

遺伝的アルゴリズムの基礎と応用〔I〕

小林 重信

1. はじめに

1960年代前半に、分岐限定法 (branch and bound method) の枠組みが提案され、現在、組合せ最適化問題に対するオペレーションズ・リサーチ的の接近の代表的かつ汎用的な方法として広く利用されている。分岐限定法の有効性は領域知識を利用した枝刈り規則 (pruning rule) の質にいちじるしく依存する。たとえば、平面 TSP については、強力な枝刈り規則が知られているが、ジョブショップスケジューリング問題では、良質の枝刈り規則は見いだされておらず、組合せの爆発 (combinatorial explosion) の問題をかかえている。

1970年代後半に、エキスパートシステム (expert system) の枠組みが提案され、現在、組合せ最適化問題に対する人工知能的接近の代表的かつ汎用的な方法として広く利用されている。エキスパートシステムの有効性は領域専門家から獲得される経験的知識 (expertise) の質と量にいちじるしく依存する。たとえば、航空機や列車のダイヤ編成を対象としたエキスパートシステムは実用化されているが、ジョブショップスケジューリング問題では、スケジューリングのための経験則の蓄積が不十分で、知識獲得 (knowledge acquisition) の問題をかかえている。

このように、分岐限定法とエキスパートシステムは、領域知識が利用可能であることを長所として、見方を変えれば、領域知識に依存せざるをえないことを短所として共有している。

1980年代に入り、自然界に存在する物理学的あるいは生物学的な現象を利用して最適化問題を解こうとする研究が盛んになってきた。Hopfield型ニューラルネット[1]、シミュレーテッドアニーリング[2]などがそうであり、本連載でとり上げる遺伝的アルゴリズムもそのような流れの1つに位置づけられる。

こばやし しげのぶ 東京工業大学 総合理工 知能科学専攻

〒227 横浜市緑区長津田4259

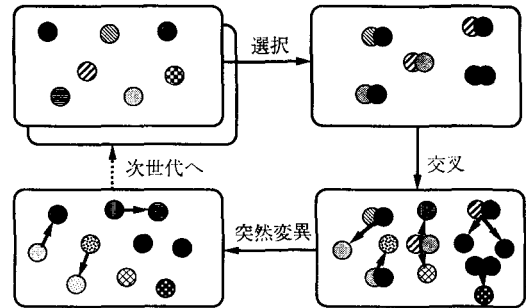


図1 遺伝的アルゴリズムの概念図

本連載は、「遺伝的アルゴリズムの基礎と応用」と題し、3回にわたり、GAの基本的な概念、GAの挙動に関する理論およびGAの応用について、できるだけ平易に解説することを通じて、本学会会員のGAに対する理解と興味を深めることを目的としている。

2. 遺伝的アルゴリズムの基本的概念

遺伝的アルゴリズム (Genetic Algorithm : GA) は、ダーウィンの自然淘汰理論 (Natural Selection Theory) に基礎をおく生物の進化過程を模倣した工学的モデルであり、人工生命 (Artificial Life) の基盤技術として使われているが、探索手法として頑健な枠組みをもち、しかも並列処理との親和性が高いことから、最近、組合せ最適化問題に対する新しいパラダイムとして、各界から大きな関心と期待が寄せられている。

GAの枠組みの提案はHollandによる[3]。図1はGAの挙動を概念的に示したもので、選択・交叉・突然変異と呼ばれる3つの遺伝的オペレータを順に適用することを繰り返して、環境にもっとも適合する個体群を見出すように、GAがふるまうことを示している。

2.1 GAのデータ構造

はじめに、GAで使われる基本的な概念をいくつか紹介する。最小構成要素を遺伝子 (gene)、遺伝子がとりうる値を対立遺伝子 (allele) という。遺伝子の並び文字列 (string) を染色体 (chromosome)、染色体上の遺伝子の位置を遺伝子座 (locus)、染色体の長さを染色

体長 (chromosome length) という。

なお、染色体の表現は、文字列に限定される必然性はない。染色体として、木構造 (tree structure)、グラフ (graph)、リスト構造 (list structure)、行列 (matrices) など、構造をもつ表現も許される。GAの特徴の1つは構造を、直接、遺伝的操作の対象としうることにある。

