

階層型問題のためのリンケージ同定手法の提案

辻 美和子[†] 棟 朝 雅 晴^{††} 赤 間 清^{††}

遺伝的アルゴリズムにおいて、ビルディングブロック破壊を防ぎ効率的に探索を行うための手段としてリンケージ同定が提案されている。リンケージ同定遺伝的アルゴリズムではビルディングブロックを構成する遺伝子座をあらかじめ調べ、単純遺伝的アルゴリズムでは暗黙に実行される問題の分割と組合せの処理を陽に実行する。しかし、実際の問題は、特にその規模が大きいとき、ビルディングブロックどうしも相互依存関係を持つような階層型の構造をとると考えられる。現在のリンケージ同定遺伝的アルゴリズムでは、遺伝子座どうしの相互依存関係は考慮されるものの、ビルディングブロックどうしは独立であるとして処理される。本論文では、遺伝子の値の摂動による適応度の変化量の非単調性に基づく単層型のリンケージ同定手法である LIEM² (Linkage Identification with Epistasis Measure considering Monotonicity) を拡張し、現実の問題に存在する階層構造のモデルの探索を可能にする。階層型リンケージ同定では、ビルディングブロックどうしの依存関係を再帰的に検出する。加えて、多様なビルディングブロック候補を保持するために、ニッチングを行う。

Linkage Identification for Problems with Hierarchical Structure

MIWAKO TSUJI,[†] MASAHARU MUNETOMO^{††} and KIYOSHI AKAMA^{††}

To avoid building block destructions, linkage identification techniques are proposed, which tries to identify a set of loci tightly-linked explicitly before performing genetic optimizations. Real-world problems, especially large-scaled complex problems, sometimes take hierarchical structures in which building blocks have recursive interdependencies. Existing linkage identification algorithms only consider interactions between loci in a same building block and assume no interdependency between building blocks. In this paper, the LIEM² (Linkage Identification with Epistasis Measure considering Monotonicity) — a single layer linkage identification algorithm based on non-monotonicity conditions — is extended to identify hierarchical multi-layered linkage groups in order to search more accurate structures of real-world problems. The hierarchical linkage identification identifies linkage groups hierarchically in a recursive manner employing niching which preserve various building block candidates.

1. はじめに

遺伝的アルゴリズム (Genetic Algorithm, GA) において最も重要なオペレータは、交叉、すなわち部分解の組合せである。このことから、GA は問題がいくつかの部分問題の集合であることを仮定して探索を行っていると考えられる。広く信じられているビルディングブロック仮説では、GA は問題を部分問題に分割し、各部分問題における適応度の高い部分解に相当するストリングの部分列 (ビルディングブロック) を組み合わせることで全体の最適解を得るアルゴリズム

ムであることが示唆されている。

部分問題への準分解とその解の組合せによって解くことができる種類の問題に対して、GA は優れた性能を発揮することができる。しかし、一点交叉あるいは多点交叉を用いた単純 GA の場合、組合せ処理は各変数のストリング上の位置に大きく依存する。同一の部分問題に属する変数がストリング上で離れて配置されているとき、交叉はこれらの変数が同一の子個体に受け継がれることを保証せず、部分解を適切に組み合わせることができない。このような困難さを克服し、より効率的な探索を行うために、同一のビルディングブロックを構成する遺伝子座のストリング上の位置をあらかじめ調べ、この情報を用いて交叉を行う GA が提案されている。この手法をリンケージ同定と呼ぶ。リンケージ同定を導入した GA は、符号化が困難であり単純 GA では解くことのできない問題に対して有効に機能する。

[†] 北海道大学工学研究科システム情報工学専攻

Division of Systems and Information Engineering,
Graduate School of Engineering, Hokkaido University

^{††} 北海道大学情報基盤センター大規模計算システム研究部門

Division of Large-Scale Computational Systems, Information Initiative Center, Hokkaido University

より問題の規模が大きいとき、問題解決は部分問題への分割とその再統合という単純な過程ではなく、複数の部分解を集めることでより大きな部分解を構成するという再帰的な繰返しによってなされると考えられる。実際、多くの現実の問題がこのような階層的な構造をとることが報告されている⁷⁾。単純 GA は、選択による有用なスキーマの割合の増加による解モジュールの次数の増加によって、暗黙に階層型の探索を行っていると考えられる。上述のリンケージ同定 GA では、階層の最下位の層に対しては解モジュールの位置を調査し、それ以外の層では単純 GA と同様に暗黙に組合せを行っていると考えられる。しかし、小さな部分問題に対して前知識による符号化が困難な問題の場合、部分解の組合せによって構成されるより大きな部分問題に関しても、適切な符号化は困難である。

本論文では、階層型構造を持つ問題に対しても効率的な探索が行えるよう、リンケージ同定手法を拡張する。これにより、密な符号化が困難であり、かつ規模の大きい現実の問題をより効率的に解けるようなアルゴリズムの構築を試みる。また、実験を行いアルゴリズムの性能を調査する。

2. リンケージ同定

2.1 リンケージ同定手法

本論文では、同一の部分問題を構成し互いに依存関係のある変数の集合がとりうる値の組合せ(スキーマ)のうち、比較的短く高い適応度に寄与するものをビルディングブロックと呼び、ビルディングブロックの遺伝子座位置をリンケージと呼ぶ。また、同一のビルディングブロックを構成する遺伝子座の集合をリンケージ集合と呼ぶ。ビルディングブロックをストリング上で密に符号化できるかどうか、GA による探索の成否を左右する。一点交叉においては、ビルディングブロックがストリング上で離れた位置に配置されている場合(たとえば、1*****1)、高確率でビルディングブロック破壊が起こり、探索の効率は低下する。また、一様交叉はストリングの順序に依存しない交叉であり、どのように配置されたビルディングブロックも同時に交換することが可能だが、ビルディングブロック破壊のほうがビルディングブロック交換よりも高確率で起こると考えられる。

