

スキーマ定理に関する一考察

6N-3

相澤 彰子

(学術情報センター)

1 はじめに

遺伝的アルゴリズム (Genetic Algorithms, 以下 GA) は、スキーマと呼ばれる部分解の高並列な処理により解空間の効率的な探索を目指すものである。

GA のスキーマ処理を数式的に表現したものが, [Holland 75] によるスキーマ定理である。スキーマ定理は多くの GA の理論的研究において参照されており, その解釈についても, だまし問題との関係を指摘した [Grefenstette 93] [Whitley 91] など多様な見方が存在する。

本稿では, スキーマ定理は GA における (1) 有望なスキーマの追求 (exploitation) と (2) 新たなスキーマの探索 (exploration) のトレードオフ関係を記述した定理であると考え, 従来より GA においては, 後者を適応的に行う能力が重視され, 前者に関して議論されることは少なかった。本稿では前者に注目し, GA における既存スキーマの追求は解空間上での山登り法とは異なるという視点から考察を行う。

2 スキーマ処理における貪欲法

2.1 山登り法と局所最適解

山登り法は, 解空間上の近傍点の中から解を改善するものを選び, その方向に探索を進める広い意味で貪欲的な方法である。山登り法による探索は, 限定された解空間中での局所的な最適解を目指す。これに対して, 探索範囲を広げるため確率的な要素を導入したものが, 焼きなまし法などの確率的な山登り法である。GA においても, 操作子として突然変異のみを用いる場合には確率的な山登り法の一つであるとみなされる。

2.2 スキーマ貪欲法と誘引スキーマ

一方, スキーマ処理に基づく探索は解空間上の近傍点に限定されておらず, 上記とは異なる定義が必要である。ここではスキーマ処理において, 解空間上での山登り法に対応する概念をスキーマ貪欲法と呼ぶ。

スキーマ処理の数学的な説明では, スキーマを正規分布でモデル化し, 各々異なる平均および分散を持つとするのが一般的である。この場合の正規分布のサンプルは, そのスキーマを含む個体の評価値であり, スキーマ定理は, 真に平均値が高いスキーマを漸近的に最適に識別するサンプル配分を定めている。

正規分布モデルに基づき本稿では, スキーマ貪欲を, 既存のスキーマの中で, サンプル平均 (真の平均値ではないことに注意) のよいスキーマをさらにサンプルすること, すなわち観測されたよい個体に含まれるスキーマを選び, さらに新しい個体を生成することであると定義する。

さて, 山登り法における局所最適解に対応して, スキーマ貪欲法を導くスキーマを誘引スキーマ (attractive schemata) と呼ぶことにする。スキーマ処理は観測されたサンプル平均に基づくため, 誘引的であるかどうかはスキーマの真の平均値だけでは決まらない。また, スキーマ処理の複雑さから, 多くのスキーマの中でどのスキーマが誘引的であることを正確に知る現実的な方法は存在しないと予想される。正規分布モデル上で行ったシミュレーションによると, (1) 平均値が大きく, かつ (2) 分散が小さいスキーマが多くの場合誘引的であることが観察された。

誘引スキーマを敢えて視覚的に表現したものが図 1 である。図中, なだらかな山 [a] では解の分散が小さく, 急峻な山 [b] では分散が大きい。このような場合にはスキーマ貪欲法は山 [a] を登りやすいといえる。これは従来のだまし問題における観察と一致している。

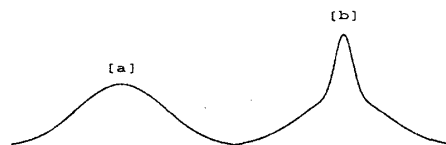


図 1: 誘引スキーマの直観的な図式化

2.3 GA 容易な問題とだまし問題

GA 容易 (GA-easy) な問題では, 誘引スキーマは最適解と多くのビットを共有すると考えられる。一方, いわゆる GA のだまし問題では, 誘引スキーマは最適解との共有ビットが少なく, 結果として探索の方向を最適解とは逆の方向に惑わせると考えられる。

ここで, GA 容易な問題の例として '1' 数え上げ問題を取り上げる。ストリング長 L として, スキーマ s の次数を $o(s)$, s の中に含まれる '1' の数を $count(s)$ で表す。ストリング x に対する評価関数 $f(x) = count(x)$ とするとき, スキーマ s の平均値 $(count(s) + \frac{L-o(s)}{2})$, 分散 $(\frac{L-o(s)}{4})$ である。すなわち, 次数が大きく '1' の数が多いスキーマほど誘引的であり, スキーマ貪欲法にとり好ましい問題であるということが出来る。

3 スキーマ貪欲な確率的探索法の構成

3.1 構成する探索法の概要

実際に、よいスキーマに対して貪欲な確率的探索法の構成を試みる。このためにまず、手順(1)として、個体集団の優れたものの中から共通ビットを取り出し有用なスキーマを抽出する。次に手順(2)として、得られた有用なスキーマに基づき確率的方法で新たな個体を生成する。以下それぞれについて簡単に述べる。