染色体によって特徴づけられる個を個体 (individual) 個体の集まりを集団 (population)、集団の大きさを集団サイズ (population size) という。

最適化問題をGA上にモデル化する場合、一般に、1つの解候補が1つの個体に対応づけられる。各個体を正の実数に写像する関数を適応度関数 (fitness function) という。最適化問題では、目的関数と制約条件を考慮したペナルティ関数を適応度関数に対応づける必要がある。

染色体によって規定される形質の外部表現形式を表現型 (phenotype)、染色体による内部表現形式を遺伝子型 (genotype) という。遺伝子型に対応する表現型が元問題の制約を充足しないとき、その染色体は致命的 (fatal) であるという。

2.2 Simple GA

GAの基本形の1つとして、Simple GAがある[4]。Simple GAは以下の手順からなる確率的過程である。[Simple GA]

(1) 初期化 (initialization)

各遺伝子座の遺伝子に対し、一様分布に従って、0または1の対立遺伝子を割り当てることにより、ランダムな染色体をもつ個体を集団サイズの数だけ生成し、初期集団とする。

(2) 選択 (selection)

ルーレット選択によって、集団中から個体を復元抽出 (sampling with replacement) することにより、集団を再編成 (reproduction) する。再編成された集団から個体を非復元抽出して、個体ペアをランダムに生成する (random pairing)。

(3) 交叉 (crossover)

(2)でつくられた各個体ペアに対して、一定の交叉率 (crossover rate) に従って、交叉させるか否かを決定する。交叉させる場合、1点交叉により、子の個体ペアを生成し、親の個体ペアと交換する。交叉させない場合、

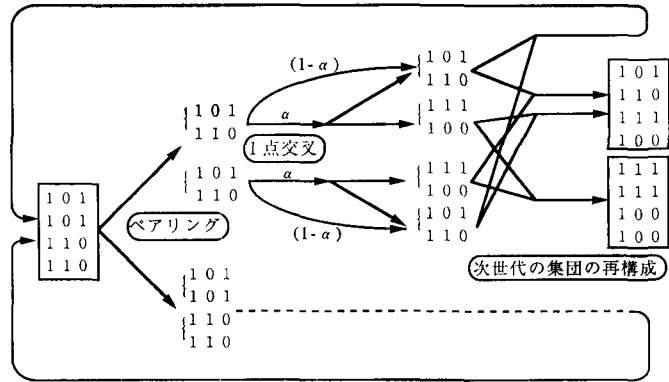


図2 Simple GA おける1点交叉

親の個体ペアはそのまま残す。

(4) 突然変異 (mutation)

各個体について、一定の突然変異率 (mutation rate) に従って、突然変異させるか否かを決定する。突然変異させる場合、遺伝子座をランダムに1つ指定し、その値を他の対立遺伝子に置き換える。

(2)~(4)の遺伝的操作 (genetic operations) の1サイクルを1世代 (one generation) という。

simple GAにおける交叉の様子を図2に示す。この例では、2種4個の個体からなる集団を扱っており、2通りのペアリングが可能で、異なる染色体をもつ個体がペアとなったとき、交叉率 α で交叉が決定され、交叉箇所が2カ所あることから、2通りの子のペアが生成可能であることを示している。

Simple GAの実装にあたっては、通常、集団サイズを数十から数百、交叉率を0.6位、突然変異率を0.01位に設定するのが、経験的に得られたおおよその目安とされている。Simple GAは、単純で小規模な問題に適用する限り、問題はないが、対象問題が少し複雑・大規模になると、最適解に到達する前に、すべての個体が同一の染色体をもつか、またはある部分空間に集団全体が陥る初期収束 (early convergence) をまねく危険性を内包している。

3. GAの構成要素および設計問題

GAがもつ潜在的な能力をフルに引き出すためには、遺伝的オペレータの役割を十分認識した上で、対象問題に対して、オペレータの設計やパラメータの調整を適切に行なうことが必要である。