現実の問題はしばしば変数間の依存関係が複雑なため、ビルディングブロックをあらかじめ密に符号化することは困難である。そこで、そのような場合でも、適切なビルディングブロック交換を保証できるように GA を改良する必要がある。この目的で、LINC

(Linkage Identification by Nonlinearity Check³⁾、LIMD(Linkage Identification by non-Monotonicity Detection³⁾、LIEM(Linkage Identification with Epistasis Measure^{1),9)}、LIEM²(Linkage Identification with Epistasis Measure considering Monotonicity³⁾の一連の研究がなされ、遺伝子の値の摂動による適応度の変化から相互依存関係のある遺伝子座を同定する手法が提案されている。以下では、ストリングとしてビット列のみを考える。

まず LINC では、遺伝子座間に存在する非線形性を検出することでリンケージ同定を行っている。2つの遺伝子の値をまとめて変化させたときの適応度の変化量が、それぞれの遺伝子の値を変化させたときの適応度の変化量の和に等しいとき、これらの遺伝子座は個別に最適化して、後から組み合わせることが可能であると考えられる。逆に、適応度の変化量の和が加法的ではなく、非線形性を示すとき、これらの遺伝子の組はまとめて扱うべきであるとされる：ストリング s において、 i 番目の遺伝子の値 s_i の摂動 ($0 \rightarrow 1$ もしくは $1 \rightarrow 0$) を \bar{s}_i と書く。このとき、適応度の変化量はそれぞれ

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

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

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

で定義される。個体群中の少なくとも1つのストリングで、これらの関係が線形すなわち、

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

ではないとき、遺伝子座の組 (i, j) はまとめて扱うべきであるとされる。

非線形性が見られる場合でも、遺伝的オペレータに対して適応度が単調に増加もしくは減少するような場合には、山登り法的な探索が可能である。このような変数は GA にとって最適化が容易であり、これらの遺伝子座も個別に最適化することが可能であると考えられる。よって、LINC の非線形条件は非単調条件へ緩和されることができる。非単調性の検出によりリンケージを同定する手法が LIMD である。単調条件は式(1)–(3)を用いて以下のように定義される。

$$\begin{aligned} &\text{if } (\Delta f_i(s) > 0 \text{ and } \Delta f_j(s) > 0) \\ &\text{then } (\Delta f_{ij}(s) > \Delta f_i(s) \text{ and } \Delta f_{ij}(s) > \Delta f_j(s)) \end{aligned} \quad (5)$$

$$\begin{aligned} &\text{if } (\Delta f_i(s) < 0 \text{ and } \Delta f_j(s) < 0) \\ &\text{then } (\Delta f_{ij}(s) < \Delta f_i(s) \text{ and } \Delta f_{ij}(s) < \Delta f_j(s)) \end{aligned} \quad (6)$$

個体群中の少なくとも1つのストリングで上記の条件が満足されなかったとき、遺伝子座の組 (i, j) はま

とめて扱うべきであるとされる。

また、現実の問題では厳密な非線形性は取扱いが困難なことが多い。よって、非線形性の強さを定量化し、この大きさを用いてリンクージ同定を行う手法が LIEM である。同様に、非単調性を定量化して利用するものが LIEM² である。本論文では、LIEM² を階層型に拡張するが、本論文とほぼ同様の手順で上述の他の手法も階層型に拡張可能である。LIEM² に関しては次節で詳述する。

2.2 LIEM²

LIEM² は、摂動による適応度の非単調性が大きい遺伝子座の組には強い相互依存関係があると仮定する。この非単調性の強さをエピスタシス値 e_{ij} と呼ぶ。エピスタシス値は、単調条件からの逸脱の度合いを用いて、以下のように定義される。

$$e_{ij} = \max_{s \in P} g(\Delta f_{ij}(s), \Delta f_i(s), \Delta f_j(s)). \quad (7)$$

$$g(x, y, z) = \begin{cases} \begin{matrix} tr(y - x) + tr(z - x), & (y > 0, z > 0) \\ tr(x - y) + tr(x - z), & (y < 0, z < 0) \end{matrix} & (8) \\ 0, & (\text{上記以外}) \end{cases}$$

$$tr(x) = \begin{cases} x, & (x \geq 0) \\ 0, & (x < 0) \end{cases} \quad (9)$$

式 (7) における $\Delta f_{ij}(s)$, $\Delta f_i(s)$, $\Delta f_j(s)$ は、式 (1)–(3) で定義される適応度の変化量である。P は個体群を表す。これらの式から計算されるエピスタシス値 e_{ij} は、線形分離の困難さを表す。遺伝子座の組 (i, j) のうち、 e_{ij} が相対的に大きな (i, j) は同一のリンクージ集合に含まれるべきであると見なされる。個体長が l のとき、リンクージ集合は遺伝子座識別番号の集合 $\{0, 1, \dots, l-1\}$ の直和分割である。交叉のアルゴリズムを図 1 に示す。交叉は、同一のリンクージ集合に属する遺伝子座を同時に交換する。完全に独立な遺伝子座に関しては、1 つの遺伝子座が 1 つの集合を構成するものとする。

3. 階層型問題への拡張

3.1 階層型問題

部分解の組合せにより、より大きな部分解を構成するという再帰的な構造をとる問題を階層型問題と呼ぶ。階層型問題のモデルを図 2 に示す。図において、最下層の S_1 が各変数に対応する。

同一の部分問題に属する変数の集合は、ある特定の

- (1) 各親個体の組 (p_1, p_2) に対して以下を繰り返す
 (a) 各リンクージ集合 i に対して以下を繰り返す
 (i) 乱数 $[0, 1) < 0.5$ のとき
 子個体 (c_1, c_2) で
 集合 i のすべての要素 j に対して
 $c_1[j] = p_2[j]; c_2[j] = p_1[j];$ とする
 (ii) 乱数 $[0, 1) \geq 0.5$ のとき
 子個体 (c_1, c_2) で
 集合 i のすべての要素 j に対して
 $c_1[j] = p_1[j]; c_2[j] = p_2[j];$ とする

図 1 リンケージ情報を用いた交叉
 Fig. 1 Crossover with linkage information.

