

## 進化的計算における並列化手法の比較検討

村尾直哉 † 棟朝雅晴 †† 赤間 清††

近年、並列遺伝的アルゴリズムに関する研究が行われているが、特定の問題に関する性能評価を行ったものが多く、問題の性質と並列化手法との関係についてはあまり知られていない。本論文では、適応度重みに基づいた問題の性質と既存の並列遺伝的アルゴリズムとの関係についての考察を行う。また、提案するリンケージ同定に基づく並列遺伝的アルゴリズムとの比較を行い、与えられた問題に対して適用すべき並列化手法とは何かという、並列遺伝的アルゴリズムを設計するための指針となるものを提示する。

### Investigations on Parallelizing Methods of Evolutionary Computation

NAOYA MURAO ,† MASAHARU MUNETOMO †† and KIYOSHI AKAMA††

Parallel genetic algorithm (PGA) is an active research area in evolutionary computation and a number of papers have been published, however, relation between the nature of the problems and the method to be applied to solve them. Most papers deal with a tailored version of PGA to solve a specific problem. In this paper, we perform empirical investigations on the relation between characteristics of problems based on building blocks weights and selection of PGA methods. We also compare conventional methods with a novel approach of PGA based on linkage identification. Final goal of the paper is to show a guideline to users choosing one of the PGA methods based on the characteristics of the problems to be solved.

#### 1. はじめに

近年、進化的計算の一種である遺伝的アルゴリズム (Genetic Algorithm, GA) が組み合わせ最適化問題などを解くために用いられており、また、GA の並列化に関する研究も数多く行われている。しかし、多くの並列 GA に関する研究は、特定の問題に対してのみ行われており、問題の性質と各並列化手法との関係についてはあまり考慮されていない。

本稿では、並列 GA の性能を調べ、問題の性質と各並列化手法との関係について考察する。また、提案するリンケージ同定に基づく並列遺伝的アルゴリズム (以下、並列リンケージ同定) との比較を行い、どの問題に対してどの並列化手法を用いるべきかという、並列 GA の設計指針について検討する。

#### 2. 遺伝的アルゴリズム

GA は自然選択と遺伝学に基づく進化的手法である。GA では、解の候補となる個体をビット列などの文字列で表現し、複数の個体を扱う (この個体数を集団サイズと呼ぶ)。また、個体を判断するための基準となる適応度を求める計算 (評価) を行う。この適応度をもとにして、良い個体を選び (選択)、その個体の良い特徴 (ビルディングブロック, BB) を組み合わせる (交叉) ことによって最適解を探索する。GA の実行過程は以下ようになる。

- (1) 複数の個体からなる集団  $P(0)$  を初期化する。
- (2)  $P(0)$  中の個体の適応度を評価する。  $t=0$  とする。
- (3)  $P(t)$  中の個体に選択を適用し、 $P'(t)$  を生成する。
- (4)  $P'(t)$  中の個体に交叉を適用し、 $P''(t)$  を生成する。
- (5)  $P''(t)$  中の個体に突然変異を適用し、 $P'''(t)$  を生成する。
- (6)  $P(t)$  と  $P'''(t)$  をあわせた集団から、次世代集団  $P(t+1)$  を生成する。
- (7) 終了条件が満たされていなければ終了。満たされていないならば、 $t=t+1$  として (3) へ。

† 北海道大学工学部情報工学科

Department of Information Engineering, Faculty of Engineering, Hokkaido University

†† 北海道大学情報メディア教育研究総合センター

Center for Information and Multimedia Studies, Hokkaido University

GA はその実行時間の大半を個体の評価に費やしており、集団サイズが大きくなるにつれ実行時間が長くなる問題がある。しかし、評価など個体単位での操作が多く、独立性が高いので並列化が容易である。次章では、並列 GA について説明する。

