

解の類似度を考慮した推定による共進化型アルゴリズム

Co-Evolution Algorithm with Estimation considered Similarity of Solutions

田村謙次

Kenji Tamura

中央学院大学

Chuo-Gakuin University

This paper proposes an Estimation of Distribution Algorithm with partial solution considered similarity of solutions. In the proposed methods, partial solutions are generated by estimated value and search local solution by co-evolution. However, these methods have a issue of searching cost by a character of optimization problem's solution that there are some optimum solutions. So it proposes an expanding method newly that searching with partial solution which have an orientation. The proposed method is applied to some optimization problems and discuss its effectiveness.

1. はじめに

近年, GA(Genetic Algorithm)などに代表される進化的計算において, 交叉や突然変異などの遺伝的操作を用いるのではなく, 遺伝子座における対立遺伝子の確率分布に着目した分布推定アルゴリズム (Estimation of Distribution Algorithm; EDA)[棟朝 08] が盛んに研究されている. 進化的計算では, 有効な部分解であるビルディングブロックが積み重ねられることにより解発見することを仮定しているが, 遺伝的操作により, その並びが破壊されるという欠点があった. EDA は, 遺伝子座における対立遺伝子の分布の確率モデルを構築し, そのモデルを用いて各遺伝子座において有効と判定される遺伝子を確率的に探索する手法であり, 局所的な探索を行うなどのさまざまな拡張手法が提案されている. また, EDA には, 遺伝子座を独立に探索する手法や BOA(Bayesian Optimization Algorithm) などのように遺伝子座間の関係を明示的に用いた手法があるが, 明示的な手法では実装の負担や解探索における計算量などに課題がある [Chickering 95]. したがって, 本論文では, 比較的実装が容易である一変量周辺分布アルゴリズム (Univariate Marginal Distribution Algorithm; UMDA) を用いる. これまでは, 解を持つ個体と, 部分解を持つ個体を導入することにより局所探索を行う共進化アルゴリズムを提案してきたが, 最適解が複数存在するような問題においては, 提案手法では解発見までに計算コストが掛かるという課題があった. そこで本論文では, 推定に用いる解集団の類似度を考慮した手法を提案し, 各適用問題における有効性の確認を行う.

2. 関連研究

2.1 ウイルス進化型遺伝的アルゴリズム (VE-GA)

VE-GA(Virus Evolutionary Genetic Algorithm)[久保田 98] では宿主個体群とウイルス個体群という2種類の個体群の共進化により解の探索を行う. 宿主は適用する問題における解候補を遺伝子列に持ち, ウイルスは部分解を遺伝子列としている. 部分解は対象とする問題の解の一部しか表現していないので適用問題の解としては成立しない. ウイルスは感染操作により

自身の遺伝子列を宿主にコピーを行い, 取込操作により宿主の遺伝子列の一部を自身の遺伝子列とする. これらの操作により VE-GA では部分解の局所探索を行い, 発見されたスキーマを感染により宿主個体群に高速に伝播させる. 一方, 宿主同士は交叉により大域的な探索を行う. 宿主はより良い解候補の探索を行い, ウイルスは宿主の適応度を向上させる部分解の探索を行う. また, ウイルスは得られた適応度から算出される生命力を持つ. ウイルスの適応度は, 感染することにより生じた宿主の適応度変化量の和であり, 宿主の適応度を改善すれば大きな値を持つこととなる. また, 生命力はウイルスの適応度に減衰率を乗じた値と, 前世代の生命力との和で求められ, 生命力が0になるとウイルスは, 取込を行うことにより, 新しい遺伝子列を持つ個体となる.