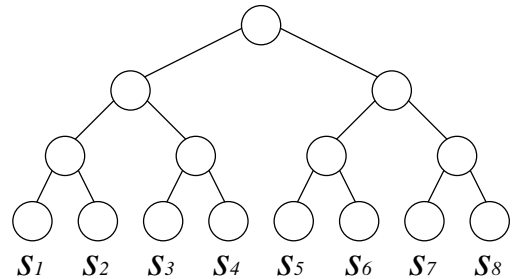


図 2 階層型問題
 Fig. 2 Hierarchical structure of variables.

値の組合せによって高い適応度に寄与する。このような値の組合せが部分解、あるいはその候補である。一般に、部分問題は複数存在し、各部分問題に対して解が存在する。部分問題の間に依存関係がなく、それぞれが完全に独立であるとき、各部分問題の最適解の組合せが問題全体の大域的最適解を与える。しかし、個々の部分問題がつねに完全に独立であるとは限らず、部分問題どうしが依存関係を持つことも考えられる。いい換えれば、ある部分問題がどの解候補を選択するかが、別の部分問題における解の適応度寄与に影響する可能性がある。

部分解の候補が選り抜かれているとき、1 つの部分問題を 1 つの変数と見なし、変数のとりうる値としてこれらの解候補を考えることができる。このように考えれば、部分問題間の関係を変数間の場合と同様にとらえることが可能である。さらに、互いに強く関連する部分問題の集合が複数存在するとき、この集合をより大きな部分問題と見なして、同様に考えることができる。

さまざまな研究から、現実の問題や実社会の構造がこのような階層性を持つことが示されている⁷⁾。本論

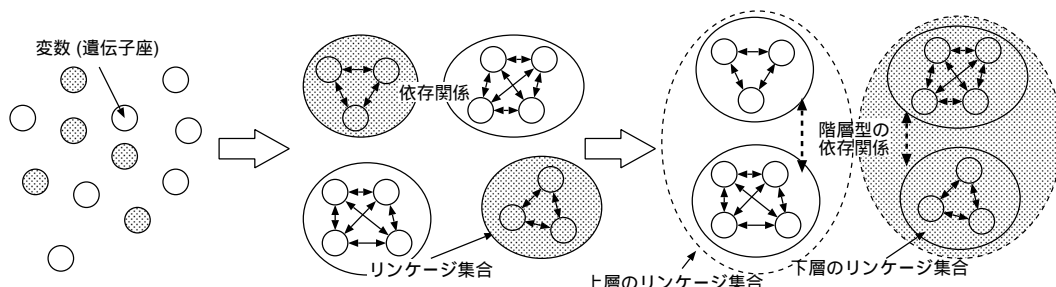


図3 一様交叉 → 従来のリンケージ同定手法 → 階層型への拡張
 Fig.3 Uniform crossover → single layer linkage identification → hierarchical linkage identification.

文では、この種類の問題に対しても効率的な探索が行えるよう、リンケージ同定手法を拡張する。

3.2 基本戦略

Watson ら⁸⁾ のビルディングブロックどうしの相互依存関係のモデル化とそれに対する GA の性能評価のための実験から、適切な符号化と個体群の多様性の保持が階層的な問題に対する解決の鍵であることが示唆された。

多くの問題においてあらかじめ適切な符号化を行うことは困難であると考えられるが、事前知識によらずに密に符号化すべき変数を同定する手法については、2章で述べたように、さまざまな研究例が存在する。図3に交叉の際の変数の扱いを示す。左は一様交叉による交叉を示す。図で \circ は各遺伝子座を表し、同色の遺伝子はそれぞれ同一の子個体に受け継がれるものとする。一様交叉の場合、変数の依存関係が考慮されず、各変数がどの子個体に受け継がれるかは独立に決定される。中央が既存のリンケージ同定手法による交叉である。 \leftrightarrow が変数の間の依存関係を示す。互いに依存関係を持つ遺伝子座はリンケージ集合としてまとめられ、交叉の際にはまとめて交換される。階層型問題に対応するためには、図の右に示すような変数の扱いが必要となると考えられる。図で点線の \leftrightarrow はリンケージ集合どうしの依存関係を示す。階層型の探索の場合、はじめは小さなリンケージ集合で定義されるビルディングブロックの組合せにより探索を行い、後に、それによって構成されたより大きなビルディングブロックの組合せによって探索を行う。これは、下層の小さな部品を組み合わせ、より上層の大きな部品を構成するような問題構造に対応する。よって、リンケージ同定は遺伝子座どうしの相互依存関係ではなく、より小さなリンケージ集合どうしの相互依存関係を対象に実行される。具体的な手法は3.3節で述べる。

上層の部分問題の探索は下層の部分問題の組合せに

よって行われる。しかし、下層の部分問題の解候補のうちで、最も高い適応度の寄与を与えるものの組合せがつねに全体の大域的最適解を構成するわけではない。上層で解を得るためには、下層の各部分解で解候補となるようなスキーマを複数用意し、これらのさまざまな組合せを試みる必要がある。よって、階層型リンケージ同定 GA には必要な解候補の多様性を保つための手法としてニッチングを導入する。具体的な方法については3.4節で述べる。