### 3. 並列遺伝的アルゴリズム

Cantú-Paz<sup>1)</sup> の研究では、並列 GA は以下の 4 つに分類される。

- マスタースレーブモデル型並列 GA
- 島モデル型並列 GA
- 超並列 GA
- 階層型並列 GA

本稿では主にマスタースレーブモデル型並列 GA と島モデル型並列 GA について検討する。

マスタースレーブモデル型並列 GA とは、1 つの主となるプロセッサが通常の遺伝オペレータ (選択・交叉・突然変異) を操作し、その他の複数のプロセッサが適応度評価を分散して行うものである。ここで、選択・交叉・突然変異に要する時間を無視し、実行時間が完全に評価時間と通信時間の和からなる場合を想定し、マスタースレーブモデル型並列 GA における 1 世代の実行時間について考える。このとき、1 世代の実行時間  $T_p$  は、

$$T_p = \frac{nT_f}{P} + PT_c$$

となる ( $T_c$  は 1 プロセッサとの通信時間、 $T_f$  は 1 個体の評価時間、 $P$  は用いるプロセッサ数、 $n$  は集団サイズ)。この式からもわかるように、評価にかかる時間はプロセッサが増加するにつれて減少するが、通信時間はプロセッサとともに増加する。これより、実行時間を最小化する最適プロセッサ数が存在し、適切なプロセッサ数を用いなければ逆に性能が悪くなってしまふ場合が生じる。最適なプロセッサ数を求めるためには、上式を  $P$  で微分したものについて、 $\frac{\partial T_p}{\partial P} = 0$  を解けばよい。この式から、最適なプロセッサ数  $P^*$  は、

$$P^* = \sqrt{\frac{nT_f}{T_c}}$$

となる。

また、並列 GA の性能を示す指針として用いられる台数効果  $S$  は、

$$S = \frac{T_s}{T_p} = \frac{nT_f}{\frac{nT_f}{P} + PT_c}$$

と書ける。この理論式において集団サイズを  $n = 1000$  とし、評価時間対通信時間の比を変化させて得た台数効果は図 1 である。この図が示すように、(評価時間)  $\gg$  (通信時間) の場合に、マスタースレーブモデル

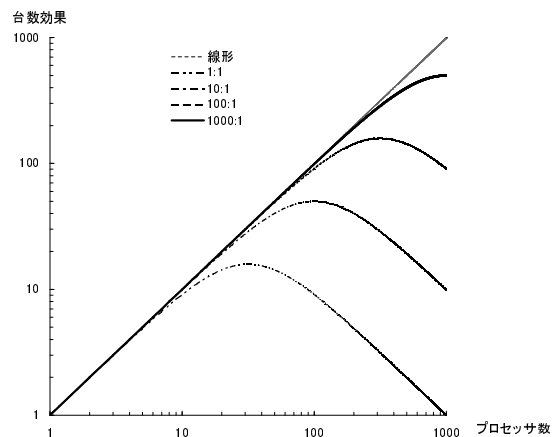


図 1 マスタースレーブモデル型並列遺伝的アルゴリズムの台数効果

型並列 GA が効果的であることを示している。このように、マスタースレーブモデル型並列 GA は、実装が簡単であるが、実行時間を大きく減少することができる。

島モデル型並列 GA は、集団をいくつかの部分集団に分割し、各部分集団で独立して GA を処理する手法である。また、ある条件 (例えば、各部分集団が収束した時) のもとで移住と呼ばれる個体交換を行うことで、各部分集団間の通信を行う。このモデルが有効に働くのは、各部分集団が独立に異なる探索を行うことで解の多様性を保存でき、局所解に陥りにくいためである。この島モデル型並列 GA の台数効果について実験的に得た結果は図 2 である。この図はマス

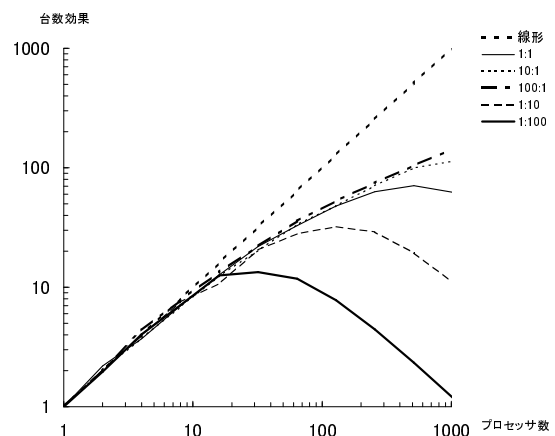


図 2 島モデル型並列遺伝的アルゴリズムの台数効果

タースレーブモデル型並列 GA と同様に、評価時間対通信時間を変化させた時の台数効果を測定したものである。結果は、評価時間が非常に大きい時にはマスタースレーブモデル型並列 GA よりも台数効果が劣る