3. 共進化型分布推定アルゴリズム

従来の VE-GA ではウイルスは遺伝的操作により宿主から取り込んだ遺伝子列を保持し, それを宿主間に感染させることで共進化を行う. しかし, EDA の特質の一つは遺伝的操作を用いずに解探索を行うことであるので, 遺伝子を直接変化させる従来の感染操作による解探索は, EDA の特性と異なる立場となってしまう. そこで, ウイルスは遺伝子列を持つのではなく, 遺伝子列から推定された推定値を持つことにより, 局所的な探索を行う個体とし, ウイルスが保持する推定値において2つの取込手法(手法1, 2)を提案してきた [田村 10]. 手法1 (図1) では, 宿主は遺伝子列とその遺伝子列の推定に用いた推定値を持ち, ウイルスはその推定値を取り込み, 感染はそれから推定された遺伝子列とその推定値を, ランダムに選択した宿主に伝播させる操作とし, 手法2 (図2) は, 宿主は遺伝子列のみを持ち, ウイルスはランダムに選択した複数の宿主の遺伝子列から推定された推定値を取り込み, 感染はその推定値を用いて部分的な遺伝子列の推定を行い, ランダムに選択した宿主の遺伝子列のみを置き換える手法である. 両提案手法により部分解の探索を行うこととなるが, 手法2では, 複数の宿主を用いて部分的な領域における解の推定を行うことで, EDA という手法における確率的な推定による解探索の設計思想を部分解にも適用し, 手法1のように単一宿主の部分解による定方向の探索を行うのではなく, 大域的な解空間において有効であると推定される解を用いて, 部分解という限定された探索領域においても, 探索方向を限定することなく, より多様な解の局所探索の

連絡先: 田村謙次, 中央学院大学,

〒270-1196 千葉県我孫子市久寺家 451 番地,

04-7182-1441, ktamura@slb.cgu.ac.jp

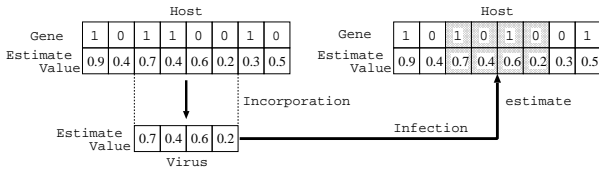


図 1: Infection and Incorporation of method1

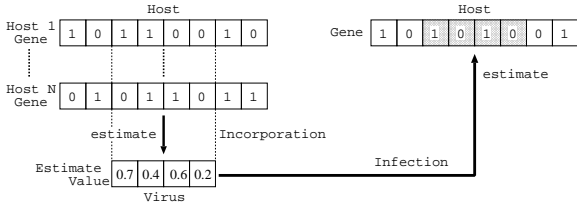


図 2: Infection and Incorporation of method2

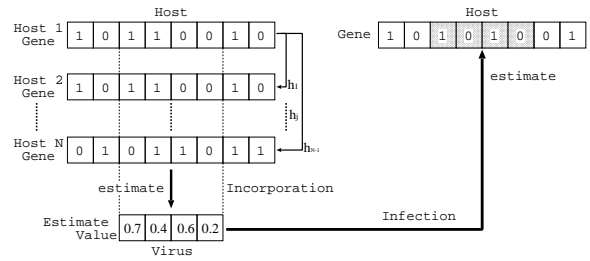


図 3: Infection and Incorporation of method3

実現が可能であった．しかし、最適解が複数存在する SubSet Sum のような問題においては、提案手法では解発見までに計算コストが掛かるという課題があった．

アルゴリズムの概要は以下のものであり、次章で述べる提案手法のアルゴリズムも同様である．

```

Begin
Generate Host and Virus population
Repeat
Select host individuals
Estimate and Generate offspring
Infection
Calculate
Host fitness, Virus fitness, Virus life
Replacement
Incorporation
Until Termination-condition = True
End.
    
```

4. 解類似度の考慮

複数の最適解の存在が比較的容易に想定できる SubSet Sum ような問題においては、全く異なる遺伝子列を持つ個体同士が、共に最適解である可能性があるため、取込に複数の解候補を用いると探索の方向性が定まらず、これまで提案してきた手法 2 では、探索コストが増加する課題があった．また、工学における多くの組合せ最適化問題において、良い解同士は似通った構造を持っているという、近接最適性原理 (Proximate Optimality Principle ; POP) が成立することが経験的に知られているが、偶奇性判定問題や数分割問題のように POP が成立しない問題が存在することが知られており、進化的計算のようなメタヒューリスティクスがうまく機能出来ないなどの課題が指摘されている [室田 95]．そのような POP における課題の 1 つの要因としては、SubSet Sum のように求めようとしている最適の解に対して、それに該当する解が複数ある場合などが考えられる．そこで本手法ではこれまでの提案してきた手法にウイルス取込の際に選択された宿主間のハミング距離を推定値に反映させることで、部分解の探索する解領域に対して方向性を持たせる手法を提案する．