3.3 階層型問題のリンケージ同定

階層型のリンケージ同定は、前の段階までで得られているリンケージ集合に関して依存関係を調べ、マージすべき集合を決定する。初期状態では、1つの遺伝子座を要素として持つリンケージ集合を遺伝子座の数だけ構成する：初期状態においては、個体長が l のとき、 l 個のリンケージ集合が存在し、各リンケージ集合は要素として1つの遺伝子座識別番号 $\{0, \{1, \dots, \{l-1\}$ を持つ。

階層型リンケージ同定の過程を図4に示す。まず、前段階での各リンケージ集合 i に関して、それぞれの集合で定義されるスキーマからビルディングブロック候補 $\{B_i, B'_i\}$ を以下のように選択する：各リンケージ集合に対して、その集合が定義するスキーマの個体群中での出現頻度を調べる。出現頻度が上位のスキーマ2つをビルディングブロック候補とする。たとえば、リンケージ集合 $i = \{1, 2, 4\}$ すなわち $ff*ff**$ に対して、スキーマ $*11*1**$, $*10*0**$ が個体群中に多数の割合を占めるとき、 $\{B_i = *11*1**, B'_i = *10*0**\}$ となる。

次に、すべてのリンケージ集合ペア (i, j) に対してエピスタシス値 e_{ij} が計算される。 e_{ij} は式(7)-(9)を用いて計算される。ただし、 $\Delta f_{ij}(s)$, $\Delta f_i(s)$, $\Delta f_j(s)$ は遺伝子の値の摂動による適応度の変化量ではなく、ビルディングブロック候補の摂動 ($B_i \rightarrow B'_i$ もしくは

- (1) 各リンケージ集合 i に関してビルディングブロック (BB) 候補 $\{B_i, B'_i\}$ を得る .
- (2) すべてのリンケージ集合の組 (i, j) に対してエピスタシス値 e_{ij} を $e_{ij} = 0$ として初期化する .
- (3) ランダムに選択した部分個体群 S に含まれるすべての個体 s に対して以下を繰り返す
 - (a) すべてのリンケージ集合の組 (i, j) に関して
 - (i) 要素数があらかじめ定めた最大値を超えるようなリンケージ集合 i あるいは j に関して：なにもしない
 - (ii) s が BB 候補 $\{B_i, B'_i\}$ のいずれかを持つとき
 - (A) s が BB 候補 $\{B_j, B'_j\}$ のいずれかを持つとき：BB 候補を入れ換え，このときのエピスタシス値 e を計算する . $e_{ij} < e$ のとき $e_{ij} = e$ に更新
 - (B) s がリンケージ集合 j に関していずれの BB 候補も持たないとき：なにもしない
 - (iii) s がリンケージ集合 i に関していずれの BB 候補も持たないとき：なにもしない
- (4) すべてのリンケージ集合 i に対して， e_{ij} を j に関して降順にソートする . 上位 k 個 (要素数 k はあらかじめ指定する) のリンケージ集合を i と強い依存関係のある集合とみなして，集合 i とマージし，リンケージ集合を更新する

図 4 階層型リンケージ同定のアルゴリズム

Fig. 4 Algorithm of hierarchical linkage identification.

$B'_i \rightarrow B_i$) による適応度の変化量として定義される .

ここで，既存の手法の場合， e_{ij} は個体群中のすべての個体を調べることで決定された . しかし，階層型の場合は，個体群全体からランダムに部分個体群を選択し，部分個体群のみを調べることで e_{ij} を決定する . 部分個体群サイズは，全体の個体群サイズに比較して小さな値が用いられる . この理由は後の 3.6 節で計算コストに関する考察において述べる .

また，エピスタシス値 e_{ij} の計算は対象とするビルディングブロック候補を含まない個体に対しては実行されない . よって，部分個体群中のすべての個体でエピスタシス計算が実行されない (i, j) の組が存在することもある . ランダムに選択された部分個体群がビルディングブロック候補を含まないとき，そのリンケージ集合に対して探索が十分に行われておらず，集合のマージを行うのは尚早であると判断される .

e_{ij} が計算されたら，この値の大きさに基づいて，リンケージ集合がマージされる . リンケージに関する情報が更新され，以降では新たなリンケージ集合に基づいた交叉が実行される .

3.4 ニッチング

それぞれの層での組合せによる探索を保証するために，ニッチングが行われる . 提案手法は適応度の情報を用いてリンケージ集合を構成しているため，適応度を变化させるようなニッチング手法は望ましくない . ニッチングは置き換え手法を以下のように改良す

ることで実行される . 置き換えでは，限定トーナメント置き換え (Restricted Tournament Replacement, RTR⁶⁾) と呼ばれる手法を用いる . RTR では，各子個体に関して，個体群から部分個体群を選択し，子個体をこの部分個体群のうちで最も類似した個体と比較することで実行される . この部分個体群の大きさはウィンドウサイズと呼ばれる .

本手法では類似度の基準としてハミング距離を考える . 部分個体群の中から，新たな個体 s とのハミング距離が最も小さくなる s' を選択し， s と s' の適応度を比較して，より適応度の高い個体を個体群に戻す .

3.5 全体のアルゴリズム

以上のように，階層型リンケージ同定 GA では，ニッチングを導入した置き換えと，再帰的にビルディングブロックどうしの相互依存関係を考慮したリンケージ同定を用いられる . 全体のアルゴリズムを図 5 に示す . 図で $Lset[t]$ は世代 t のすべてのリンケージ集合である . 各世代で，それまでのリンケージ集合を対象にリンケージ同定を行いリンケージ集合を更新する . その後，新たなリンケージ集合に基づいて交叉が実行される .