3.1 コード化/交叉問題

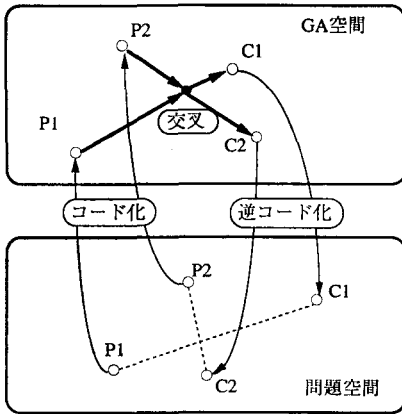


図 3 問題空間と GA 空間の対応

表現型 (phenotype) によって規定される解候補の集まりからなる空間を問題空間, 遺伝子型 (genotype) によって規定される空間を GA 空間と呼ぶことにする. 問題空間と GA 空間の対応関係をコード化 (encoding/decoding) と呼ぶ. コード化問題とは, 問題空間から GA 空間への写像 (encoding) および GA 空間から問題空間への逆写像 (decoding) を規定する問題をいう. 図 3 は問題空間と GA 空間の対応関係を示している.

図 3 では, 問題空間から GA 空間に向かうアークは表現型から遺伝子型へのコード化を, GA 空間から問題空間に向かうアークは遺伝子型から表現型への逆コード化を表わしている. GA 空間上で個体 P 1 と P 2 の間での交叉により, 子の個体が生成され, これを逆コード化することにより, 問題空間上に, 親 P 1 と P 2 の性質を何らかの意味で継承した個体 C 1 と C 2 が生成されることを図 3 は示している.

simple GA でも採用されている 1 点交叉 (one point crossover) では, 2 つの染色体の間で, ランダムに選ばれた交叉箇所の後半部分を交換する. これを複数の交叉点に拡張した操作を多点交叉 (multi point crossover) といい, あらゆる多点交叉が一様に起こるような操作を一様交叉 (uniform crossover) という.

1 点交叉に対して, GA の性能は遺伝子座の染色体上へのコード化に敏感であるが, 一様交叉に対しては遺伝子座の位置には無関係なことが知られている. 一般に, コード化および交叉の設計は, GA の基本的性能を決定づける重要な課題である.

コード化および交叉が不適切であれば, GA 空間で交叉によって生成された個体を問題空間に逆写像したとき

実行不可能解になるかもしれない. また, 実行可能であっても, 両親とはまったくかけ離れた子の個体が生成されてしまうかもしれない. 不適当な個体が生成される可能性が高いほど, GA はランダムサーチより劣る性能を示すことになるであろう.

コード化および交叉の評価規範として次の条件が考えられる [5].

[コード化評価規範]

1) 完備性 (completeness)

問題空間上の解候補はすべて染色体として表現できることが望ましい.

2) 健全性 (soundness)

GA 空間上の染色体はすべて問題空間の実行可能な解候補に対応づけられることが望ましい.

3) 非冗長性 (nonredundancy)

染色体 (GA 空間) と解候補 (問題空間) は 1 対 1 に対応づけられることが望ましい.

[交叉の評価規範]

4) 形質遺伝性 (character preservingness)

親の形質は交叉によって生成される子に適切に継承されることが望ましい.

完備性は自然な要求である. 致死遺伝子の生成を排除する健全性は, 一般に, これを満たすことが困難ではあるが, 交叉に限定を加えることにより, 充足できる場合がある. 致死遺伝子を許容するコード化は, 無駄な探索を強いられるが, 致死遺伝子の生成を抑制するための計算コストが節約され, 結果的に効率よく最適解を見出すことができる場合もあり得る. GA 空間と問題空間の対応関係が多対多の場合, 対応づけのあいまいさを解消するための規則が必要とされる.

図 3 に示すように, コード化と交叉は, 不可分の関係にあり, GA が動作する空間を定義する. コード化が上で述べた評価規範を満足していても, 交叉が不適切であれば, 好ましい解候補を生成することができず, GA はランダムサーチと同程度またはそれ以下の性能しか示さないであろう.

形質遺伝性は交叉が満たすべき評価規範として妥当と思われる. 形質の定義は問題領域に依存することから, 問題領域ごとに形質を定義したうえで, 形質遺伝性を満たすコード化と交叉の方法を見出すことが要求される.

コード化と交叉については, たとえば, 巡回セールスマン問題 (Traveling Salesman Problem : TSP) を対象に, これまで次のような工夫が試みられてきた.

