遺伝的アルゴリズム（以下 GA）は大域的な探索を実現していると考えられているが、実際に試行するとしばしば初期収束が見られ、その挙動は必ずしも明確ではない。またGAの探索空間はn次元超立方体で表わされる。しかしこの表現では探索の推移を見るのは困難である。ここでは集団分布の確率的挙動に注目してGAの時間発展について調べる。

2. モデル

世代tにおける集団分布の期待値を P^t とした場合の P^t の推移を関数 $G: P^t \rightarrow P^{t+1}$ について考える。例えば、ルーレット方式による選択、交叉率100%の一様交叉、突然変異率u、集団サイズが十分大きい場合のモデルについては以下のように書ける。

t: 世代（各操作毎に+1/3, 1世代で+1）

x_i : 遺伝子。0または1 ($i = 1, 2, \dots, n$)

$P_{x_1 x_2 \dots x_n}^t$: 個体が $x_1 x_2 \dots x_n$ である確率

$f_{x_1 x_2 \dots x_n}$: $x_1 x_2 \dots x_n$ の評価値

選択: (ルーレット方式)

$$P_{x_1 \dots x_n}^{t+1/3} = P_{x_1 \dots x_n}^t f_{x_1 \dots x_n} / \sum_{y_1=0}^1 \dots \sum_{y_n=0}^1 P_{y_1 \dots y_n}^t f_{y_1 \dots y_n}$$

交叉: (一様交叉)

$$P_{x_1 \dots x_n}^{t+2/3} = \frac{1}{2^n} \sum_{i_1=0}^1 \dots \sum_{i_n=0}^1 \sum_{y_1(1-i_1)=x_1} \dots \sum_{y_n(1-i_n)=x_n} P_{y_1 \dots y_n}^{t+1/3} P_{y_1 \dots y_n}^{t+1/3}$$

上の和は交叉後に $x_1 x_2 \dots x_n$ になる場合のすべてについての和である。

突然変異: (ランダム)

$$P_{x_1 \dots x_n}^{t+1} = \sum_{y_1=0}^1 \dots \sum_{y_n=0}^1 u^d (1-u)^{n-d} P_{y_1 \dots y_n}^{t+2/3}$$

ただしdは突然変異の生じる遺伝子の数で

$$d = \sum_{i=1}^n |x_i - y_i| \text{である。}$$

我々はこの式より定常分布 ($P^{t+1} = P^t$) を直接計算で求めることを提案した。[3]

ここでは $P^{t+1} = G(P^t)$ と $P^{t+1} = P^t$ から P^t の推移について考える。[1] 両者の交点は定常分布である。図1は $P^{t+1} = G(P^t)$ と $P^{t+1} = P^t$ の関係を2次元で表した概念図である。

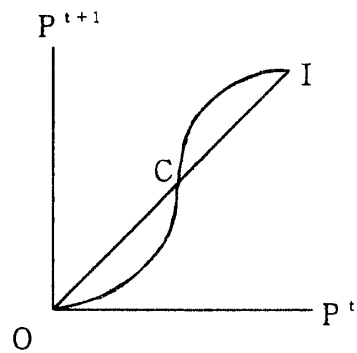


図1. 集団の状態推移関数G

これにより、期待値分布の推移の方向が解る。図1の場合では点Cが不安定平衡点であり、それより大きい場合はIに小さい場合はOに向かって収束する。(安定平衡点)

3. 議論

従来GAはグローバルな探索をしているという見方があるが、図1の立場から見るとローカルな山登りをしていると考えることもできる。つまりGAは超空間上の評価値の山(fitness landscape)を山登りするのではなく、関数 $G(P^t) - P^t$ を与えるポテンシャルの山登りをしていると見なせる。言い換えれば、現在の集団の状態を頼りにその状態空間を近傍探索をしていることになる。