3.6 計算コスト

既存の LIEM² の計算コストは個体長が l のとき $O(l^2)$ である . ここでは以下のような単純な階層型のモデルを考えて，hLIEM² のコストを評価する . 部分問題がすべての層で同一であるような再帰的な構造を

- (1) 個体群を初期化する
- (2) リンケージ集合の集合 $\text{lset}[0]$ を初期化する
- (3) 世代数 $t = 1$ とする
- (4) 終了条件が満たされるまで以下を繰り返す
 - (a) リンケージ集合の集合 $\text{lset}[t-1]$ から新たに $\text{lset}[t]$ を構成する
 - (b) $\text{lset}[t]$ を用いた交叉を適用する
 - (c) ニッチングを導入した置き換え (RTR)
 - (d) 突然変異を適用する
 - (e) ニッチングを導入した置き換え (RTR)
 - (f) $t = t + 1$ として (a) へ

図 5 全体のアルゴリズム
Fig. 5 Overall algorithm.

仮定する．個体長を l ，各部分問題の次数を k とすれば，総階層数は $\log_k l$ である．また，第 i 層 ($i = 0, 1, \dots$) での部分問題の数は l/k^{i-1} ，変数の数は l/k^i である．

部分問題の最大次数が k のとき，リンケージ集合を検出するために必要な個体数 n は $c2^k$ であることが知られている³⁾． c は定数である．ただし，上述のように階層型のリンケージ同定では部分個体群に対してしかエピスタシス計算を実行しない．部分個体群の個体数を n_1 とすれば，1 度のリンケージ同定で検出することができるリンケージ集合は全体の n_1/n である．次のリンケージ同定で，残りのリンケージ集合を検出することになる． i 層に関する j 回目 ($j = 0, 1, \dots$) のリンケージ同定で摂動の対象となるスキーマの数は $l/k^i(1 - j\frac{n_1}{n})$ 個である．1 度のリンケージ同定で $l/k^i(1 - j\frac{n_1}{n})$ 個の集合から可能な 2 つの組をすべて調査する．ゆえに，第 i 層に関する適応度評価回数 E_i は

$$E_i = \sum_{j=0}^{n/n_1-1} n_1 \left(l/k^i (1 - j\frac{n_1}{n}) \right)^2 = n_1 \left(\frac{l}{k^i} \right)^2 \left(1 + \frac{1}{6} \left(2\frac{n}{n_1} - 3 + \frac{n_1}{n} \right) \right) \quad (10)$$

である．これをすべての層に関して合計し，

$$E = \sum_{i=0}^{\log_k l-1} E_i = n_1 \left(1 + \frac{1}{6} \left(2\frac{n}{n_1} - 3 + \frac{n_1}{n} \right) \right) \times \sum_{i=0}^{\log_k l-1} \left(\frac{l}{k^i} \right)^2$$

$$= n_1 \left(1 + \frac{1}{6} \left(2\frac{n}{n_1} - 3 + \frac{n_1}{n} \right) \right) \frac{(l^2 - 1)}{(1 - \frac{1}{k^2})}$$

を得る．上式から，階層型に拡張したリンケージ同定の適応度評価回数は $O(l^2)$ であり，単層の場合と比較してオーダは増加しない．

また，項 $n_1 \left(1 + \frac{1}{6} \left(2\frac{n}{n_1} - 3 + \frac{n_1}{n} \right) \right)$ から，次数 k のリンケージの同定のために本来必要とされる個体数 n と比較してより小さい部分個体群 n_1 を選択することで，より小さい評価回数を得ることができる．しかし， n_1 が極端に少ないとき，正確なリンケージを同定する前にビルディングブロック破壊により重要なスキーマが失われる可能性もある．ただし，ニッチングによりこの可能性をある程度軽減することができる．

4. 実験

4.1 テスト関数

テスト関数として Watson ら⁸⁾ により提案された HIFF 関数と Pelikan ら⁶⁾ の BOA の階層型問題への拡張のテストとして提案された H-TRAP 関数を用いる．

4.1.1 Hierarchical If-and-Only-IF (HIFF)

HIFF の構造は 2 進木である． $\text{height}(x)$ により，ノード x からその前項の葉の 1 つまで距離を表現する．各葉は適応度に 1 だけ寄与する．各親ノード x は，その子が両方とも 1 か 0 だったとき，そしてそのときのみ，全体の適応度に $2^{\text{height}(x)}$ 寄与する．それ以外の $(0,1)$ ， $(1,0)$ や $-(\text{null})$ が含まれるとき，寄与は 0 である．2 つのシンボルを用いて親ノードへの入力定義される．この新たなシンボルの定義は変換関数 (transform function) によってなされる．変換関数は，入力が 2 つともが 0 のとき 0，2 つともが 1 のとき 1，それ以外は $-(\text{null})$ を出力する．葉はストリングによる入力をそのまま受け取る．この例を図 6 に示す．

4.1.2 Hierarchical TRAP Functions

階層型トラップ関数は，平衡 k -ary 木を構造として用いる．ただし， $k \geq 3$ である．変換関数はすべてのブロックが 0，もしくは 1 のとき，それぞれ 0 と 1 を出力し，それ以外は $-(\text{null})$ を出力する．

各階層の寄与関数は， k ビットのトラップ関数とする．ただし， k ビットの入力のうち 1 つでも $-(\text{null})$ があるとき，その関数は 0 を返す．トラップ関数はブロック中の 1 の数を u として次式で与えられる (図 7)．

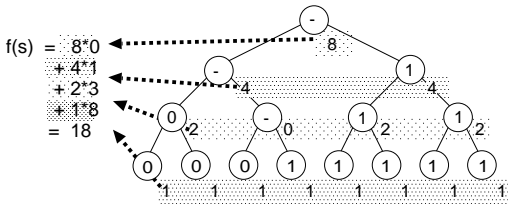


図 6 HIFF 関数
Fig. 6 HIFF function.

