



Title	進化計算におけるリンケージ同定
Author(s)	棟朝, 雅晴
Description	「進化計算の新展開特集号」解説
Citation	システム/制御/情報, 52(10), 362-367
Issue Date	2008-10-15
Doc URL	https://hdl.handle.net/2115/44483
Type	journal article
File Information	system_munetomo.pdf



進化計算におけるリンケージ同定

棟朝 雅晴*

1. はじめに

遺伝的アルゴリズムにおいて最も重要な遺伝的操作である交叉を効果的に適用するためには、互いに関連する遺伝子を密に符号化することが必要とされている。なぜなら、ビルディングブロックを構成する遺伝子が密に符号化されていない場合、単純な交叉によりそれら有用なビルディングブロックが容易に破壊され、交叉が正しく働かないためである。しかしながら、一般の問題において密な符号化を保証することは困難な場合も多いので、互いに関連する遺伝子を同定し、その情報をもとにした交叉を行うリンケージの同定・学習に関する手法が数多く開発されている。以下では、我々が開発してきた、非線形性尺度および非単調性尺度にもとづくリンケージ同定手法を中心として、リンケージ同定に関連する手法をサーベイする。

2. リンケージとは

遺伝学において、「リンケージ (linkage)」は「2つもしくはそれ以上の対立遺伝子が次の世代へ同時に継承される傾向」と定義される [1]。現実の生物の遺伝子を観測、解析するという観点からは、この定義は意味のあるものであるが、これをそのまま進化計算へ適用すると、それら遺伝子が符号化された文字列内で近くに存在するという意味となり、単なる符号化の結果を示すだけで、問題を解くという観点からは有用な定義とは言い難い。我々に求められているのは、進化計算において問題を解くため、いかにしてビルディングブロックを生成、交換するかであるため、我々は進化計算における「リンケージ」の定義を「2つもしくはそれ以上の遺伝子座における遺伝子が、解くべき問題に関して有用な部分解 (ビルディングブロック) を構成する傾向」と定義することとする。また、「密なリンケージ」(tight linkage) とは有用な部分解を構成する遺伝子が個体の文字列上で互いに近くに符号化され、存在することを意味し、1点交叉など古典的な交叉オペレータを用いる際に重要な前提条件と

なる。

例えば、以下に示される2つの個体において、下線部の遺伝子が有用な部分解を構成するものと仮定すると、

(a)	(b)
1101 <u>01</u> 00101	<u>11</u> 010100 <u>01</u> 01
0010 <u>11</u> 01100	<u>00</u> 101101 <u>10</u> 0

1点交叉を適用するとき、(a)の場合には問題が生じないが、(b)の場合には高い確率で部分解が破壊されてしまう。(a)のような状況をリンケージが密である (tight) といい、(b)のような状況をリンケージが疎である (loose) という。リンケージが疎となる符号化を避けることが、遺伝的アルゴリズムにおける古典的な指針の一つである。しかしながら、現実の問題において密なリンケージを保証する符号化が困難な場合も多く、いくつかの問題においては原理的に不可能である。例えば、ネットワーク構造の最適化問題などでは、2次元のネットワーク構造を1次元の文字列に強制的に符号化すると、問題の有する隣接関係をすべて保持することは不可能である。遺伝的操作を工夫することで、有用な部分解を保持できる場合もあるが、そのような遺伝的操作を開発することは、すなわち問題を人手により部分的に解くことに相当する。そこで、人手により符号化や遺伝的操作を工夫する代わりに、より包括的なアプローチにより有用な部分解の破壊を防ぎつつ効果的に探索をすすめる手法の開発が求められている。リンケージ同定手法とは、解くべき問題の情報として、適応度のみを用いてリンケージを検出、もしくはリンケージ集合を生成することで、交叉によるビルディングブロック破壊を防ぎ、効果的な遺伝的探索を実現する方法論の総称である。

3. リンケージ同定以前

密なリンケージの重要性は遺伝的アルゴリズムが提案された当初から認識されており、逆位 (Inversion) 操作によるリンケージの探索が試みられていたが、リンケージの探索とビルディングブロックの探索を同時に行うことができないため、適切なリンケージを生成することが困難であった。

Goldberg による messy GA [2] では、ビルディングブ

* 北海道大学情報基盤センター Information Initiative Center, Hokkaido University; North 11, West 5, Kitaku, Sapporo city, Hokkaido 060-0811, JAPAN

Key Words: evolutionary computation, linkage identification, estimation of distribution algorithms.