3.2 手順(1)：有用なスキーマの抽出手順

時点 t における大きさ M の個体集団を P_t とするとき、前述のようにスキーマ s は P_t 中で s を含む個体部分集合の平均値で評価される。

これより、有用なスキーマの抽出ではまず、 P_t 上で可能なすべての個体部分集合を考え、これを平均評価値のよいものから順位づけする。このために、 P_t に含まれる個体を評価値のよいもの順に並べ、 $c_1, \dots, c_M$ のようにインデックスをつける。そして P_t の部分集合間の半順序関係を利用して、部分集合 $\{c_1\}$ から出発して図2の矢印の順序にしたがい候補集合を生成、順位づけリストにソート順に挿入すると、時間オーダー $M \log(M)$ で上位から M 個の部分集合が取り出せる。

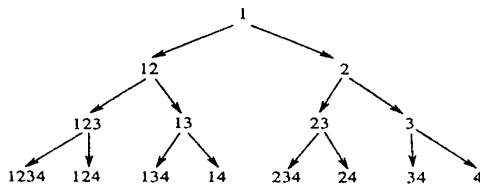


図2: インデックス集合の部分順序関係

次に、この順位づけした M 個の部分集合について、要素間で共有する最大のスキーマを取り出す(図3)。この手順ではスキーマの選択はGAのように確率的ではなく、上位から一意に定まるとい意味で貪欲的である。

population P_t	ordered schemata												
c_1 <table><tr><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr></table>	0	1	0	0	1	1	$s(\{c_1\})$ <table><tr><td>0</td><td>1</td><td>0</td><td>0</td><td>1</td><td>1</td></tr></table>	0	1	0	0	1	1
0	1	0	0	1	1								
0	1	0	0	1	1								
c_2 <table><tr><td>0</td><td>0</td><td>1</td><td>0</td><td>1</td><td>1</td></tr></table>	0	0	1	0	1	1	$s(\{c_1, c_2\})$ <table><tr><td>0</td><td>*</td><td>*</td><td>0</td><td>1</td><td>1</td></tr></table>	0	*	*	0	1	1
0	0	1	0	1	1								
0	*	*	0	1	1								
c_3 <table><tr><td>1</td><td>1</td><td>0</td><td>0</td><td>0</td><td>1</td></tr></table>	1	1	0	0	0	1	$\vdots$						
1	1	0	0	0	1								
$\vdots$	$s(\{c_1, c_2, c_3\})$ <table><tr><td>*</td><td>*</td><td>*</td><td>0</td><td>*</td><td>1</td></tr></table>	*	*	*	0	*	1						
*	*	*	0	*	1								
	$\vdots$												

図3: 部分個体集合とスキーマの対応

3.3 手順(2)：新しい個体の生成手順

上記により順位づけしたスキーマを基に次世代の個体集団 P_{t+1} を生成する。

まず、上位 M 個の各スキーマの '*' ビット部にランダムに '0' または '1' を割り当て、 M 個の個体を生成する。ここで、新たに生成した個体は、対応する個体部分集合の共有スキーマを含むことから、この操作は、

親の数を2個から任意個に拡張した均一交叉 (uniform crossover) となっている。さらに、集団中に新しいスキーマを導入するため、突然変異を用いてランダムにビット反転をする。このとき突然変異率はアルゴリズムの制御パラメタとする。

4 実験による動作比較例

4.1 実験の条件

実験では、(1) 世代ごとに個体を入れ換える単純GA、(2) 操作子として突然変異のみを用いる確率的山登り法 (Stochastic Hill Climber)、および本稿で述べた(3) 確率的スキーマ貪欲法 (Stochastic Schemata Exploiter) の3者について動作を比較した。設定したパラメタの値は、個体集団の大きさ 100、突然変異率 0.001、また単純GAについて交叉率 0.6 である。

4.2 GA 容易問題に適用した結果

ストリング長 160 ビットの '1' 数え上げ問題に適用した結果を図4に示す。確率的スキーマ貪欲法は、単純GAや確率的山登り法と比較して解の収束が速いことがわかる。また、DeJongのテスト問題やナップザック問題等を用いた動作比較においても同様の結果が得られている。

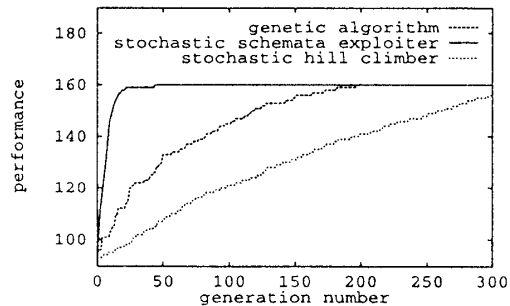


図4: '1'数え上げ問題による動作比較

5 おわりに

本稿では、スキーマ処理に基づく探索法におけるスキーマ貪欲性について考察を行い、確率的スキーマ貪欲法の構成を試みた。実験によると確率的スキーマ貪欲法は、はやい収束特性を示しており、また制御パラメタの数も少ないため、GAの特性を考察する上で有効であると考えられる。スキーマ定理が示唆するように、適応度に応じたスキーマの確率的な選択が有効かなどの検討が今後の課題である。

参考文献

- [Holland 75] Holland, J.H.: "Adaptation in Natural and Artificial Systems," the University of Michigan Press (1975).
- [Grefenstette 93] Grefenstette, J.J.: "Deception Considered Harmful," *Foundations of Genetic Algorithms 2*, Morgan Kaufman Publisher (1993).
- [Whitley 91] Whitley, L.D.: "Fundamental Principles of Deception in Genetic Search," *Foundations of Genetic Algorithms*, Morgan Kaufman Publisher (1991).