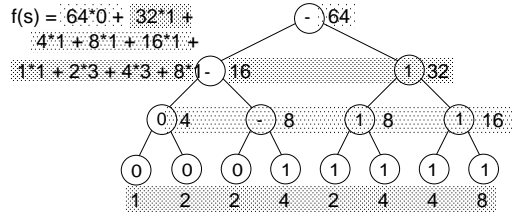


図 9 適応度の寄与が指数的な HIFF 関数
Fig. 9 HIFF function with exponential fitness contribution.

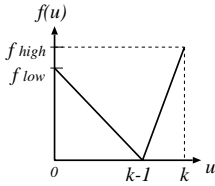


図 7 トラップ関数
Fig. 7 Trap function.

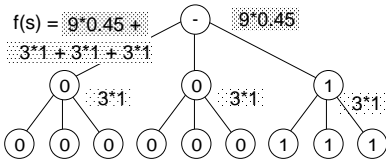


図 8 H-TRAP 関数
Fig. 8 Hierarchical trap function.

$$f(u) = \begin{cases} f_{high} & (u = k) \\ f_{low} - u \frac{f_{low}}{k-1} & (u \neq k) \end{cases} \quad (11)$$

階層型トラップ関数は図 8 のような関数である。葉は適応度に寄与しない。それ以外のノードの適応度関数への寄与は

$$\begin{aligned} f_{high} &= 1.0, \quad f_{low} = 0.9 \quad (\text{根ノード}) \\ f_{high} &= 1.0, \quad f_{low} = 1.0 \quad (\text{それ以外}) \end{aligned}$$

を用いて式 (11) で計算される。各部分関数には $k^{\text{height}(x)}$ で、重みがつけられる。最適解はすべての遺伝子が 1 をとるときである。以下では、この関数を H-TRAP1 と呼ぶ。例を図 8 に示す。

次に、根ノード以外のノードで f_{low} を f_{high} より大きくすることで、より強力にすべて 0 に向かうバイアスのかけた階層型トラップ関数を定義する。式 (11) は

$$\begin{aligned} f_{high} &= 1.0, \quad f_{low} = 0.9 \quad (\text{根ノード}) \\ f_{high} &= 1.0, \quad f_{low} = 1 + 0.1/K \quad (\text{それ以外}) \end{aligned}$$

を用いて計算される。K は階層型木構造の葉ではないノードの数である。各部分関数の重みは H-TRAP1 と同様である。この関数の最適解は H-TRAP1 と同

様にすべての遺伝子が 1 をとるとき得られる。この関数を H-TRAP2 と呼ぶ。

4.1.3 適応度の寄与が指数的な HIFF

部分関数の適応度寄与の寄与の違いは探索性能に影響を与える。同じ高さの層の部分関数の寄与が異なる関数を考える。ここでは、極端な場合として指数関数的に適応度の寄与が変化する HIFF を設定した。

この関数では、適応度寄与が $w(x_p)$ である親ノードの左右の子ノード x_r, x_l の重み $w(x_r), w(x_l)$ は

$$w(x_r) = w(x_p)/2 \quad (12)$$

$$w(x_l) = w(x_r)/2 \quad (13)$$

で与えられる。根ノード x_r の重み $w(x_r)$ は

$$w(x_r) = l^2 \quad (14)$$

で与えられる。l = 8 の場合の例を図 9 に示す。上の層への変換関数は通常の HIFF と同様である。

4.2 実験結果

HIFF および $k = 3$ の H-TRAP 関数に対して hLIEM² を適用し、hBOA (Hierarchical Bayesian Optimization Algorithm) との性能比較を行う。BOA は、適応度の高い部分個体群を多変量確率分布モデルであるベイシアンネットワークとしてモデル化し、交叉や突然変異の代わりに、モデルに従って新個体を生成するアルゴリズムである。変数間の依存関係は、ベイシアンネットワーク上の条件付き確率として表現される。この BOA を階層型の依存関係を扱うことができるように拡張したものが hBOA である。階層型の依存関係が探索可能なアルゴリズムは少なく、中でも hBOA は優れた性能を示しているため、提案手法の比較対称として hBOA を用いた。

実験は、Pelikan ら⁶⁾ と同様の手法で実行した。手法の性能は、問題を解くために必要な適応度評価回数で測定される。さまざまなサイズ (ストリング長) のテスト関数に対して 30 回の試行を行い、そのすべてで正しい解が得られたときの適応度評価回数の平均を調べた。個体数は正しい解が得られるまで徐々に増やした。

表 1 実験パラメータ

Table 1 Parameters for the experiments.

	n_1	p_c	p_m	w	k	MLE
HIFF	1	1.0	0.0	l	2	$l/2$
H-TRAP1	1	1.0	0.0	l	3	$l/3$
H-TRAP2	1	1.0	0.0	l	3	$l/3$
HIFF2	1	1.0	0.0	l	2	$l/2$

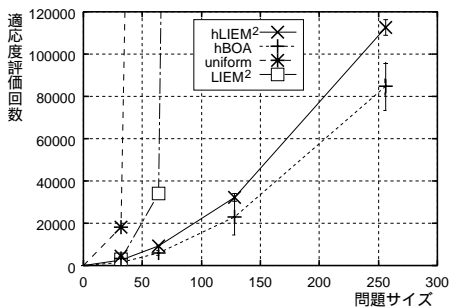


図 10 実験結果 HIFF
Fig. 10 Result of HIFF.