ロックを直接処理することで、リンケージの問題を解決することができる。しかしながら、そのために、初期個体群として長さ k のスキーマとしてすべての組み合わせを生成する必要があり、実用的なアルゴリズムとはならなかった。

Harikらによる Linkage Learning GA[3]においては、円環状に符号化された個体と、2点交叉に類似した遺伝的操作を用いることで、あるビルディングブロックに着目した場合に、それを構成する遺伝子が自動的に密に集まることを理論的に示している。ただし、この条件が満たされるためには、それぞれのビルディングブロックの適応度全体に関する寄与度合いに十分な差が必要であり、一様な適応度寄与を有するビルディングブロックについては、うまく動作しないことが知られている¹。

また、Karguptaによる gene expression messy GA[4]では、それぞれの遺伝子について摂動を行い、その局所最適性を調べることで、間接的にリンケージの情報を抽出している。具体的には、それぞれの遺伝子に対して、それが発現するかどうかの重みを割り当て、摂動による適応度の悪化度合いをもとに重みを計算している。

4. 非線形性尺度にもとづく手法

リンケージ同定の基本的な考え方は「遺伝的アルゴリズムにとって探索が易しい遺伝子座のペアについては、交叉によりそれぞれ独立して探索可能であるため、分離可能である。逆に難しい遺伝子座については、交叉により有用なビルディングブロックが失われる可能性があるため、リンケージとしてまとめる必要がある。」というものである。

リンケージ同定手法として最初に開発された LINC (Linkage Identification by Nonlinearity Check)[5]では、任意の遺伝子座のペア (i, j) に対して遺伝子の摂動を行い、それによる適応度の変化量が、非線形の効果となるかどうかを判別することで、リンケージの同定を行う。具体的には、十分なサイズを有する集団 P 内のすべての個体 s に対して、以下を計算し、

$$\Delta f_i(s) = f(\dots \bar{s}_i \dots) - f(\dots s_i \dots) \quad (1)$$

$$\Delta f_j(s) = f(\dots \bar{s}_j \dots) - f(\dots s_j \dots) \quad (2)$$

$$\Delta f_{ij}(s) = f(\dots \bar{s}_i \dots \bar{s}_j \dots) - f(\dots s_i \dots s_j \dots), \quad (3)$$

それらを用いて、

$$\Delta f_{ij}(s) \neq f_i(s) + f_j(s) \quad (4)$$

となる s が存在する場合に、遺伝子座のペア (i, j) についてリンケージが存在するものと判断する。現実の問題においては、厳密に等式が成立するかどうかを調べる代わりに誤差 $\epsilon > 0$ を許し、

¹この問題点を解決するため、イントロン(解釈されない遺伝子)を挿入する方法などが提案されている。

$$|\Delta f_{ij}(s) - (\Delta f_i(s) + \Delta f_j(s))| > \epsilon \quad (5)$$

が満たされるかどうかを調べる。

ここで重要なのは、1つの個体で判定するのではなく、十分なサイズの集団中の個体すべてに対して非線形性を有するかどうかの判定を行う点である。理論解析の結果、ビルディングブロックの最大長が k で抑えられるとき²、 $O(k2^k)$ という比較的少ない個体数で正しいリンケージが同定できることが示されている。

具体的な LINC のアルゴリズムは以下の通り単純なものとなる [24]。

- (1) 初期集団 P として $O(k2^k)$ 個の個体をランダムに生成する。
- (2) $i = 1, \dots, l$ に対して (3)~(6) を繰り返す
- (3) $j = i + 1, \dots, l$ に対して (4)~(6) を繰り返す
- (4) 集団 P 内の全ての個体 s に対して以下を繰り返す
- (5) $\Delta f_i(s), \Delta f_j(s), \Delta f_{ij}(s)$ を計算する
- (6) $|\Delta f_{ij}(s) - (\Delta f_i(s) + \Delta f_j(s))| > \epsilon$ を満たせば (i, j) のペアをリンケージとして保存する。
- (7) 得られたリンケージのペア (i, j) からリンケージ集合を生成する。