4.1 ハミング距離を用いたウイルス取込手法

本手法では、ウイルス取込の際に選択された宿主の中から、ランダムに一つの個体を選択し、その個体を基準として、その他の宿主とのハミング距離を計算する．そのハミング距離が基準となる宿主との遺伝子型における距離であり、適応度が高くても、探索方向が異なる個体であると考えられる．したがって、ウイルスの遺伝子列となる推定値を求める際には、距離が離れている個体からの影響を小さくすることにより、解探索において方向性を持つ部分解となることが期待される (図 3)．

取込の際に、宿主が選択されるまでは、これまでの手法と同様である．まず、基準とする個体としてランダムに選択された宿主とその他の宿主とのハミング距離を用いて、解の探索方向に対する影響度を求める (式 (1))．

$$r_i = 1 - \frac{(1 - \alpha)h_i}{L} \quad (1)$$

r_i は個体 i による解への影響度、 h_i は基準とする宿主と個体 i とのハミング距離、 L は遺伝子列全体の長さ、 α は影響度の基準点を表している．遺伝子列の値は、0 または 1 としているので、本研究では α を中間値である 0.5 とする．

次に、影響度を考慮した推定値 p_j を求める (式 (2))．

$$p_j = \frac{\sum_{i=0}^N r_i \cdot g_{ij}}{N} \quad (2)$$

g_{ij} は個体 i における j 番目の遺伝子の値、 $\sum_{i=0}^N r_i \cdot g_{ij}$ は遺伝子座 j における推定値の和、 N は選択された宿主個体数を表し、求められた p_j の値の並びがウイルスの遺伝子列となる．本論文では、本提案手法を手法 3 とする．

5. 最適化問題を用いた実験

実験は、OneMax, Isopeak, Plateau, SubSet Sum による比較実験を行った．各問題の規模は SubSet Sum 以外では 24, 50, 100, SubSet Sum では 50, 100, 150, 200 とし、各規模において 60 回の試行を行い、解が発見されるまでの世代数を比較した．各パラメータは、宿主個体数 100, ウイルス個体数 5, 感染数 5, 世代ギャップ 5 とした．感染数とは 1 つのウイルスが感染を行う宿主の数を表し、世代ギャップとは感染が実行される世代間隔を表している．また、手法 1, 2 および提案手法 (手法 3) での感染や取込における推定コストを考慮して世代数の評価を行うものとし、取込の際に選択する宿主数は試行錯誤的に決定した．

各問題の特徴としては、OneMax は式 (3) のように定式化され、線形的で遺伝子座間に相互作用はなく、山登り的な探索で最適解を求めることが出来る．

$$f_{OneMax}(\vec{x}) = \sum_{i=1}^n x_i \quad (3)$$

Isopeak は、下表と式 (4) のように定式化され、遺伝子座間に相互作用が存在する問題である。

\vec{x}	00	01	10	11
f_{Iso1}	m	0	0	$m-1$
f_{Iso2}	0	0	0	m

$$f_{IsoPeak}(\vec{x}) = f_{Iso2}(x_1, x_2) + \sum_{i=2}^m f_{Iso1}(x_{2i-1}, x_{2i}) \quad (4)$$

Plateau は、下記の式 (5), (6) のように定式化され、遺伝子座間に相互作用が存在する問題である。

$$f_{Plateau}(\vec{x}) = \sum_{i=1}^m g(\vec{s}_i) \quad (5)$$

$$g(x_1, x_2, x_3) = \begin{cases} 1 & \text{if } x_1 = 1 \ \& \ x_2 = 1 \ \& \ x_3 = 1 \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

SubSet Sum は、設定された和に対して、その和を構成するための数の部分集合を求める問題であり、求める数について複数の解が存在することがあるため、大域的な探索を行うことによる各最適解への探索を行うと、計算コストが多くなってしまふ場合がある。