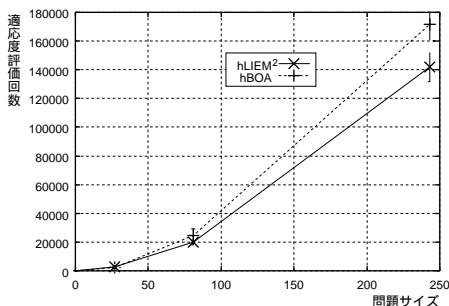


図 11 実験結果 H-TRAP1
Fig. 11 Result of H-TRAP1.

リンケージ同定のための個体数 n_1 , 交叉確率 p_c , 突然変異確率 p_m , ウィンドウサイズ w , 問題の次数 k , リンケージ集合の最大要素数 (MLE) を表 1 に示す . 表で l はストリングの長さである . 交叉の性能を測定するために , 突然変異確率は 0 に設定した . また , 今回用いたテスト問題では各部分関数が 2 もしくは 3 という低次数であるため , リンケージ同定に用いる個体数 n_1 は 1 に設定した .

また , hBOA の実験は Pelikan らの BOA のソースコード⁴⁾ を基に階層型に改造したものをを用いた . ウィンドウサイズ (すべて l) とトーナメントサイズ (すべて 2) は文献 5) と同様の値を用いた .

どちらのアルゴリズムに対しても , 探索の終了条件は少なくとも 1 つの最適解が発見されたときとした .

それぞれのテスト関数における問題サイズ (ストリング長) と評価関数の関係を図 10 , 図 11 , 図 12 , 図 13 に示す . 横軸は問題サイズ , 縦軸は各問題サイ

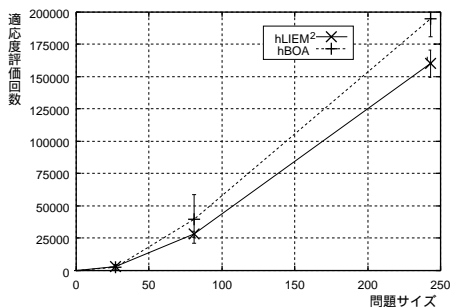


図 12 実験結果 H-TRAP2
Fig. 12 Result of H-TRAP2.

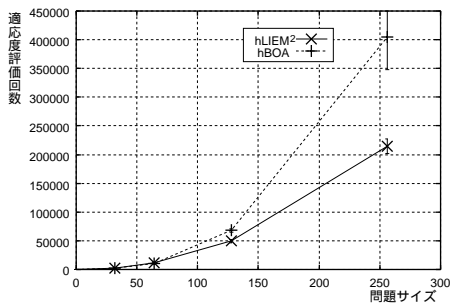


図 13 実験結果 適応度の寄与が指数的な HIFF
Fig. 13 Result of HIFF with exponential fitness contribution.

ズに関する評価回数の 30 回の実行における平均値 , 垂直線 (エラーバー) は標準偏差である .

HIFF に対しては , hLIEM² と hBOA のほかに , 一樣交叉による GA と単層の LIEM² による実験を行った . 結果を図 10 に示す . 一樣交叉のように変数の依存関係を考慮しない交叉では , 問題のサイズや層数が小さい $l = 32$ の場合ですら大きな計算回数が必要とする . 単層の LIEM² では , 最下層のリンケージのみが同定される . このリンケージ同定によって , $l = 32$ の HIFF は , $l = 16$ の HIFF に還元される . このサイズの小さい HIFF は , 階層型の場合とほぼ同等の計算回数で解くことができる . より問題サイズが大きいとき , 単層のリンケージ同定のみでは不十分であり , 問題を解くことができない . hBOA⁶⁾ と本手法とを比較すると , 適応度寄与が一樣な場合 (図 10 ~ 図 12) には HIFF 関数では hBOA が , H-TRAP1 , 2 関数では本手法が , より少ない評価回数で解を得ることができている .

これは , HIFF 関数は次数が低く , かつ騙し的でない比較的簡単な問題であり , ペイジアンネットワークによるモデル化が容易であるからであると考えられる . 一方で , 本手法は , 問題の性質が比較的簡単であるときも , 問題がトラップ性などの最も困難な性質を持つ

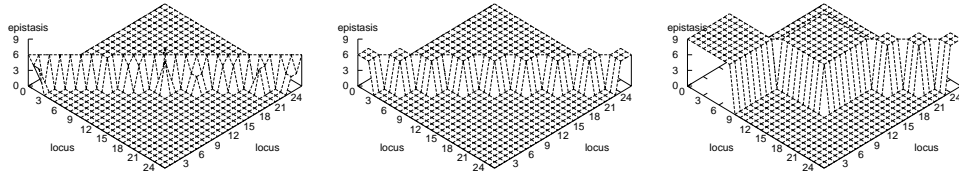


図 14 $l = 27$ H-TRAP1 でのエピスタシス値 : 1, 4, 8 世代
Fig. 14 Epistasis value in H-TRAP1, $l = 27$.

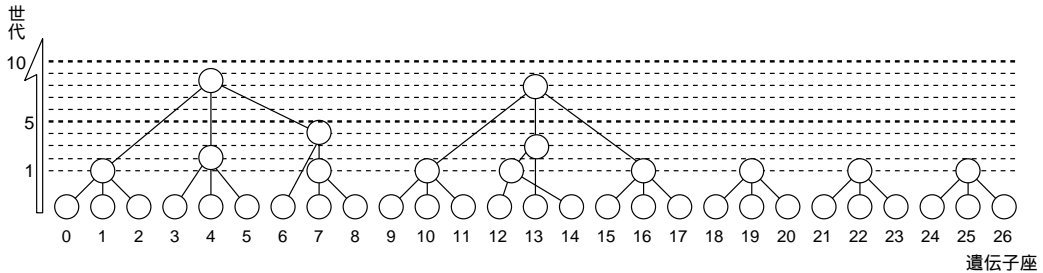


図 15 $l = 27$ H-TRAP1 でのリンケージ集合
Fig. 15 Linkage groups in $l = 27$ H-TRAP1.