より一般的な枠組みとして、非線形性の尺度(エピスタシス尺度)を定義することで、より柔軟なリンケージ同定を行う LIEM (Linkage Identification with Epistasis Measure) が提案されている [7]。LIEM のエピスタシス尺度は以下の式により定義される。

$$e_{ij} = \max_{s \in P} |\Delta f_{ij}(s) - (\Delta f_i(s) + \Delta f_j(s))| \quad (6)$$

ここで、 e_{ij} は遺伝子座ペア (i, j) に関するエピスタシス尺度であり、この値が大きいほど、非線形性が強くリンケージとなる可能性が高いことを示している。逆に $e_{ij} = 0$ の場合には、LINC の条件において完全な線形性に相当し、リンケージが存在しないことを示している。LIEM のアルゴリズムは以下の通りとなる [24]。

- (1) 初期集団 P として $O(k2^k)$ 個の個体をランダムに生成する。
- (2) 全ての遺伝子座のペア (i, j) に関して、式 (6) にしたがいエピスタシス尺度 e_{ij} を計算する。
- (3) それぞれの遺伝子座 i に関して、以下の (4)~(5) を実行する。
- (4) i を固定して e_{ij} を j に関して値の大きい順にソートする。
- (5) e_{ij} の値の大きい順に、その遺伝子座 j を k 個取出し、 i に関するリンケージ集合とする。

²ビルディングブロックの最大長 k は、GA 一般においてその問題の難しさを定義する上で本質的なパラメータであり、 $l \gg k$ (l : 個体の長さ) となることが選択交叉型の GA で問題が解けるための条件である。これは、GA の初期集団サイズが $O(2^k)$ となる理論式からも導き出される [24]。

5. 非単調性尺度にもとづく手法

線形・非線形の条件は、遺伝的アルゴリズムにとって必ずしも本質ではない。たとえば、ある遺伝子の摂動による適応度の改善と別の遺伝子の摂動による改善がなされる場合、それらの遺伝子を同時に摂動させた時、必ずしも適応度の改善が線形和とならなくてもよく、線形和以上の改善がなされるような場合にも、遺伝的操作により、容易に探索を行うことができる。

以上の考えをもとに開発されたのが、LIMD (Linkage Identification with non-Monotonicity Detection)[6]であり、以下の単調性条件が満たされない場合に遺伝子座ペア (i, j) を同じリンケージ集合に属するものと判断する。

$$\begin{aligned} &\text{if } (\Delta f_i(s) > 0 \text{ and } \Delta f_j(s) > 0) \\ &\text{then } (\Delta f_{ij}(s) > \Delta f_i(s) \text{ and } \Delta f_{ij}(s) > \Delta f_j(s)) \end{aligned} \quad (7)$$

$$\begin{aligned} &\text{if } (\Delta f_i(s) < 0 \text{ and } \Delta f_j(s) < 0) \\ &\text{then } (\Delta f_{ij}(s) < \Delta f_i(s) \text{ and } \Delta f_{ij}(s) < \Delta f_j(s)) \end{aligned} \quad (8)$$

これにより、非線形性条件に比べ、より広範囲の問題について正しいリンケージ情報を検出することが可能となった¹。

また、LIEMにおけるエピスタシス尺度を、LIMDの条件をもとにしたものと置き換えた手法として、LIEM² (Linkage Identification with Epistasis Measure considering Monotonicity)[8]が提案されている。LIEM²におけるエピスタシス尺度は、LIMDの条件が満たされる点からの距離として求められ、以下の式で定義される。

$$e_{ij} = \max_{s \in P} g(\Delta f_{ij}(s), \Delta f_i(s), \Delta f_j(s)). \quad (9)$$

$$g(x, y, z) = \begin{cases} tr(y-x) + tr(z-x), & (y > 0, z > 0) \\ tr(x-y) + tr(x-z), & (y < 0, z < 0) \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

$$tr(x) = \begin{cases} x, & (x \geq 0) \\ 0, & (x < 0) \end{cases} \quad (11)$$

LIMD、LIEM²のアルゴリズムについては、それぞれLINC、LIEMと同様となる。

6. さまざまな拡張

リンケージ同定について上記のような「古典的な」手法が提案された後、さまざまな拡張が試みられている。代表的な手法として、Coffinらにより提案されたgLINC[9]では、遺伝子単体ではなく、遺伝子のグループ単位での、