結果になっているが、これは集団を分割することにより、各部分集団の収束時間は減少するが、各部分集団で得られる異なる解を移住により交換し、再び遺伝オペレータを処理することで全体として実行時間が長くなっているためであると考えられる。また、島モデル型並列 GA では、マスタースレーブモデル型並列 GA ほど頻繁に通信が行われないので、通信時間が大きい時には島モデル型並列 GA が有効的であるといえる。

次節では、本稿で提案する並列化手法について説明する。

#### 4. リンケージ同定に基づく並列遺伝的アルゴリズム

通常の GA では、交叉により各個体中の BB が破壊される可能性がある。これを防ぐために、これらの BB を形成する部分 (リンケージ) をまとめて扱うことが考えられる。これまでの研究では、遺伝子座間の摂動を用いて相互依存性を検出するリンケージ同定に基づく遺伝的アルゴリズム (LINC,<sup>2)</sup>LIMD<sup>3)</sup> が提案されている。具体的には、ある個体  $s$  とその個体の遺伝子座  $i$  を摂動させた個体  $s_i$  の適応度の差  $\Delta f_i$  を計算する。同様にして、 $\Delta f_j, \Delta f_{ij}$  を計算し、その値が非線形性を示す遺伝子座の組をリンケージ集合として扱う。

$$\begin{aligned}\Delta f_i &= f(s_i) - f(s) \\ \Delta f_j &= f(s_j) - f(s) \\ \Delta f_{ij} &= f(s_{ij}) - f(s) \\ \Delta f_{ij} &\neq \Delta f_i + \Delta f_j\end{aligned}$$

これを、すべての遺伝子座の組  $(i, j)$  について行い、また 1 個体では現れない非線形性が存在する可能性があるため、用いる  $O(c2^k)$  個の個体について適用してリンケージ集合を作成する ( $c$  はある定数、 $k$  は BB のサイズ)。これにより得られたリンケージ集合をもとに、GA を適用することで、符号化が適切に行われていない問題についても解くことができることがわかっている。しかし、リンケージ同定ではすべての遺伝子座の組、すべての個体について調べているので、個体評価に  $O(nl^2)$  だけかかる ( $n$  は集団サイズ、 $l$  は個体長) が、各遺伝子座の組における非線形性を検出する過程は独立して行うことができ、並列化が容易である。また、リンケージ同定により得られるリンケージ集合内で GA を処理する過程においても、各リンケージ集合を分散して処理することが可能である。これらの処理を並列化することにより、高速に解を求めることが可能となる。図 3 は、並列リンケージ同定の台数効果について、評価時間対通信時間の比を変えて実験的に

得た結果を示している。この図が示すように、並列リ

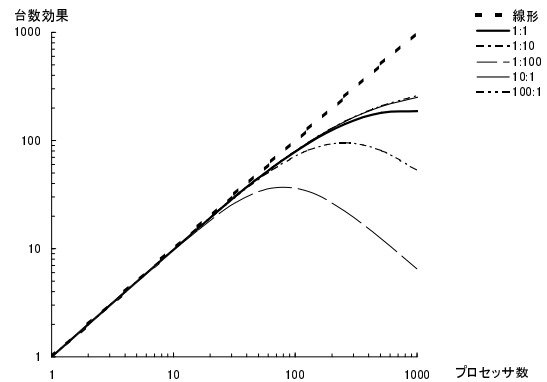


図 3 評価時間対通信時間に関する並列リンケージ同定の台数効果

ンケージ同定では 100 台近くのプロセッサを用いた場合でも、線形に近い台数効果を得ることができている。また、リンケージ同定は各遺伝子座の組ごとで並列化が可能であることから、個体長によってもその台数効果が変化することが考えられる (図 4)。これらの結果