ている場合と同等の処理を行うため、HIFF 関数の評価回数が hBOA と比較して多くなっているものと考えられる。

部分関数の寄与度が異なる問題(図 13)に関しては、両方の手法で通常の HIFF よりも大きな適応度評価回数を必要とした。しかし、増加の度合いは hLIEM² のほうが小さい。hLIEM² では、問題モデルは適応度の差分を計算して作成しているため、他の部分関数からの影響をある程度は無効にすることができる。よって適応度寄与の大きさの違いはリンケージ同定に対して、特に最初の層において、大きく影響を及ぼさない。ただし、適応度の大きさによって個体の置き換えが行われるため、各部分関数に対する探索全体の性能は適応度寄与の大きさの違いの影響を受ける。hBOA では、探索とモデル構成の双方で適応度の大きさが強く影響するため、寄与度の異なる問題では探索性能が低下すると考えられる。

問題サイズが 27 のときの H-TRAP1 関数におけるエピスタシス値(式(7))を図 14 に示す。 x 軸 y 軸の交点が各遺伝子座の組を表し、 z 軸が e_{ij} の大きさを表す。また、このときのリンケージ同定の過程を図 15 に示す。それぞれの円が遺伝子座を表し、円の下番号は遺伝子座のインデックスを表す。縦軸は世代数を表す。スキーマごとの収束のばらつきから階層構造の同定には時間差があるが、3 ビットからなる最下層のリンケージ集合が徐々にまとまって 9 ビット、27 ビットからなる上層の集合を構成しており、提案手法は問題の階層構造を適切に把握しているといえる。

5. おわりに

リンケージ同定の階層型への拡張を行った。階層構造を持ち、かつ各層で騙し的な問題に関して、リンケージ同定とニッチングを行った本手法は、有効に働いた。提案手法を hBOA と比較すると、HIFF 関数では BOA のほうが優れた性能を示したが、より困難な騙し性を持つ H-TRAP1, H-TRAP2 関数では本手法のほうが少ない評価回数で最適解に到達することができた。さらに、部分関数の寄与度の異なる階層型問題に関して、提案手法はより少ない評価回数の増加で解を得ることができた。

提案手法によって、これまでの GA では解くことが難しかった、大規模で問題どうしが依存関係を持つ問題を効率的に解くことができる。

参考文献

- 1) Munetomo, M.: Linkage Identification Based on Epistasis Measures to Realize Efficient Genetic Algorithms, *Proc. 2002 Congress on Evolutionary Computation*, pp.445–452 (2002).
- 2) Munetomo, M.: Linkage Identification with Epistasis Measure Considering Monotonicity Conditions, *Proc. 4th Asia-Pacific Conference on Simulated Evolution and Learning* (2002).
- 3) Munetomo, M. and Goldberg, D.E.: Identifying Linkage Groups by Nonlinearity/Non-monotonicity Detection, *Proc. 1999 Genetic and Evolutionary Computation Conference*, pp.433–440, Morgan Kaufmann Publishers

- (1999).
- 4) Pelikan, M.: A C++ Implementation of the Bayesian Optimization Algorithm with Decision Graphs, Technical Report IlliGAL Report No.2000025, Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL (2000).
 - 5) Pelikan, M.: Bayesian Optimization Algorithm: From Single Level to Hierarchy, Doctoral dissertation, University of Illinois at Urbana-Champaign (2002).
 - 6) Pelikan, M. and Goldberg, D.E.: Escaping Hierarchical Traps with Competent Genetic Algorithms, Technical Report IlliGAL Report No.2001003, Illinois Genetic Algorithms Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL (2001).
 - 7) Ravasz, E. and Barabasi, A.-L.: Hierarchical organization in complex networks, *Physical Review E*, Vol.67, No.2, pp.026112-1 – 026112-7 (2003).
 - 8) Watson, R.A., Hornby, G.S. and Pollack, J.B.: Modeling Building-Block Interdependency, Parallel Problem Solving from Nature – PPSN V, Lecture Notes in Computer Science 1498, pp.97–106, Springer-Verlag (1998).
 - 9) 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 - 004, pp.6–13 (2002).

(平成 15 年 4 月 9 日受付)

(平成 15 年 6 月 2 日再受付)

(平成 15 年 7 月 11 日採録)



辻 美和子

昭和 53 年生。平成 14 年北海道大学工学部情報工学科卒業。現在同大学大学院工学研究科システム情報工学専攻修士課程在学中。



棟朝 雅晴 (正会員)

昭和 43 年生。平成 8 年北海道大学大学院工学研究科情報工学専攻修士後期課程修了。同年同大学院工学研究科システム情報工学専攻助手。平成 10 年～平成 11 年イリノイ大学基礎工学科遺伝的アルゴリズム研究室客員研究員。平成 11 年北海道大学情報メディア教育研究総合センター助教授。平成 15 年同大学情報基盤センター助教授 (大規模計算システム研究部門)。博士 (工学)。進化的計算, 分散システムに関する研究に従事。IEEE 会員。



赤間 清 (正会員)

昭和 48 年東京工業大学工学部制御工学科卒業。昭和 50 年同大学大学院修士課程修了。昭和 54 年同大学院博士課程単位取得退学。同年同大学助手。昭和 56 年北海道大学文学部講師。平成元年同大学工学部助教授。平成 11 年同大学情報メディア教育研究総合センター教授。平成 15 年同大学情報基盤センター教授 (大規模計算システム研究部門)。工学博士。人工知能, 知識処理, 等価変換に基づく問題解決の研究に従事。ソフトウェア科学会会員。