初期集団分布依存性について

以上の議論からGAに見られる初期集団分布依存性の様相を考えることができる。初期分布が不安定平衡解を境にした領域のどちらに入るかによって収

束する場所が異なる。例えば、表1に示した評価関数では、わずかな初期分布の違いによって収束先が異なる。(ここでは遺伝子型11は00と同じ評価値としたので省略。突然変異率は0とする。)このために、GAを実行する際、大集団で1回行うよりも、小集団で数回行ったほうが良いという経験則が見られたのである。また表2の例では、初期分布の偏りのため評価値が等しい2つの山の一方に分布が集まる様子を表わしている。これは、集団遺伝学でいう遺伝的浮動(選択の偶然性等の理由で遺伝子頻度が揺らぐ現象)により適応度に差がない一方の遺伝子が消滅する現象[2]に対応している。

表1. 初期分布依存性1

遺伝子型	00	01	10
評価値 f	2	21	20
初期分布1	0.25	0.224344	0.275656
t=60分布1	0.000044	0.999878	0.000035
初期分布2	0.25	0.224343	0.275657
t=60分布2	0.000115	0.000178	0.999384

表2. 初期分布依存性2

遺伝子型	00	01	10
評価値 f	1	10	10
初期分布3	0.25	0.26	0.24
t=30分布3	0.000001	0.999997	0.000001

表3. 集団サイズの違い

遺伝子型	00	01	10	11
評価値 f	10	1	1	10
初期集団1	0	1	1	0
初期集団2	0	2	2	0

表4. 分散の違い

t=1	期待値分布	ハミング距離の分散	ユークリッド距離の分散
初期集団1	1:3:3:1	1.41	0.94
初期集団2	1:3:3:1	0.70	0.78

集団サイズについて

以上の議論はあくまでも期待値の推移についてである。実際のGAでは期待値からずれる訳であるが、そのずれは集団サイズが小さいほど大きく現れやすい。(すなわち、分散が大きい。)この場合、不安定平衡解の境を越え別の領域に移る場合もありうる。集団サイズと分散の例を表3、表4に示した。こう考

えると、集団のサイズが小さい方が小領域から脱出しやすいといえる。

突然変異について

突然変異率が小さい場合、関数Gを描く際、とりあえず選択と交叉だけで描き、突然変異は(集団サイズの影響と同様に)期待値の僅かなずれとみなしたほうが実用的である。(ただし突然変異は分布を均す方向に働く確率が高い点には注意。)この場合 $P_i = 1, P_{j \neq i} = 0$ が安定平衡点の自明解となるメリットがある。

4. おわりに

本論文では、集団分布の期待値推移関数について形状から終状態を予測するという手法を用いて、GA探索の初期集団依存性、すなわちGAがある意味でローカルな探索であるという見方を示した。また他のタイプのGAについても定式化が可能である。理論上の困難としては、突然変異率が大きい場合は収束性そのものを保証できないことであるが、その場合は探索としての意味が薄れるので、実用上には問題は無いと考えている。また何故GAがNP完全問題を(遺伝子長に対して多項式倍程度の)実用的な時間で解くことができるのかという点についても、このモデルは示唆を与えている。GAでは遺伝子長nを1増やす毎に、探索空間が2倍になるのに対して、不安定平衡解によって分割された領域の数も2倍以上になる。ただし、各領域の体積は一定ではない。例えば、 $f_i = 10, f_{j \neq i} = 1$ のような場合、 $P_i = 1$ に収束する領域は $P_{j \neq i} = 1$ などを除いてほぼ全領域である。このため、探索にはやはり指数倍の時間がかかるのである。現在この点について詳しく検討している。

参考文献

- [1]甘利俊一(1978)神経回路網の数理. 産業図書
- [2]丹羽、田中(1994)遺伝的アルゴリズムにおける遺伝的浮動に関する一考察. 第49回情処全国大会
- [3]吉澤、橋本(1996)GAにおける個体分布の推移と定常集団分布について. 1996年電子情報通信学会総合大会