¹例えば、LINCで正しいリンケージ集合が検出可能な、部分関数の線形和で表される関数(ただし、それぞれの部分関数の遺伝子がランダムに符号化されていても良い)に加えて、それを n 乗した関数(それぞれの部分関数が正である場合)についても正しいリンケージ集合が検出可能であることが示されている。

より高次の摂動を行うことでリンケージ同定を行っている。また、Heckendornらは、摂動の概念をより一般化した Limited Probing[10]を提案している。

さらに、階層型の問題構造に対応した、階層型のリンケージ同定手法 hLIEM (hierarchical LIEM)が提案されている[12]。hLIEMにおいては、徐々にリンケージ集合をマージしていくことで、階層型の問題構造に対応した木構造のリンケージ集合を生成することができる。hLIEMはネットワークの設計問題に適用され、従来型のGAや単純なリンケージ同定と比べて、高い最適化性能を有することが示されている[13]。

リンケージ同定においては、それぞれの遺伝子座のペア (i, j) についてリンケージが存在するかどうかを判定するが、そこで得られた結果をもとに、リンケージ集合(互いに関連し、密に符号化されるべき遺伝子座の集合)を生成する必要がある。リンケージに重複が存在しない場合は問題ないが、重複が存在する場合はそれを除去した上で、互いに重複しないリンケージ集合を生成する必要がある。LIMD-TD (Tightness Detection)[6]では、それぞれの遺伝子座において、それらの関連性の強さを求め、それをもとに重複を除去する。具体的にはLIMDにより生成された遺伝子座 (i, j) のペアに関するリンケージの情報から、それぞれの遺伝子座 i に関するリンケージ集合を生成した上で、以下で定義される密な度合 (tightness) を求め、それが一定の閾値未満の場合に、リンケージ集合から除去する。

$$tightness(i, j) = \frac{n1(i, j)}{n1(i, j) + n2(i, j)} \quad (12)$$

ここで、 $n1(i, j)$ は i, j の両方を含むリンケージ集合の数、 $n2(i, j)$ は i, j のいずれかを含むリンケージ集合の数である。この方法は閾値の設定が難しいという欠点がある。

YuらによるDSM-GA[11]は、LINCなどにより得られたリンケージ同定の結果を、依存関係行列 (Dependency Structure Matrix, DSM) $[d_{ij}]$ (リンケージの条件を満たす場合 $d_{ij} = 1$, 満たさない場合 $d_{ij} = 0$) として表現し、以下に示される評価尺度を最小とするリンケージ集合への分割を求める手法である。

$$f_{DSM}(M) = (n_c \log(n_c) + \log(n_n) \sum_{i=1}^{n_c} cl_i) + (|S|(2 \log(n_n) + 1)) \quad (13)$$

ここで、 M はモデル (リンケージ集合)、 n_c はリンケージ集合の数、 n_n は遺伝子座の数、 cl_i はそれぞれ i 番目のリンケージ集合に含まれる遺伝子座の数、 S は $d_{ij} = 1$ となるが、リンケージ集合による分割に当てはまらない要素の数である。

さらに複雑なリンケージの重複構造を有する場合には、後述するリンケージ重複を考慮した交叉手法が提案

されている。

7. 実数値 GA におけるリンケージ同定

以上で紹介されてきたリンケージ同定手法は、基本的に二進符号化によるビット列を対象としている。実数値 GA に対するリンケージ同定手法についても、いくつかの手法が提案されている。