パス表現 (path representation) と呼ばれるコード化では、都市名を遺伝子として、起点とする都市から巡回する順に都市名を列挙した文字列を染色体とする。パス表現された個体ペアに対し、1点交叉を適用すると、多くの場合、致死遺伝子を生じてしまう。

Grefenstette は、致死遺伝子を抑制するコード化における工夫として、順序表現 (ordinal representation) と呼ばれるコード化を提案している[6]。順序表現では、次に巡回する都市が未訪問都市リストの中で何番目にあるかを調べ、その番号を遺伝子とし、起点都市から順にこれを並べた文字列を染色体とする。順序表現された個体ペアに対し、1点交叉を適用すると、致死遺伝子の生成は抑制されるが、親から子に継承される形質は、いずれも交叉箇所の前半部分だけであり、後半部分に相当する親の形質は失われてしまうことに問題がある。

Goldberg は、致死遺伝子を抑制する交叉の工夫として、部分写像交叉 (partially mapped crossover) を提案している[7]。部分写像交叉では、パス表現された個体ペアに対し、1点交叉により、染色体の前半と後半を交換した後、致死遺伝子を生じないように、部分的な入れ替え、すなわちパッチを当てることにより、致死遺伝子を抑制する。

致死遺伝子の抑制を重視したこれらの方法は、形質遺伝性に関しては不完全であるために、ランダムサーチと同程度の性能しか示さないことが知られている。

筆著らは、形質遺伝子を重視した交叉の工夫として、サブツアー交換交叉 (subtour exchange crossover) と呼ばれる方法を提案している[5]。サブツアー交換交叉は、2点交叉の特殊なもので、パス表現された個体ペアに対し、サブツアーに含まれる都市集合が一致するときに限って交叉を行なう。サブツアー交換交叉によって生成される子はいずれも、実行可能解であり、しかも両親の形質すなわちサブツアーをできる限り継承することが保証される。

平面 TSP では、交叉経路をもつツアーは最適解にはなりえない。サブツアー交換交叉を実装した GA では、500都市までの問題において、順序表現や部分写像交叉で排除できなかった交叉経路を排除できることが報告されている[5]。

3.2 突然変異の役割

突然変異の基本は、ある確率で選ばれた遺伝子座の値を他の対立遺伝子に置き換えることにある。突然変異の変型として、逆位 (inverse) がある。逆位とは、染色体

のある部分文字列に着目して、その順序を反転させることをいう。染色体上の異なった位置にある2つの文字または部分文字列の位置を交換することも突然変異の1つの方法である。

突然変異は、染色体上のある遺伝子座の値を他の対立遺伝子に置き換えることにより、個体の近傍に新しい個体を生成するもので、その意味で突然変異は、局所的なランダムサーチの一種とみなせる。突然変異は致死遺伝子を生じる危険性を内包している反面、集団として喪失した対立遺伝子の回復に寄与することも期待でき、集団の多様性を維持するうえで有効な1つの手段ともいえる。

交叉と突然変異の間には、相補的な側面と競合的な側面が存在する。集団全体が探索空間のある超平面に陥ってしまった場合、突然変異により超平面からの脱却が可能となることが期待できることおよび突然変異による局所的な探索が有効に働くこともあり得るという意味において、突然変異は交叉に対して相補的である。しかし、交叉によって形成された優れた形質が突然変異により破壊されてしまうことがあるという意味において、突然変異は交叉に対して競合的である。

3.3 選択戦略および選択圧力

選択の目的は、集団における個体の適応度の分布にもとづいて、交配のための候補を選抜するための確率的決定を行なうことにある。

Simple GA で採用されている代表的な選択戦略であるルーレット選択 (roulette selection) では、ある個体を選択される確率は集団の適応度の総和に対するその個体の適応度の比によって与えられる。

選択によって集団が適応度の高い個体群に収束していく様子を比喩的に選択圧力 (selection pressure) という。ルーレット選択は、同じ個体の復元抽出 (sampling with replacement) を許容していることから、他と比べて適応度が突出した個体が存在するとき、その個体が繰り返し選ばれる可能性が高く、選択圧力が高くなりやすい方法といえる。