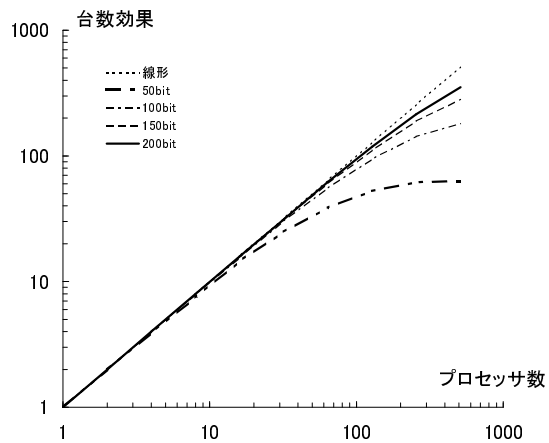


図 4 個体長に関する並列リンケージ同定の台数効果

と、先に示した 2 つの並列 GA の台数効果を比較してわかるように、並列リンケージ同定の台数効果は非常に高いものである。これは、各遺伝子座の組ごとに並列化できることから、その並列化可能なプロセッサ数が非常に多く、より多くのプロセッサ数を用いることができるためである。また、通信回数も少ないことが並列リンケージ同定の台数効果を高くしている。さらに、個体長が長くなるにつれ並列化可能なプロセッサ数が増加するため、台数効果を高くなっている。

次章では、ここで提案した並列リンケージ同定と先に説明した既存の並列 GA について、問題の性質との

関係を考察するための比較実験を行う。

## 5. 実験

この章では、適応度重みの基づく問題の性質<sup>4)5)</sup>と並列化手法の性能についての比較実験を行う。対象関数として5ビットトラップ関数を用いる ( $u$  は5ビット中の'1'の数)。

$$\begin{aligned} \text{if } u = 5, \quad & f_i(u) = 1.0 \\ \text{else,} \quad & f_i(u) = \frac{4-u}{5} \end{aligned}$$

この部分関数を複数個 ( $n$  個) 用い、さらに次の式のように重み  $\omega_i (i = 1, \dots, n)$  をつけ、その総和を個体の適応度  $F$  とする。

$$F = \omega_1 f_1(u) + \dots + \omega_n f_n(u) \quad 0 < \omega_i \leq 1$$

この重みの分布がほぼ一様であるとき ( $\omega_{i+1} = \omega_i, i = 1, \dots, n-1$ ), GA は異なる探索過程を示し、島モデル型並列 GA において解の多様性を保存したまま探索を進めることが可能である。一方、重み  $\omega_i$  の分布が指数的である場合 ( $\omega_1 = 1.0, \omega_i \leq r\omega_{i-1}, i = 2, \dots, n, r$  はある定数), もっとも重みの大きな部分から順に解が求められていくことが予想され、島モデル型並列 GA が効果的に動作せず、解を得るためには非常に大きな集団サイズを必要とする。

まず、重みの分布と島モデル型並列 GA の性能に関する実験を行う。この図は係数  $r$  を変更させた場合に

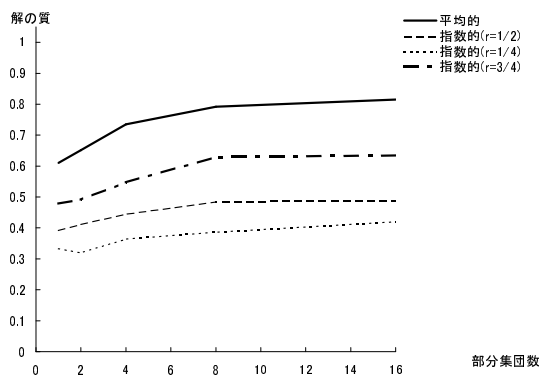


図 5 重みの分布と島モデル型並列 GA の効果

ついて調べており、係数  $r$  が 1 に近いほど、重みの分布は一様に近くなることを表している。この図が示すように、 $r$  が小さくなるほど島モデル型並列 GA の効果は低いことがわかる。この原因は、上で述べたように解の多様性を保てないためであると考えられる。このことを示すために、 $r = 1$  と  $r = 1/2$  の場合に関して、それぞれの探索過程を調べる。図 6,7 は、各部分集団においてそれぞれの部分解 (BB) が正しく得られ

た割合 (解の質と示す) の推移を示しているが、重みの分布が一様な場合では異なる部分解から探索ができているが、重みの分布が指数的な場合にはほぼ同一の部分解から探索をしていることがわかる。