Tsutui らは、実数値 GA における単体交叉と組み合わせでリンケージを同定する LISS (Linkage Identification with Single Stage Evolution) と LIMS (Linkage Identification with Multi Stage Evolution) を提案し、比較検討を行っている [14]。これらのアルゴリズムでは、PICI (Piecewise Interval Correlation by Iteration) アルゴリズムを用いた相関関係の検出にもとづいてリンケージの同定を行っている。

また、実数値 GA におけるリンケージ同定手法として、手塚らにより LINC-R および LIDI-R が提案されている [15]。LINC-R においては、LINC における非線形条件の判定と同じ条件を用いており、遺伝子の摂動を実数値に合わせて設計している。LIDI-R は適応度差分の符号に着目してリンケージ同定を行う手法であり、差分の符号独立性に関する条件 (LIMD の単調性に類似した条件) をもとにそれが成り立たない場合に、リンケージが存在するものと判断する。

摂動を行う場合、ビット列においては $0 \rightarrow 1$ または $1 \rightarrow 0$ となるが、実数値の場合には摂動に関する任意性があり、完全に解空間を網羅することが原理的にできない。上記の手法においては、適切な摂動を行うため、リンケージを検出可能な確率をもとにした理論的考察を行い、必要な集団サイズを求めている。

8. 確率モデルとの融合

リンケージ同定に関連する手法として、確率モデル構築型の GA (分布推定アルゴリズム, Estimation of Distribution Algorithm (EDA) と呼ばれる) があげられる。ここでは、その詳細について議論しないが、例えば、BOA (Bayesian Optimization Algorithm) [16] では、集団から選択を行った後の「有望な」個体集団の遺伝子分布から、それを最もよく説明する (具体的には Bayesian-Dirichlet メトリックなどの尺度を基準に判断する) ベイジアンネットワークを生成し、それによる条件付き確率モデルをもとにして、次世代の個体集団を生成する。この場合、リンケージに相当する情報はベジアンネットワークの構造として学習がなされることとなる。

Harik による ECGA (Extended Compact GA) [17] においては、最小記述長 (Minimum Description Length) 原理にもとづいた複雑度を定義し、それを最小化するモデルとしてリンケージ集合を求めている。具体的には、確率モデルを記述するために必要な記述長としてのモデル複雑度と、集団をそのリンケージ集合で分割した場合

にエントロピーをもとにした集団複雑度の和を最小化するリンケージ集合を求めている。

また、確率モデルの考え方と、リンケージ同定における摂動を組み合わせた手法として D^5 (Dependency Detection for Distribution Derived from Df) [18] が提案されている。 D^5 は従来型のリンケージ同定と確率モデル構築型の GA の両方の良さを取り入れたリンケージ同定手法の実現を目指して開発された。LINC, LIMD など従来型のリンケージ同定においては、遺伝子座のペアに対して遺伝子の摂動を行う必要があるため、個体の長さを l とすると、 $O(l^2)$ の適応度評価が必要となる。また、確率モデル構築型の GA では、選択後の個体に対して確率モデルの構築を行うため、適応度全体への寄与が少ないビルディングブロックに関するリンケージ情報の検出が難しくなること、モデル構築の計算量が大きくなるなどの問題点がある。

D^5 においては、遺伝子座のペアではなく、それぞれの遺伝子に対して摂動を行った後、その摂動による適応度の変化量にもとづいて集団内の個体のクラスタリングを行い、生成されたそれぞれのクラスタにおいて、エントロピー基準にもとづいたリンケージ同定を行う。 D^5 のアルゴリズムを以下に示す。

- (1) n 個の個体からなる集団 P を初期化する
- (2) $i = 1, 2, \dots, l$ について以下の (3)~(5) を繰り返し実行する
- (3) P 内のそれぞれの個体 s について i 番目の遺伝子を摂動させた場合の適応度の変化量 $\Delta f_i(s)$ を計算する
- (4) P 内の個体を $\Delta f_i(s)$ の値に基づきクラスタリングする
- (5) クラスタリングされた個体のエントロピーに基づき、リンケージ集合を生成する