選択圧力を調整する手っ取り早い手段としてスケールリング (scaling) が考えられる。集団内に突出した適応度をもつ個体が存在したり、反対に集団内での適応度のバラツキが小さいとき、適応度に対し線形または非線形の変換を施すことにより、各個体の選択確率が適応度の違いに応じて適切に配分されるように調整する。

集団内に同じ染色体をもつ個体が複数個存在すると

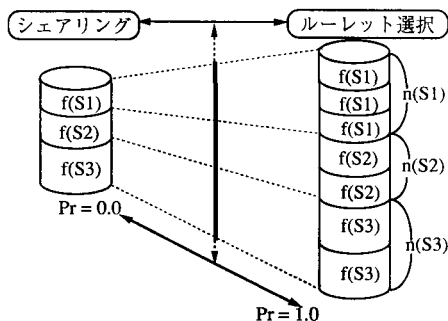


図 4 選択圧力制御モデル

き、ルーレット選択では、その個体の選択確率が複数倍されてしまうため、同じ個体の重複抽出が加速される効果が生じる。同じ個体の行き過ぎた重複抽出を抑制するために、集団内に同じ個体が複数個存在していても、選択に際しては、同じ個体は1個しか存在しないものとして選択するシェアリング (sharing) が提案されている。シェアリングは、ルーレット選択に比べて、選択圧力は低い方法といえる。

図4は、シェアリングとルーレット選択を線形結合した選択圧力モデルの一例を示している。図4では、個体の種類を S_i 、種類 S_i の適応度を $f(S_i)$ 、種類 S_i の個体数を $n(S_i)$ で、それぞれ表わしている。選択圧力パラメータ P_r は0~1の間の値を取り、 $P_r=0$ のときシェアリングに一致し、 $P_r=1$ のときルーレット選択に一致する。

一般に、探索の初期局面では、探索空間を広く探索するために、選択圧力を低く設定することが望ましく、探索の終期局面では、収束を加速するために、選択圧力を高く設定することが望ましい。図4の選択圧力制御モデルを採用する場合、選択圧力パラメータ P_r の値を徐々に上げていく制御則が考えられる。この制御則は、シミュレーテッドアニーリングにおける温度パラメータ T を徐々に下げていくことに対応づけられる。

一方、選択が適応度の分布に影響されることを排除するために、適応度の順位だけにもとづいて選択確率を決定するランクベース選択 (rank based selection) がある。

また、確率的選択では、適応度最大の個体であっても淘汰されることを許容しているが、最適化問題のタイプによっては、適応度最大の個体を無条件で次世代に残すエリート保存戦略 (elitest preserving strategy) を併用することも考えられる。

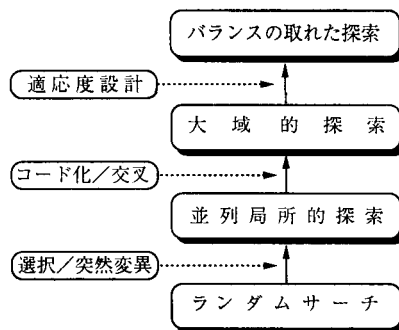


図 5 GAの性能レベルと設計レベルの対応関係

選択は、集団の中から個体を取捨選択するだけであり、それ自身は新しい個体を生成する能力をもたないが、問題に合った適切な制御戦略を設定し、選択圧力を適切に変化させることにより、大域的探索から局所的探索への円滑な制御を実現する役割をになうことができる。

4. GAの性能レベルと特徴

GAは確率的かつ並列的なアルゴリズムであり、局所的探索と大域的探索を効果的に組み合わせることにより、既存の探索手法と比べて、頑健な挙動を示すことが期待できる。しかし、GAの潜在的な能力をフルに引き出すためには、GAのそれぞれの設計段階において構成要素やパラメータを適切に設計することが必要である。

4.1 GAの性能レベル[8]

図5にGAの設計レベルと性能レベルの関係を示す。モンテカルロ法に代表されるランダムサーチでは、ある一定の分布に従って、解候補をサンプリングする探索方法であり、近似解法としてはもっとも一般性が高いが、すでに発見された解候補の情報を利用しないことから、高い性能を期待することは困難である。GAは少なくともランダムサーチが示す性能レベルをクリアしていなければならない。