これに対し、並列リンケージ同定では、リンケージ同定により得られるリンケージ集合ごとの探索を行うので、それぞれの部分解を適切に求めることが可能となる。これらのことをもとに、解を得るために必要な十分な大きさの個体群を用意して、島モデル型並列 GA と並列リンケージ同定では、どちらが高速に解を求めることができるかを比較したものが図 8, 9 である (用いたプロセッサ数は 8 個)。重みの分布が指数的な場合、個体長が長くなるにつれ島モデル型並列 GA よりも並列リンケージ同定が高速に解を求めることができることを示している。これは、個体長が長くなるにつれ、島モデル型並列 GA が必要とする個体サイズが指数的に大きくなり、そのために評価回数も多くなるためである。一方、並列リンケージ同定では、個体長が長くなることで評価回数が増加するが、用いる個体数はほとんど変化しない。このように、重みの分布が指数的な場合には並列リンケージ同定が優れている。しかし、重みの分布が一様である場合、島モデル型並列 GA は解の多様性を保ちながら効果的に動作することができ、並列リンケージ同定よりも高速に解を求めることができる。

これらの図が示すように、重みの分布が指数的である場合には並列リンケージ同定を用いると高速に解を求めることができ、重みの分布が一様である場合には島モデル型並列 GA を用いると良いことがわかるが、先に述べたように並列リンケージ同定の台数効果は大きいので、重みの分布が一様である場合に多くのプロセッサを用いることで、その評価回数の差を小さくすることができ、島モデル型並列 GA とほぼ同様の時間で解を求めることが可能となる。図 10 は、個体長を 200 ビットとした場合に、用いるプロセッサ数を大きくして得た結果である。

## 6. 結論

本稿では、マスタースレーブモデル型並列 GA と島モデル型並列 GA、提案する並列リンケージ同定について、その並列化効果について考えてきた。また、適応度重みに基づく問題の性質から、適用すべき並列化手法を決定するための比較実験を行ってきた。前節の結果から、通常の GA では適応度寄与の高い BB の影に隠れてしまい探索が困難になる BB が存在する、重みの分布が指数的な場合に関してもリンケージ同定を

用いることで、各 BB ごとの探索ができるので解を適切に求めることができる。またリンケージ同定は並列化が容易であり、台数効果が非常に大きく、評価対通信比だけでなく、解空間の大きさ(個体長)によってもその台数効果を大きくできる。

本稿の結果は以下のようなものである。

- (評価時間)≫(通信時間) である場合には、マスタースレーブモデル型並列 GA を用いることで実行時間を短縮することができる。
- 重みの分布が一様である場合には、島モデル型並列 GA を用いることで、解の多様性を保存したまま探索ができ有効的である。
- 重みの分布が指数形である場合には、並列リンケージ同定を用いることで各 BB ごとの探索が可能となり、島モデル型並列 GA よりも高速に解を求めることができる。
- 利用可能なプロセッサが非常に多い場合には、重みの分布にかかわらず並列リンケージ同定を用いることで解を高速に求めることができる。

今後の課題として、重みに関する考察をより詳細に進めていくことや、他の進化的計算における並列化手法との比較を行うつもりである。

### 参 考 文 献

- 1) Erick Cantú-Paz. Designing Efficient and Accurate Parallel Genetic Algorithms. Technical Report IlliGal Report No.99017, University of Illinois at Urbana-Champaign, July 1999.
- 2) Masaharu Munetomo and David E. Goldberg. Designing a Genetic Algorithm Using the Linkage Identification by Nonlinearity Check. Technical Report IlliGal Report No.98014, University of Illinois at Urbana-Champaign, December 1998.
- 3) Masaharu Munetomo and David E. Goldberg. Linkage identification by non-monotonicity detection for overlapping functions. Technical Report IlliGal Report No.99005, University of Illinois at Urbana-Champaign, January 1999.
- 4) David E. Goldberg. Using Time Efficiency: Genetic-Evolutionary Algorithms and the Continuation Problem. Technical Report IlliGal Report No.99002, University of Illinois at Urbana-Champaign, January 1999.
- 5) Ravi P. Srivastava. Time Continuation in Genetic Algorithms. Technical Report IlliGal Report No.2001021, University of Illinois at Urbana-Champaign, June 2002.