D^5 ではクラスタリングとして重心法を用いているが、他の手法も用いることができる。上記のアルゴリズムの (5) において、クラスタリングされたそれぞれのクラスタ C_p においてエントロピー基準に基づき、以下のアルゴリズムを実行して、 i 番目の遺伝子座に関するリンケージ集合 V_i を生成する。ここで、 k はリンケージ集合のサイズの最大値、 $E(W)$ は集合 W の個体集団のエントロピーである。

- (1) すべてのクラスタ C_p について (2)~(7) を実行する。
- (2) $|W_2| < k$ を満たす間、(3)~(6) を繰り返す。
- (3) $W_1 = \{1, 2, \dots, i-1, i+1, \dots, l\}$, $W_2 = \{i\}$ とする。
- (4) すべての $j \in W_1$ について $E_j = E(W_2 \cup \{j\})$ (エントロピー) を計算する。
- (5) $h = \operatorname{argmin}_{j \in W_1} E_j$ を求める。
- (6) $W_1 = W_1 - \{h\}$; $W_2 = W_2 \cup \{h\}$ とする。
- (7) $W_p = W_2$; $E_p = E(W_2)$ とする。
- (8) $p' = \operatorname{argmin}_p E_p$ を求める。
- (9) リンケージ集合を $V_i = W_{p'}$ として求める。

D^5 の特長として、適応度全体に対して寄与度合いの小さなビルディングブロックを正しく検出するとともに、リンケージ同定に必要な適応度の評価回数を $O(l \log l)$ に抑えられる点があげられる。特に解くべき問題のサイズが大きくなった場合に、 D^5 は他のリンケージ同定手法、確率モデル構築型GAと比べて大きなメリットを有することとなる。

9. リンケージ重複を考慮した交叉

リンケージ同定によりリンケージ集合を求めたのち、その情報をもとに交叉を適用する。求められたリンケージ集合に重複が存在しない場合には、単純にリンケージ集合ごとにまとめて遺伝子を交換することでビルディングブロックを破壊することなく交叉を行うことができるが、リンケージ集合に重複が生じた場合には交叉によりビルディングブロックが破壊される可能性がある。LIMD-TDやDSM-GAでは、遺伝子座ペアに関するリンケージの情報から重複のないリンケージ集合を生成することで、ビルディングブロックの破壊を回避しているが、解くべき問題が本質的に重複するリンケージ構造を有するような難しい問題である場合、強制的に重複のないリンケージ集合を生成することで、問題の有する情報が失われてしまう可能性がある。そこで、重複するリンケージ集合の生成を許しつつ、できる限りビルディングブロックの破壊を防ぐための手法が提案されている。

Yuらによる手法[19]では、リンケージ集合をノード、それらの重複関係をリンクとして表現したネットワークを生成し、その最小カットを求めて、それにもとづいた分割により交叉を行うことで、ビルディングブロックの破壊をできる限り防ぐ交叉手法が提案されている。この手法はリンケージ集合の重複に柔軟に対応できるが、問題によっては最小カットのパターンが限られ、交叉の多様性が失われる可能性があるという欠点が存在する。

Yuらの手法を改良したCDC (Context Dependent Crossover)[20]では、リンケージ集合の重複関係を表すネットワークの最小カットを求める際に、実際に交叉を行う個体ペアの遺伝子の情報を考慮してネットワーク構造を再構成することで、ビルディングブロック破壊を防ぎつつ、多様な交叉パターンを確保することができる。具体的には、ある個体ペアに対して交叉を行う際に、Yuらの手法により生成されたリンケージ構造のネットワークから、その個体ペアにおいて交叉を行った場合にビルディングブロック破壊が起こらない(すなわち、リンケージの重複部分における遺伝子の値が個体ペア間で同一であり、交叉において重複の影響が生じない)重複関係を相当するリンクを除去した後に最小カットを求める。

リンケージ同定手法とCDCを併用することで、複雑なリンケージ構造を有し、従来型のGAやリンケージ同定手法などで解くことが極めて難しいような問題に対して、重複したリンケージ構造を最大限活用しつつ、適切