交叉を排除した選択と突然変異だけのGAは、シミュレーテッドアニーリングの並列バージョンに相当する。選択圧力を制御することにより、局所解からの脱却はある程度可能ではあるが、大域的探索能力をもたないことに限界がある。

GAにおける大域的探索能力は交叉に求められる。3.1節で議論したように、交叉が有効に機能するためには、問題に適したコード化が必要である。コード化/交叉設計に失敗すると、ランダムサーチと同程度の性能しか得られない場合がある。

コード化/交叉が適切に設計されても、適応度関数の設計に失敗すると、騙し問題 (deceptive problem) が生じて、局所解に陥る場合がありうる。

図5は、コード化/交叉の設計、適応度の設計が適切に行なわれ、さらに実装パラメータの調整が適切に行なわれることにより、大域的探索と局所的探索のバランスのとれた探索が実現されることを示している。

バランスのとれた探索は、特に組合せ最適化問題に要求されるものであり、GAの枠組みはそのような潜在能力をもつものと考えられる。

4.2 GAの特徴[9]

探索手法としてのGAは、次のような特徴をもつ。

(1) 確率探索法 (Stochastic Search)

GAにおける選択操作は、確率的であることから、適応度の低い個体であっても、交配の候補として選抜される可能性を許容している。これは、シミュレーテッドアニーリングにおける確率的な状態遷移に相当する。選択圧力を動的に制御することにより、SAにおける温度スケジューリングをシミュレートすることが可能である。

(2) 多点探索法 (Multi-Point Search)

GAにおける集団サイズは、ビームサーチ (beam search) におけるビーム幅に相当する。ビームサーチでは、多点情報を新しい探索点の生成に利用していないが、GAでは、交叉を通じて新しい探索点の生成に多点情報を利用している。

(3) 直接探索法 (Direct Search)

GAは、適応度関数の微分値を陽に利用しないことから、直接探索法 (Direct Search) の一種である。さらに、ランクベース選択を採用すれば、適応度関数の値を序数的にしか利用しないことから、解候補群を人間の評価者が主観によって順序づけることさえすれば、GAは動作可能であるので、GAは対話型解法としても利用できる。

(4) 並列探索法 (Parallel Search)

GAにおける遺伝的操作のうち、交叉と突然変異の操作は並列処理が可能であり、超並列マシンとの親和性も高いことから、超並列マシンまたは専用マシンを用いた

GAの高速化は、きわめて近い将来に、実現・普及するものと見込まれる。

今回は、GAの基本的概念、構成要素およびGAの設計問題を中心に議論した。次回には、組合せ最適化問題、特にスケジューリング問題を対象として、GAによるモデル化の実際を紹介する。

参 考 文 献

- [1] Hopfield, J. J. and Tank, P. W. : Neural Computation of Decisions in Optimization Problems, *Biological Cybernetics*, Vol.81 (1985) 141-152.
- [2] Kirkpatrick, S., et al. : Optimization by Simulated Annealing, *Science*, Vol. 220 (1983) 671-680.
- [3] Holland, J. H. : *Adaptation in Natural and Artificial Systems*, The Univ. Michigan Press, 1975 and MIT Press, 1992.
- [4] Goldberg, D. : *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, 1989.
- [5] 山村, 小野, 小林 : 形質の遺伝を重視した遺伝的アルゴリズムに基づく巡回セールスマン問題の解法, *人工知能学会誌*, 7, 6 (1992), 1049-1059.
- [6] Grefenstette, J., Gopal, R., Rosmaita, B.J. and Van Gucht, D. : Genetic Algorithms for the Traveling Salesman Problem, *Proc. of ICGA '85*, 1985, 16-168.
- [7] Goldberg, D. and Lingle, R. : Alleles, loci, and the travelling salesman problem, *Proc. of ICGA '85*, 1985, 154-159.
- [8] 山村, 小林 : 遺伝的アルゴリズムによる組合せ最適化, *シミュレーション*, 12, 1 (1993), 75-81.
- [9] 小林 : 遺伝的アルゴリズムの現状と課題, 計測と制御, 32, 1 (1993), 2-9.