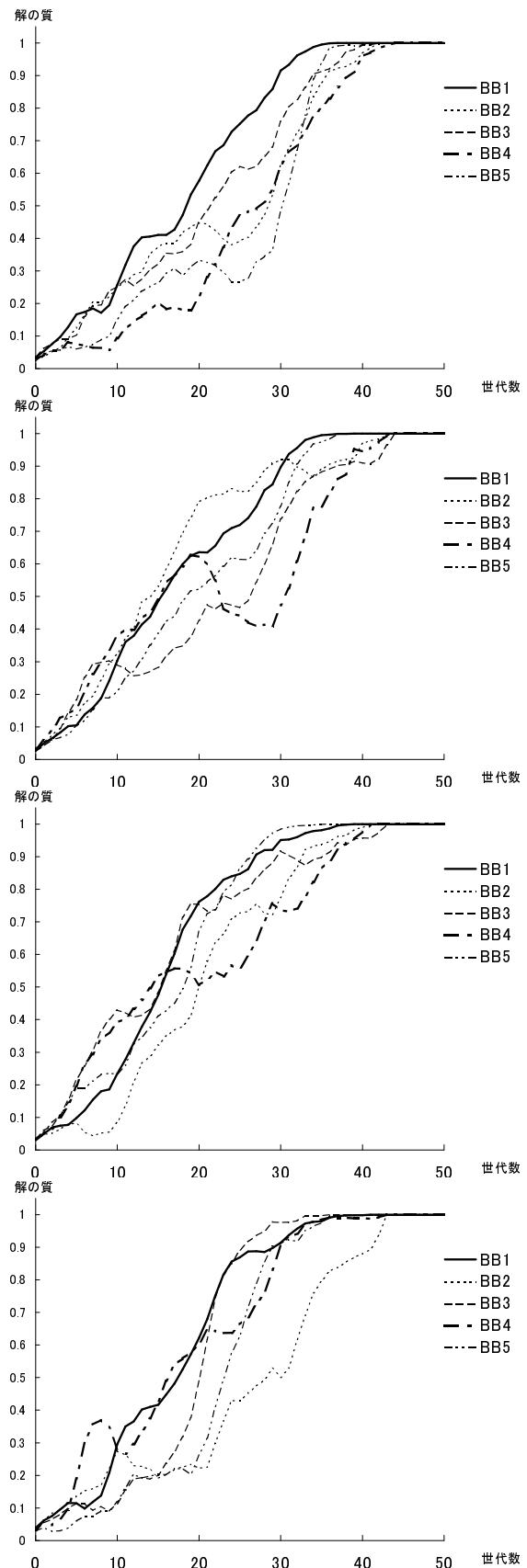


図 6 重みの分布が一様な場合の島モデル型並列 GA の探索過程

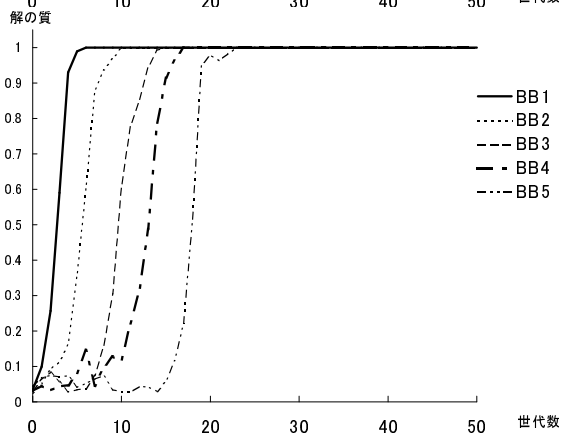
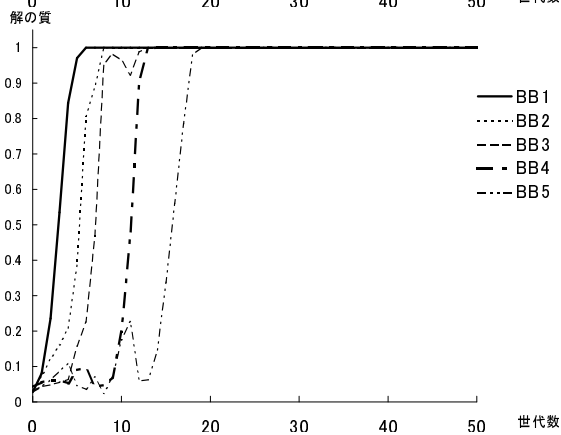
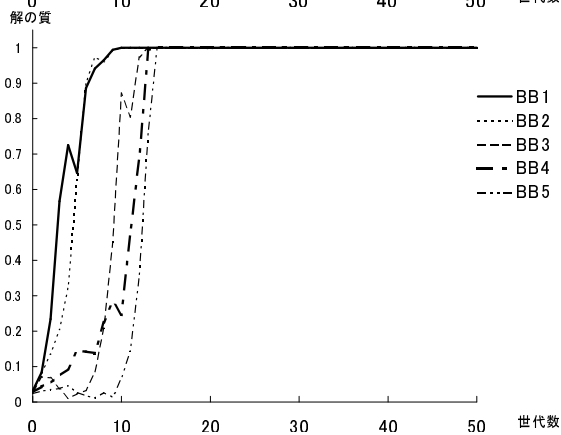
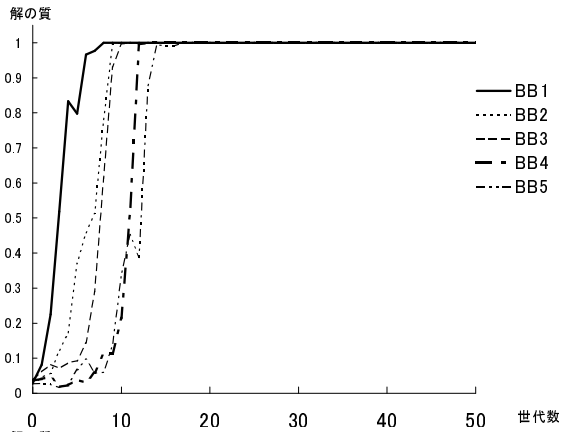