な交叉を適用することが可能となった。

具体的な適用例としては、 D^5 にCDCを組み合わせた最適化アルゴリズム $D^5 + \text{CDC}$ により、地理的な制約、通信リンクの制約、トラフィック条件など複雑な制約条件を有するネットワーク設計問題において、従来型のGAや、単純なリンケージ同定、階層型のリンケージ同定のすべてを上回る品質の解を得ることに成功している[21]。

10. リンケージ同定の並列化

確率モデル構築型GAにおいては、確率モデルに関する情報を共有する必要があるため、並列化には工夫が必要であるが、リンケージ同定アルゴリズムは、それぞれの遺伝子座(ペア)ごとに処理を行うため、並列化が極めて容易である。確率モデル構築型GAの並列化については、PBOA(Parallel BOA)[22]などが提案されている。PBOAでは、BOAにおいて計算負荷の多くを占めているベイジアンネットワークの探索を並列化している。リンケージ同定の並列化についてはいくつかの手法が提案されているが、代表的なものは、pLINC (parallel LINC)[23]である。また、 D^5 の並列化に関する研究もなされており、問題長が数十万ビットに達するような大規模問題において、リンケージ構造を同定し、最適解を得ることに成功している。

11. おわりに

本解説記事においては、進化計算において重要な研究分野の一つとなりつつあるリンケージ同定について、その古典的な手法から、最新の成果まで概略を述べた。最新のアルゴリズムである、 $D^5 + \text{CDC}$ では、問題の有する複雑な構造を、リンケージ集合という形で少ない計算量により効果的に同定するとともに、重複を有するリンケージ集合に対応した高性能な交叉オペレータにより、ビルディングブロックの破壊を防ぎつつ、高品質な解を生成することができる。

リンケージ同定や確率モデル構築型GAに関する研究の進展により、問題の構造を「遺伝子解析」の形で分析しつつ最適化を行う次世代の進化計算アルゴリズム、最適化アルゴリズムの発展が期待される。今後の研究の方向性としては、実問題へのさらなる適用と、その効果の検証、問題の有する構造とリンケージ同定との関係に関する理論的研究、確率モデル構築型GAとリンケージ同定との関連に関する研究、多目的最適化におけるアルゴリズムの開発などがあげられる。

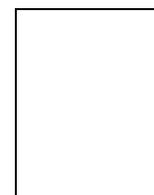
参考文献

- [1] P. C. Winter, G. I. Hickey, H. L. Fletcher: Instant Notes in Genetics, Springer (1998).
- [2] D. E. Goldberg: Messy genetic algorithms: Motivation, analysis, and first results, Complex Systems, No. 3, pp.493-530 (1989)