Table.1 は、各手法における実験結果を示している。共進化による手法については、既存の手法よりも良好な結果となり、部分解を用いることで局所的な探索を行うことができたと考えられるが、問題規模が小さいものについては、既存の手法の方が良い結果となるものもあった。共進化による手法を比較すると、手法 1 よりも手法 2, 3 の方が良好な結果となった。これは、手法 1 ~ 3 は取り込んだ部分解に適応度を改悪する遺伝子を含んでいても、改良する部分の割合が大きければ、解の伝播が行われてしまうが、手法 2, 3 では、複数の宿主から推定される解を持つので、そのようなことが生じる確率が低く、手法 1 よりも良好な結果となったと考えられる。また、手法 2 と 3 を比較すると、OneMax と Plateau では手法 2, Isopeak と SubSet Sum では手法 3 が概ね良好な結果となった。

各問題において考察すると、遺伝子座間の相互作用がない OneMax や、遺伝子座間に弱い相互作用がある Plateau においては、手法 2 が良好な結果となったが、これは両問題とも山登りの探索で解が発見できるため、手法 3 は、解空間における探索の明確な方向性を決定するまでに、手法 2 よりも多くのコストが掛かっていると考えられる。遺伝子座間に相互作用がある Isopeak や、複数の最適解が存在する SubSet Sum においては、手法 3 が良好な結果となった。Isopeak と SubSet Sum は準最適解または最適解が複数あり、山登りの探索では素早く解を発見することができないため、手法 2 のように複数の方向に対して均等に影響を受ける手法では、解が明確な方向性を持たないため、一定方向の進化を行われるまでに探索コストを要してしまうと考えられる。一方、手法 3 においても、複数の方向性に対して探索を行うが、各部分解においては手法 2 に比べて明確な方向への解探索を行うため、良好な結果になったと考えられる。また、SubSet Sum においては、手法 2 よりも手法 1 の方が良好な結果となったが、これは手法 3 と同様に、手法 1 は、1 つの宿主からの部分解による取込を行うため、手法 2 と比較して明確な方向性を持つ特徴があるためだと考えられる。

表 1: Experimental Results

Problem	Size	Method			
		UMDA	1	2	3
OneMax	24	14.8	17.2	17.3	18.3
	50	88.5	96.8	88.3	115.2
	100	3611.0	2110.6	755.7	1050.4
Isopeak	24	264.6	221.4	198.4	231.8
	50	4564	1104	1037	965.0
	100	402692	71299	34228	8490.2
Plateau	24	21.75	21.17	21.9	23.2
	50	141.53	122.92	114.8	119.9
	100	4614.4	2582.2	842.3	951.3
SubSet Sum	50	15.2	15.9	16.8	11.1
	100	33.3	43.5	39.8	29.3
	150	67.4	60.5	61.5	50.2
	200	135.9	105.5	109.9	101.6

6. おわりに

本稿では、解の類似度を考慮する取込手法を導入した共進化型分布推定アルゴリズムを提案し、従来手法との比較において適用問題の性質の違いにより良好な結果が得られたことを報告した。提案したハミング距離を利用することによる解の類似度を考慮した手法においては、Isopeak のような遺伝子座間に相互作用がある問題、および、SubSet Sum のように最適解が複数存在する問題では、これまでの手法よりも良好な結果となったが、OneMax や Plateau のように、山登りの探索が効果的な問題においては、ハミング距離による解同士の距離を考慮する処理を施したことにより、改悪する結果となった。

今後の課題としては、解の遺伝子座間における依存関係の強弱の違いがある問題への適用や、他の EDA 手法との比較や提案手法の遺伝子座間に相互作用がある手法への適用、および、有効に機能する問題の性質や大規模な問題における有効性についての検証などが考えられる。

参考文献

- [棟朝 08] 棟朝雅晴: 遺伝的アルゴリズム, 森北出版 (2008).
- [室田 95] 室田一雄 他: 離散構造とアルゴリズム IV, 近代科学社 (1995).
- [Chickering 95] David M. Chickering, Dan Geiger, David Heckerman: Learning Bayesian Networks: Search Methods and Experimental Results, In Preliminary Papers of the Fifth International Workshop on Artificial Intelligence and Statistics, pp.112-128(1995)
- [久保田 98] 久保田直行・福田敏男: ウイルス進化とファジィ理論, 日本ファジィ学会誌, Vol.10, No.4, pp.626-636 (1998).
- [田村 10] 田村謙次・鳥居隆司: 推定による部分解を用いた共進化型分布推定アルゴリズム, 電気学会論文誌 C, Vol.130, No.1 (2010).