図7 重みの分布が指数的な場合の島モデル型並列 GA の探索過程

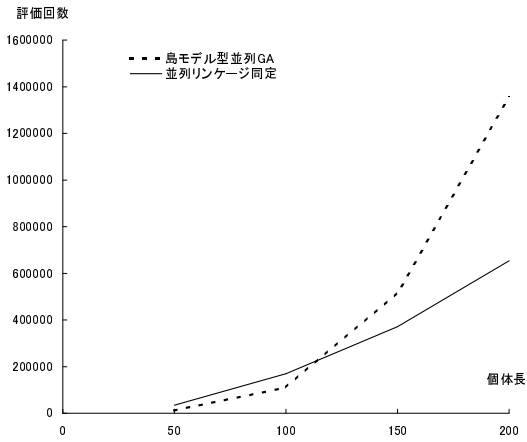


図8 重みの分布が指数的な場合の島モデル型並列 GA と並列リンクージ同定の性能比較 .

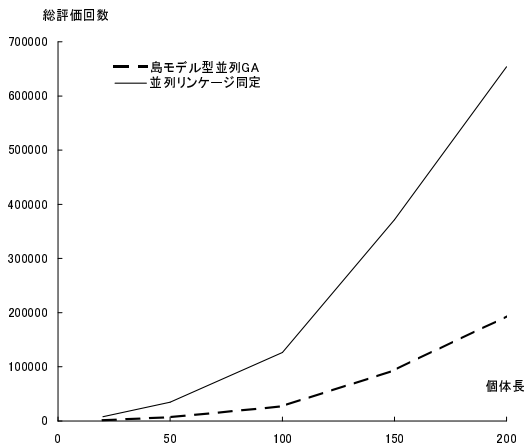


図9 重みの分布が一樣な場合の島モデル型並列 GA と並列リンクージ同定の性能比較 .

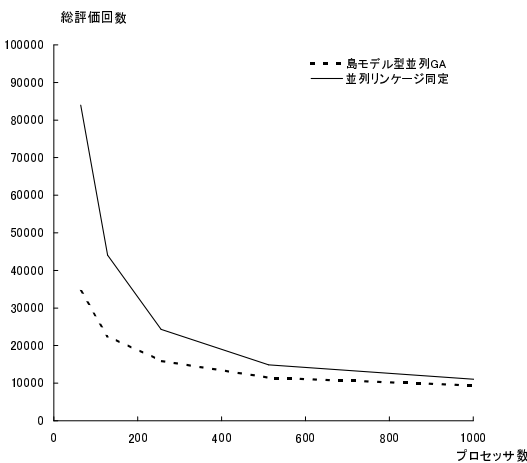


図10 重みの分布が一樣な場合に用いるプロセッサを大きくした場合についての島モデル型並列 GA と並列リンクージ同定の性能比較 .