- [3] G. R. Harik and D. E. Goldberg. Learning linkage. *Foundations of Genetic Algorithms 4*, pp. 247-262. (1997)
- [4] Hillol Kargupta: The gene expression messy genetic algorithm. *Proceedings of the 1996 International Conference on Evolutionary Computation*, pp. 814-819 (1996)
- [5] Masaharu Munetomo, David E. Goldberg: Identifying Linkage Groups by Nonlinearity/Non-monotonicity Detection, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO99)*, pp.433-440 (1999)
- [6] Masaharu Munetomo, David E. Goldberg: Linkage Identification by Non-monotonicity Detection for Overlapping functions, *Evolutionary Computation*, vol.7, No.4, pp.377-398, MIT Press (1999)
- [7] 棟朝雅晴: エピスタシス尺度に基づくリンケージ同定手法の提案, *情報処理学会論文誌「数理モデル化と応用」*, Vol.43, No.SIG10(TOM7), pp.6-13 (2002)
- [8] Masaharu Munetomo: Linkage Identification with Epistasis Measure Considering Monotonicity Conditions, *Proceedings of the 4th Asia-Pacific Conference on Simulated Evolution and Learning*, pp.550-554 (2002)
- [9] David J. Coffin, Christopher D. Clack: gLINC: identifying composability using group perturbation, *Proceedings of the 2006 Genetic and Evolutionary Computation Conference (GECCO2006)*, pp.1133-1140 (2006)
- [10] Robert B. Heckendorn, Alden H. Wright: Efficient Linkage Discovery by Limited Probing, *Proceedings of the 2003 Genetic and Evolutionary Computation Conference (GECCO2003)*, pp.1003-1014 (2003)
- [11] Tian-Li Yu and David E. Goldberg: Dependency structure matrix analysis: Offline utility of the dependency structure matrix genetic algorithm, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO2004)*, Vol. 2, Lecture Notes in Computer Science 3103, pp. 355-366. Springer-Verlag (2004)
- [12] 辻美和子, 棟朝雅晴, 赤間清: 階層型問題のためのリンケージ同定手法の提案, *情報処理学会論文誌「数理モデル化と応用」*, Vol.45, No.SIG2(TOM10), pp.22-31 (2004)
- [13] Miwako Tsuji, Masaharu Munetomo, Kiyoshi Akama: Metropolitan Area Network Design Using GA Based on Hierarchical Linkage Identification, *Genetic and Evolutionary Computation - GECCO2003 Part 2, Lecture Notes in Computer Science 2724*, pp.1616-1617, Springer (2003)
- [14] Tsutsui, S. and Goldberg, D. E.: Simplex Crossover and Linkage Identification: Single-Stage Evolution VS. Multi-Stage Evolution, *Proceedings of the 2002 Congress on Evolutionary Computation (CEC2002)*, pp. 974-979 (2002)
- [15] 手塚大, 棟朝雅晴, 赤間清: 目的関数の加法分解性および差分の符号独立性にもとづく実数値遺伝的アルゴリズムのリンケージ同定, *情報処理学会論文誌「数理モデル化と応用」*, Vol.47, No.SIG 14, pp.43-53 (2006)
- [16] Martin Pelikan, David E. Goldberg, and E. Cant ´ u-Paz. BOA: The bayesian optimization algorithm. *Proceedings of the 1999 Genetic and Evolutionary Computation Conference*, pages 525-532 (1999)
- [17] G. R. Harik: The compact genetic algorithm. Technical Report IlliGAL Report No.97006, University of Illinois at Urbana-Champaign (1997)
- [18] Miwako Tsuji, Masaharu Munetomo, Kiyoshi Akama: Linkage Identification by Fitness Difference Clustering, *Evolutionary Computation*, Vol.14, No.4, pp. 383-409, MIT Press (2006)
- [19] Tian-Li Yu, Kumara Sastry, and David E. Goldberg. Linkage learning, overlapping building blocks, and systematic strategy for scalable recombination. *Proceedings of the 2005 Genetic and Evolutionary Computation Conference*, pp. 1217-1224 (2005)
- [20] 辻美和子, 棟朝雅晴, 赤間清: 複雑なビルディングブロック重複を持つ問題に対する交叉手法の提案, *情報処理学会論文誌「数理モデル化と応用」*, Vol.48, No.SIG15, pp.23-33 (2007)
- [21] Miwako Tsuji, Masaharu Munetomo, Akama Kiyoshi: A network design problem by a GA with linkage identification and recombination for overlapping building blocks, *Proceedings of the 2007 IEEE Congress on Evolutionary Computation*, pp.349-356 (2007)
- [22] Jiri Ocenasek: Parallel Estimation of Distribution Algorithms, PhD Thesis, Brno University of Technology (2002)
- [23] Masaharu Munetomo, Naoya Murao, Kiyoshi Akama: A Parallel Genetic Algorithm Based on Linkage Identification, *Genetic and Evolutionary Computation - GECCO2003 Part 1, Lecture Notes in Computer Science 2723*, pp.1222-1233, Springer (2003)
- [24] 棟朝雅晴: 遺伝的アルゴリズム その理論と先端的手法, 森北出版 (2008)

著者略歴

むね とも まさ はる
棟 朝 雅 晴 (非会員)



1968年9月10日生。1996年3月北海道大学大学院工学研究科情報工学専攻博士後期課程修了。同年4月北海道大学大学院工学研究科助手。1998年6月イリノイ大学基礎工学部客員研究員。1999年10月北海道大学情報メディア教育研究総合センター助教授。2003年4月北海道大学情報基盤センター助教授。2007年4月同准教授となり現在に至る。進化計算および大規模計算システムに関する研究に従事。IEEE, ACM SIGEVO, 情報処理学会などの会員。